Supporting Information

Guha *et al.* 10.1073/pnas.0806158105

SI Methods

Cell Culture, SILAC, and Lysate Preparation. HBECs were grown in keratinocyte serum free medium (KSFM), supplemented with EGF and bovine pituitary extract (BPE) as described in ref. 1. Arginine- and lysine-free KSFM were bought from Invitrogen. A detailed protocol for SILAC media preparation can be obtained from http://www.silac.org. The Arg/Lys-free KSFM was supplemented with light (normal Arg/Lys), medium ($^{13}C_6$ Arg/D^4 Lys; +6 Da and + 4 Da shift for labeled arginine- and lysine-containing peptides, respectively), and heavy ($^{13}C_6{}^{15}N_4$ Arg/$^{13}C_6{}^{15}N_2$ Lys; +10 Da and + 8 Da shift for labeled arginine- or lysine-containing peptides, respectively) amino acids for adaptation of HBECs expressing WT EGFR, KRAS G12V, and DelE746-A750 EGFR, respectively, in one experiment, and WTEGFR, L858R EGFR, and DelE746-A750 EGFR, respectively, in another experiment. The adenocarcinoma cells H2030 and H1650 were grown in RPMI-containing light (normal Arg/Lys) and heavy ($^{13}C_6$ Arg/$^{13}C_6$ Lys; +6Da and + 6Da shift for labeled arginine and lysine, respectively) amino acids. All adenocarcinoma cell lines, except H3255 (a kind gift from Bruce Johnson), were purchased from ATCC.

The cells were adapted for at least five generations before scaling up for experiments. Between 5×10^7- 10^8 cells from each "state" (relatively equal number of cells from each state in each experiment) were serum-starved overnight before lysis in modified RIPA buffer (50 mM Tris·HCl, pH 7.4 containing 1% Nonidet P-40, 0.25% Na-deoxycholate, 150 mM NaCl, 1 mM EDTA, protease inhibitor mixture tablets (Roche), 1 mM Na3VO4, NaF 1 mM). Protein estimates were made by the Lowry method (BioRad), and equal amounts of protein from lysates of each state were mixed together for phosphotyrosine immunoprecipitation.

For the experiment using the lung adenocarcinoma lines, H2030 and H1650, $\approx 10^8$ cells were serum starved overnight and then stimulated with 10 ng/ml EGF before harvesting in modified RIPA buffer. Equal amounts of protein from each of the two states were then mixed together.

Antibodies, Immunoprecipitation, and Western Blots. Tyrosine-phosphorylated proteins and proteins strongly interacting with phosphorylated proteins were immunoprecipitated from mixed lysates by a mixture of two antiphosphotyrosine antibodies, 4G10-agarose (Millipore) and RC20 biotin (BD Biosciences). Proteins were eluted with 100 mM phenyl phosphate and dialyzed against water.

PhosphoScan kit (Cell Signaling Technology) was used for phosphopeptide enrichment from mixed cellular lysates according to the manufacturer's protocol. For enrichment of phosphotyrosine-containing peptides, the Cell Signaling kit was used for immunoprecipitation, following the manufacturer's instructions. Immunoprecipitations were performed with commercial antibodies to EGFR (clone 528 hybridoma, ATCC) and to ERBB2 (Santa Cruz). The 4G10-HRP conjugate (Millipore) was used for Western blot analysis for phosphotyrosine.

Mass Spectrometry. The phosphotyrosine immunoprecipitates were eluted with phenyl phosphate, dialyzed against water, run in 1D SDS/PAGE gel, and stained with colloidal Coomassie (Invitrogen). Gels were cut into 20–30 pieces, followed by in-gel trypsin digestion (Promega) and extraction of peptides. For enrichment of tyrosine phosphorylated peptides, we used in-solution trypsin digestion with the PhosphoScan kit (Cell signaling technology), following the manufacturer's instructions.

The peptide samples from both strategies were analyzed using reversed phase nanoscale liquid chromatography, interfaced with a QSTAR Pulsar (Applied Biosystems) quadrupole-time-of-flight mass spectrometer. RP-LC system consisted of a trap column (75 μm $\times$ 3 cm, C18 material 5–10 μm, 120Å, YMC) and an analytical column (75 μm $\times$ 10 cm, C18 material 5 μm, 120Å, YMC) with an emitter tip 8 μM (New Objective) attached. The peptides were eluted using an organic solvent gradient from 5 to 40% acetonitrile in 0.1% formic acid, for 30 min, with a flow rate of 300 nl/min. The MS spectra from m/z 350 to 1200 were acquired in a data-dependent manner, targeting the three most abundant ions in the survey scan using a dynamic exclusion of 45 s. Identification and quantitation of phosphopeptides were done after formatting the mass spectrometry data. LC-MS/MS data acquired using AnalystQS 1.1 (MDS Sciex) were searched using Mascot v2.2.0 (Matrixscience). Reverse database RefSeq26r was searched to find false discovery rate, which was $<1\%$ in all experiments. While searching the database, carbamidomethylation of cysteine was set as a fixed modification and phosphorylation of tyrosine, serine, and threonine; $^{13}C_6$-arginine, lysine 2H4, $^{13}C_6{}^{15}N_4$-arginine, and $^{13}C_6{}^{15}N_2$-lysin; and oxidation of methionine were allowed as variable modifications. The mass tolerance was set to 0.3 atomic mass units for precursor and to 0.4 atomic mass units for fragmented ions. Relative quantitation of phosphotyrosine peptides and peptides derived from phosphotyrosine IP of proteins was performed using MSQuant downloaded from http://msquant.sourceforge.net, which computes protein quantitation ratio using averaged ratios from peptides eluted at different chromatographic points (2). Mascot search results were parsed with LC-MS/MS instrument data file using MSQuant. The quantitation data were verified by manual inspection of heavy, medium, and light peptide derived MS and MS/MS spectra, in MSQuant.

Tryptic digest samples from adenocarcinoma cells were analyzed in an Agilent ion-trap mass spectrometer. MS analyses were performed using an LC system (Agilent 110 series) interfaced with an MS system (Agilent Technologies) and using a capillary reversed phase column, integrated with picotipemitter. Criteria for data acquisition and gradient system used for reversed phase LC were explained earlier. Automatic MS2 scan (m/z 50–2,200) was performed with the following setup. Precursor (350–1,200 m/z) threshold value, 10,000; maximum precursor ions chosen, five; exclusion time, 2 min; Smart target value, 200,000 with a maximum accumulation time of 150 ms. Singly charged ions were excluded from the system. Mass spectrometry data were searched against National Center for Biotechnology Information RefSeq database using Spectrum Mill Proteomics Workbench Version A.03.03 (Agilent Technologies). The following variable modifications were allowed: phosphorylation of serine, threonine, and tyrosine residues; and oxidation of methionine, arginine $^{13}C_6$, and lysine $^{13}C_6$. Peptides were identified with the score cutoff of 8.0, and relative quantitation was carried out using Spectrum Mill. The complete list of peptides identified in the lung adenocarcinoma experiment is shown in Table S5.

Functional Analysis. We performed functional annotation using the 2007 version of DAVID, the Database for Annotation, Visualization and Integrated Discovery, available on the Web at http://david.abcc.ncifcrf.gov. We first uploaded a list of 426 Entrez Gene IDs corresponding to the proteins for which we had

results, selected GOTERM BP ALL only, and performed DAVID Functional Annotation Clustering. We downloaded analysis results and filtered for clusters that contained one or more GO category overlap, with a P value ≤ 0.05. Clusters were named and ranked according to the GO category, with the smallest P value. Cluster results appear in Fig. S2.

Bayesian Network Analysis. Data from our study and the eight published ERBB signaling-related phospho-proteomic studies were used for network modeling (3–10). We first converted the protein identifiers to their equivalent HUGO gene symbols. We identified 20 EGFR signaling-related tyrosine phosphorylation ratios from these 9 proteomic studies, which we called "observations." The phosphorylation levels for each protein were categorized to three states for each observation: up-regulation (ratio >1.5), down-regulation (ratio <0.66), and no change (ratio between 0.66 and 1.5). For observations with multiple measurements, such as time-course or duplicate measurements, we used the following categorization rule. The protein is up-regulated if any of the measurements are up-regulated, the protein is down-regulated if any of the measurements are down regulated, otherwise the protein has no change in regulation. There were 50 proteins that were identified in 50% or more of the observations. Their discrete phosphorylation states were clustered with hierarchical Euclidean clustering using Cluster 3.0 (11). The heatmap of this clustered data are shown in Fig. S8*A*. There were 18 proteins identified in 70% or more of the observations. We estimated missing data in the studies, where any of the 18 proteins were not identified, with a "nearest neighbors" algorithm using a discrete distance metric, defined as the total number of differences in the phosphorylation states, and represented the phosphorylation level in a heatmap (Fig. 5*A*). Bayesian networks (BN) were generated from this data using the BN structure learning software Banjo 2.0.0, with a simulated annealing edge searcher that considered one billion potential networks (12). EGFR and ERBB2 were constrained to only have outgoing edges, except between each other, since it is known *a priori* that these proteins are upstream of all of the others. The top scoring network is the one that has the highest probability generating the observed data and is shown in Fig. 5*B*. The edges with positive influence scores returned by Banjo are black and represent direct correlation (12).

1. Sato M, *et al.* (2006) Multiple oncogenic changes (K-RAS(V12), p53 knockdown, mutant EGFRs, p16 bypass, telomerase) are not sufficient to confer a full malignant phenotype on human bronchial epithelial cells. *Cancer Res* 66:2116–2128.
2. Schulze WX, Mann M (2004) A novel proteomic screen for peptide-protein interactions. *J Biol Chem* 279:10756–10764.
3. Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* 22:1139–1145.
4. Bose R, *et al.* (2006) Phosphoproteomic analysis of Her2/neu signaling and inhibition. *Proc Natl Acad Sci USA* 103:9773–9778.
5. Huang PH, *et al.* (2007) Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc Natl Acad Sci USA* 104:12867–12872.
6. Kratchmarova I, Blagoev B, Haack-Sorensen M, Kassem M, Mann M (2005) Mechanism of divergent growth factor effects in mesenchymal stem cell differentiation. *Science* 308:1472–1477.
7. Olsen JV, *et al.* (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127:635–648.
8. Wolf-Yadlin A, *et al.* (2006) Effects of HER2 overexpression on cell signaling networks governing proliferation and migration. *Mol Syst Biol* 2:54.
9. Zhang Y, *et al.* (2005) Time-resolved Mass Spectrometry of Tyrosine Phosphorylation Sites in the Epidermal Growth Factor Receptor Signaling Network Reveals Dynamic Modules. *Mol Cell Proteomics* 4:1240–1250.
10. Guo A, *et al.* (2008) Signaling networks assembled by oncogenic EGFR and c-Met. *Proc Natl Acad Sci USA* 105:692–697.
11. de Hoon MJ, Imoto S, Nolan J, Miyano S (2004) Open source clustering software. *Bioinformatics* 20:1453–1454.
12. Yu J, Smith VA, Wang PP, Hartemink AJ, Jarvis ED (2004) Advances to Bayesian network inference for generating causal networks from observational biological data. *Bioinformatics* 20:3594–3603.

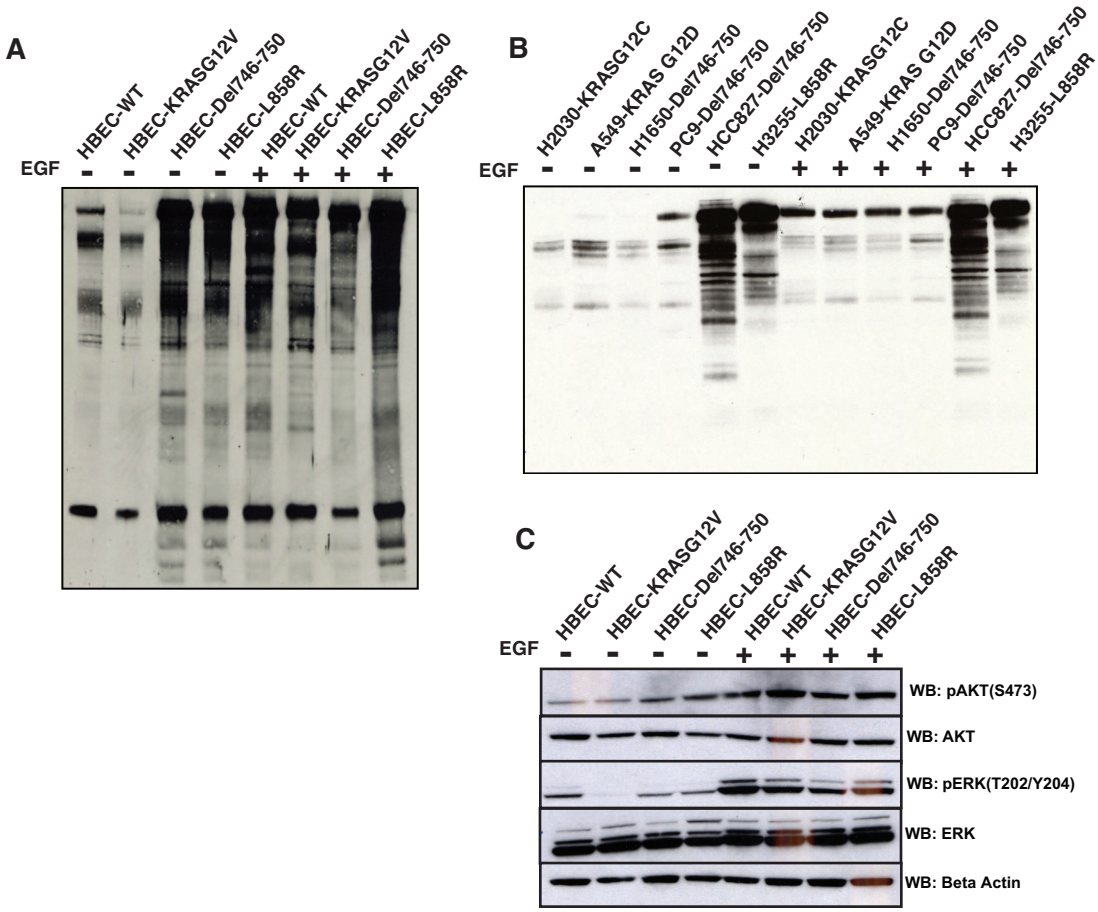

Fig. S1. Increased constitutive anti-phosphotyrosine immunoreactivity in HBECs expressing WT EGFR, KRAS G12V, L858R EGFR, and Del E746-A750 EGFR (*A*) or lung adenocarcinoma cells expressing the mutant EGFRs as indicated (*B*). Immunoprecipitation was done with anti-phosphotyrosine antibody (4G10) on lysates from serum starved or EGF stimulated (3 min) cells, and Western blot was done with the same antiphosphotyrosine antibody. (*C*) Western blot of lysates from serum-starved or EGF-stimulated (for 3 min) HBECs show that both pAKT (S473) and pERK (T202/Y204) immunoreactivity increase upon EGF stimulation in these cells, suggesting canonical EGFR signaling is preserved in these cells.

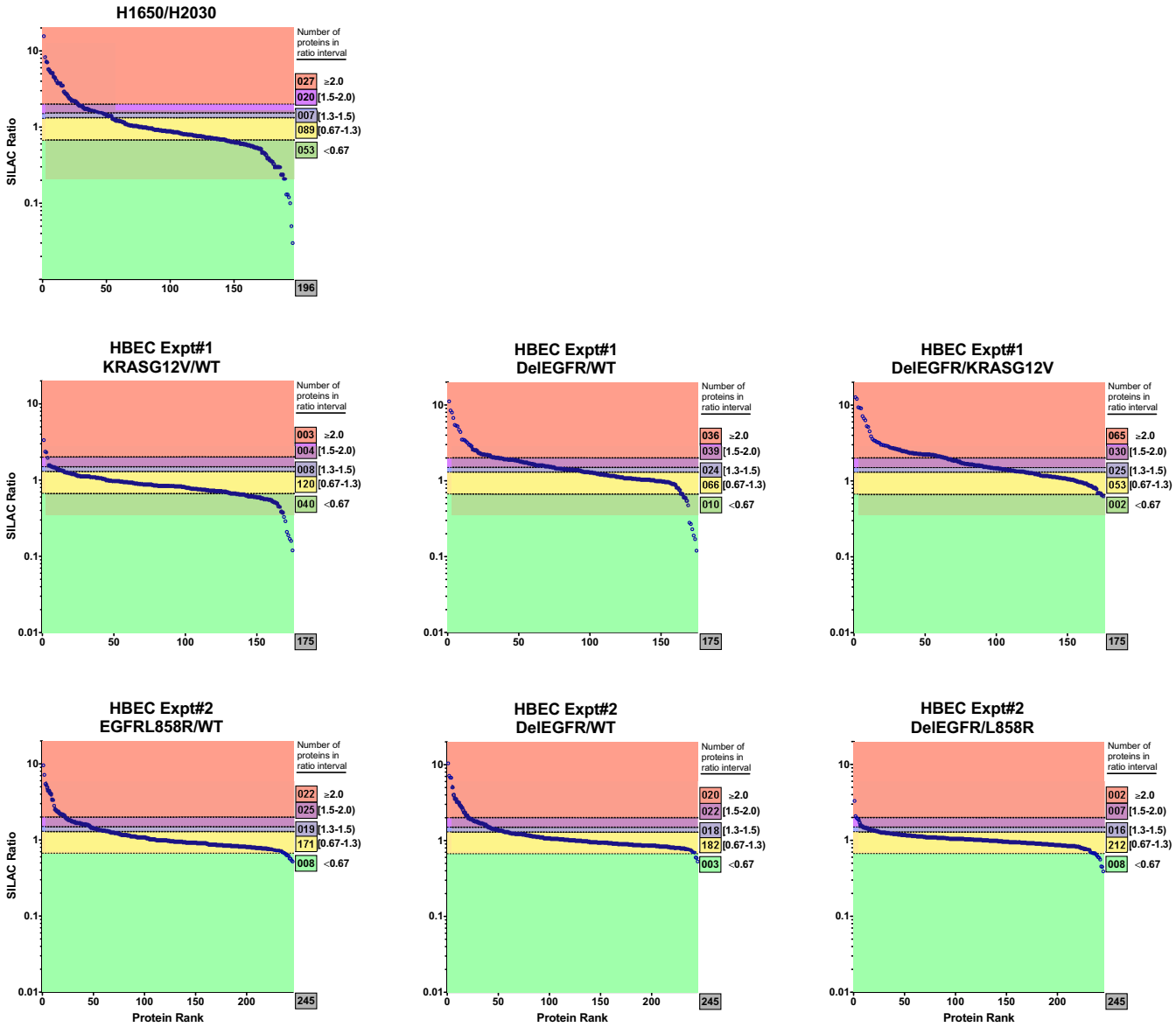

Fig. S2. Tyrosine phosphorylation levels of individual proteins as quantified by SILAC. SILAC ratios are shown for each protein (blue circles) in a separate plot for each pairwise sample comparison. Ratios for each comparison were ranked in decreasing order and then plotted on a log scale. Subject and reference samples are indicated in the plot title, in the numerator and denominator positions, respectively. The colored backgrounds indicate the given ratio ranges. The numbers of proteins within these ratio intervals are indicated to the right of each figure, with the total number of proteins boxed to the lower right. Proteins with phosphorylation ratios $> = 1.5$ (violet and orange) are typically considered increased in the subject sample relative to the reference sample, and proteins with ratios <0.67 (green) are typically considered decreased.

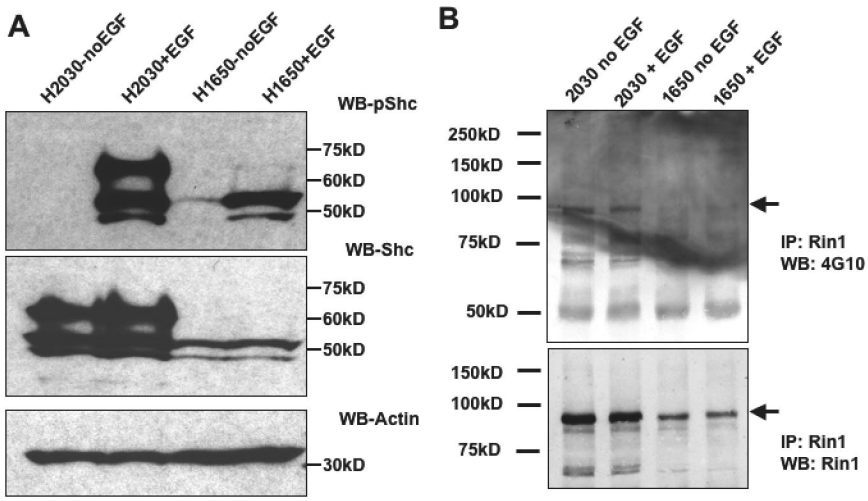

Fig. S3. Decreased protein expression of p66SHC and RIN1 in H1650 compared with H2030 and hence decreased overall phosphorylation of each protein. (*A*) Western blot of lysates for pSHC (Y239/Y240) and total SHC. (*B*) Immunoprecipitation of RIN1 followed by Western blot analysis with 4G10 (pY) and RIN1.

INSR	NP_000199	1029	LGQGSFGM-VYEGNARDIIK- 1047
MET	NP_000236	1084	IGRGHFGC-VYHGTLLDN--- 1100
IGF1R	NP_000866	1005	LGQGSFGM-VYEGVAKGVVK-1023
ERBB2	NP_001005862	696	LGSGAFGT-VYKGIWIP---- 711
ERBB4	NP_001036064	724	LGSGAFGT-VYKGIWVP---- 739
INSR	NP_001073285	1017	LGQGSFGM-VYEGNARDIIK- 1035
ABL2	NP_001093578	279	LGGGQYGE-VYVGVW------ 292
MST1R	NP_002438	1088	IGKGHFGV-VYHGEYIDQ--- 1104
ERBB2	NP_004439	726	LGSGAFGT-VYKGIWIP---- 741
ABL2	NP_005149	258	LGGGQYGE-VYVGVW------ 271
EGFR	NP_005219	718	LGSGAFGT-VYKGLWIP---- 733
ERBB4	NP_005226	724	LGSGAFGT-VYKGIWVP---- 739
ABL2	NP_009298	294	LGGGQYGE-VYVGVW------ 307
INSRR	NP_055030	985	LGQGSFGM-VYEGLARGLEA-1003

Fig. S4. Alignment of Y727 of EGFR with closely related RTKs shows that Y727 of EGFR is conserved across the various RTKs. Multiple alignment was performed using the ClustalW tool and National Center for Biotechnology Information protein reference sequences (RefSeq) for selected, closely-related RTKs. Gene symbols and RefSeq accessions for these proteins are indicated in the figure.

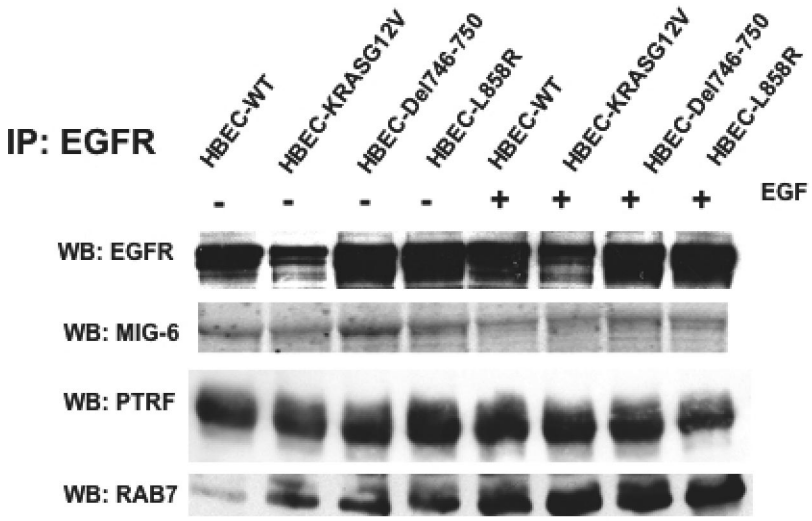

Fig. S5. EGFR interacts with Mig-6, PTRF, and RAB7 in wild type EGFR-, mutant EGFR-, and mutant RAS-expressing HBECs. Immunoprecipitation was done by EGFR monoclonal antibody 528, and Western blot analysis was done with MIG6-, PTRF-, and RAB7-specific antibodies.

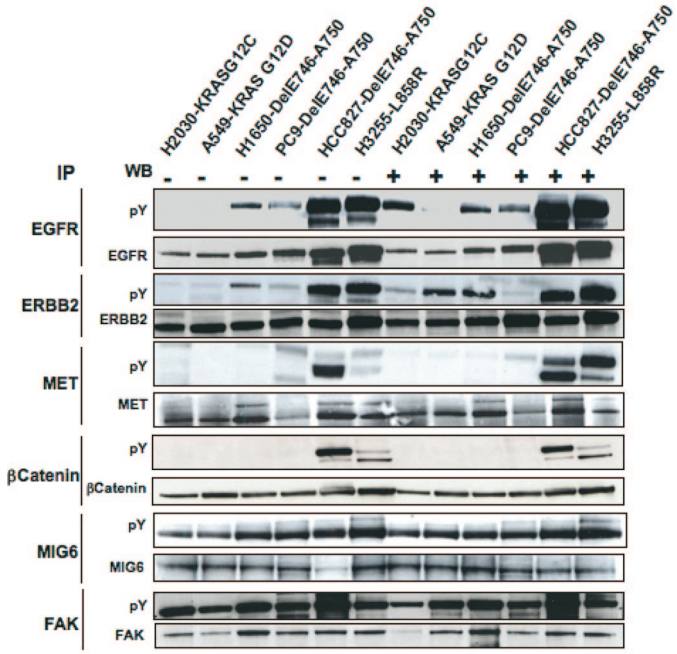

Fig. S6. Validation of mass spectrometry-based quantitation by immunoprecipitation and Western blots of representative proteins. Immunoprecipitation was done with antibodies to the indicated proteins from lysates of a panel of lung adenocarcinoma cell lines with mutations in either *EGFR* or *KRAS* genes, and Western blots done with anti-phosphotyrosine and protein-specific antibodies.

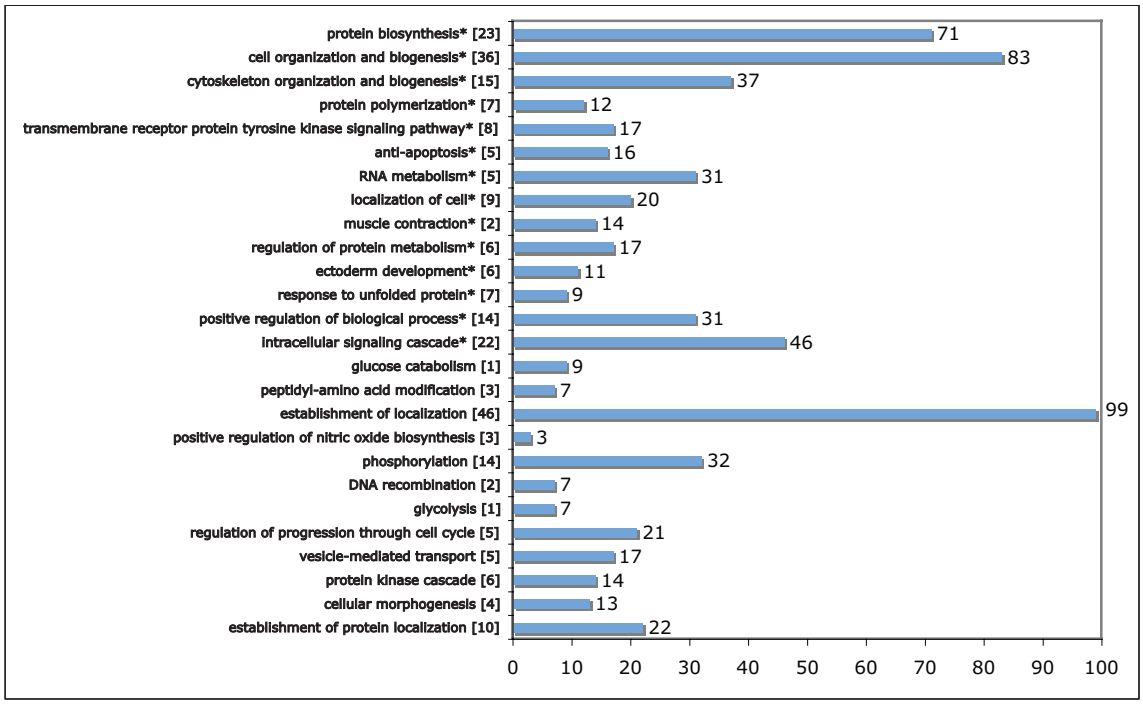

Fig. S7. Functional classification of proteins identified in at least one of the three ''large-scale'' phosphotyrosine immunoprecipitation experiments at the whole protein-level, using filtered results from the 2007 version of DAVID (*D*atabase for *A*nnotation, *V*isualization and *I*ntegrated *D*iscovery). The categories are listed in order of significance from top-down. Clusters with asterisks are also significant ($P \leq 0.05$) by the Benjamini multiple testing correction test. The number in brackets is the number of genes with a phosphotyrosine IP ratio $\geq$1.5. The length of the bar (and the number to the right of the bar) is the number of overlapping genes in the most significant cluster category.

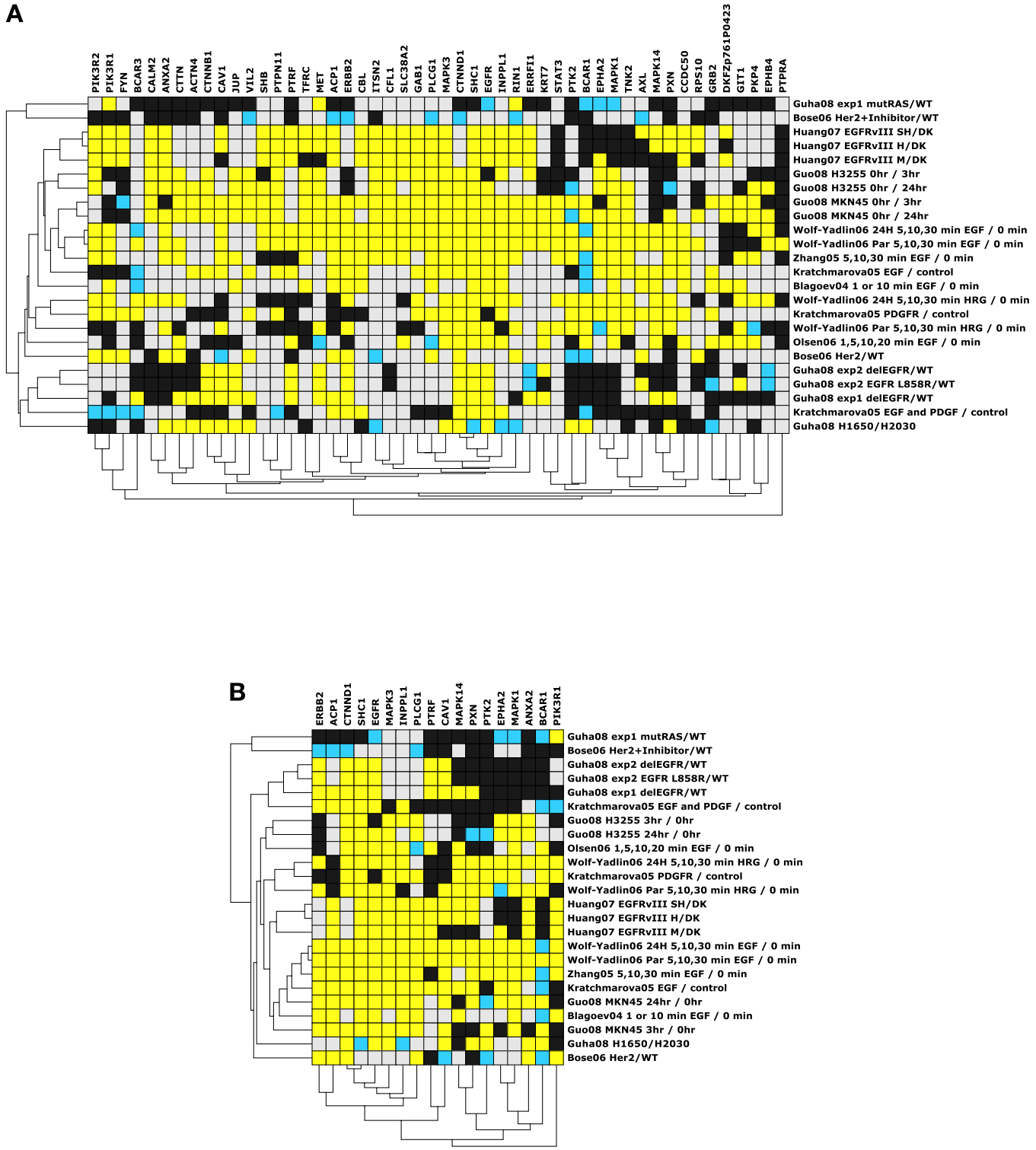

Fig. S8. (*A*) Heatmap of the discrete phosphorylation levels of 50 proteins that show up in 50% or more of the 24 observations from our study and the eight other published proteomic studies. The *y*-axis lists the 24 observations and their respective parent studies. The *x*-axis lists the proteins. The bottom *x*-axis and left *y*-axis show hierarchical clustering based on the discrete distance between the observations and between the proteins. Yellow is up-regulation, blue is down regulation, black is no regulation change, and gray represents no data. (*B*) Heatmap for the phosphorylation level of 18 proteins identified in 70% or more of the datasets.

Table S1. Summary of tyrosine phosphorylated peptides and the site of tyrosine phosphorylation identified by mass spectrometry from pY-IPs of peptides from lysates of HBECs expressing WTEGFR, L858R EGFR and DelEGFR

SILAC ratio	HBEC-Expt#1 DelEGFR /KRASG12V	HBEC-Expt#1 DelEGFR /WTEGFR	HBEC-Expt#1 KRASG12V /WTEGFR	HBEC-Expt#2 L858R /WTEGFR	HBEC-Expt#2 DelEGFR /L858R	H1650/H2030 DelEGFR /KRASG12C
>1.5	95	75	7	45	9	47
0.66–1.5	78	90	128	190	228	96
<0.66	2	10	40	8	8	53
Total	175	175	175	245	245	196

The ratio is an average of three experiments involving three distinct pY-IPs and mass spectrometry analyses from same lysates. The standard deviation obtained from the ratios obtained in each of the three separate analysis is also shown.

Other Supporting Information Files

Table S2
Table S3
Table S4
Table S5
Table S6