Science Signaling
AAAS

www.sciencesignaling.org/cgi/content/full/9/436/rs6/DC1

Supplementary Materials for

Conservation of protein abundance patterns reveals the regulatory architecture of the EGFR-MAPK pathway

Tujin Shi, Mario Niepel, Jason E. McDermott, Yuqian Gao, Carrie D. Nicora, William B. Chrisler, Lye M. Markillie, Vladislav A. Petyuk, Richard D. Smith, Karin D. Rodland, Peter K. Sorger, Wei-Jun Qian, H. Steven Wiley*

*Corresponding author. Email: steven.wiley@pnnl.gov

Published 12 July 2016, *Sci. Signal.* **9**, rs6 (2016)
DOI: 10.1126/scisignal.aaf0891

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencesignaling.org/cgi/content/full/9/436/rs6/DC1)

Table S4 (Microsoft Excel format). Transcriptomics analysis of EGFR-MAPK pathway components across seven cell lines.
Table S5 (Microsoft Excel format). SRM analysis of all of the EGFR-MAPK pathway proteins.
Table S6 (Microsoft Excel format). Variance analysis of signaling proteins in the EGFR-MAPK pathway.
Table S7 (Microsoft Excel format). Copy number gain or loss frequency for genes in the EGFR pathway.

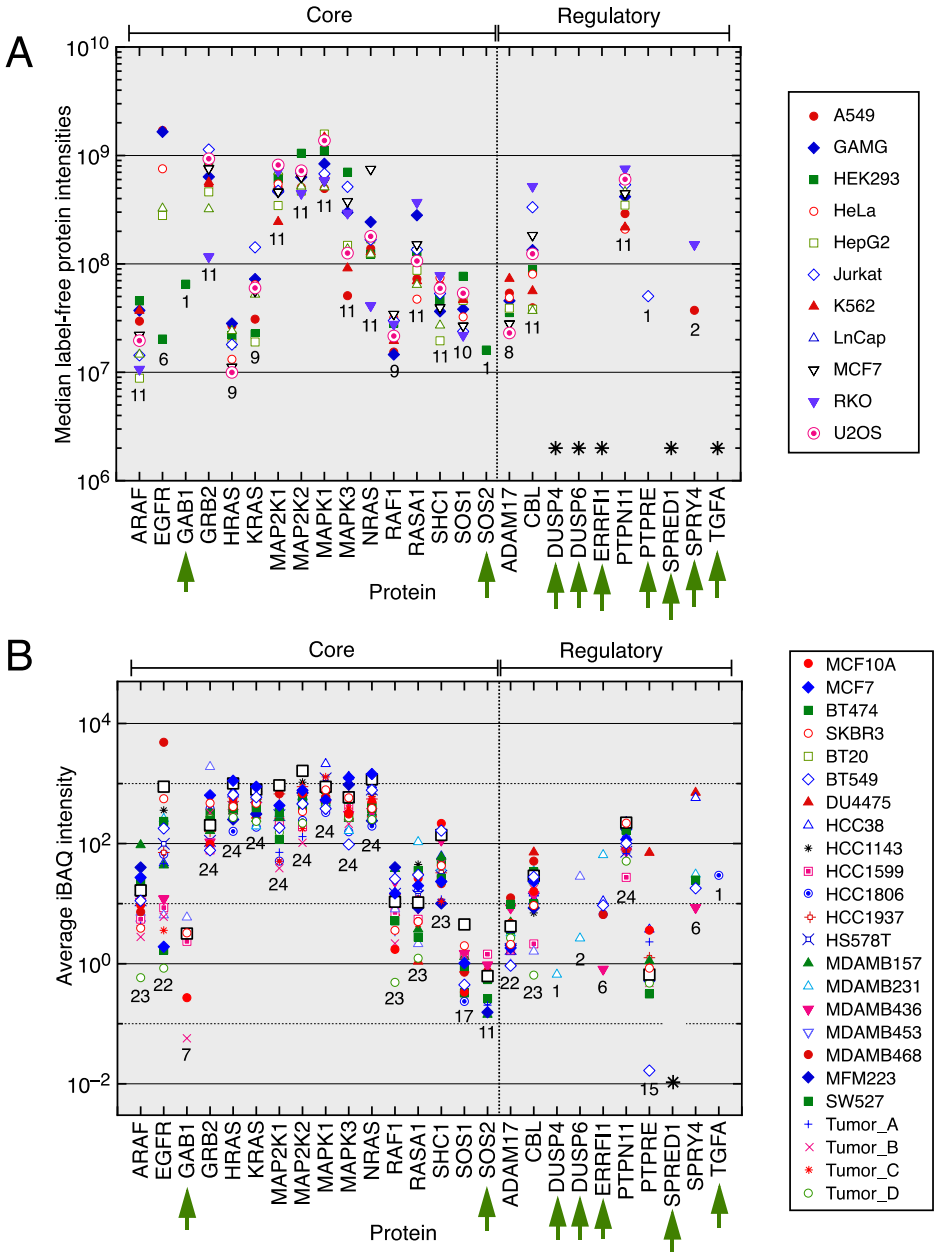

Figure S1: Relative abundance of proteins of the EGFR-MAPK signaling pathway as assessed by deep proteomics surveys. (A) Reported abundances of proteins in the EGFR-MAPK pathway from the study of Geiger et al. (*46*). Proteins are grouped into either core components or feedback regulators as described in the text. Within groups, proteins are listed alphabetically. The LFQ intensity values for both core and regulated components of the EGFR-MAPK pathway are shown for the indicated cell types. Numbers below each protein group indicates the number of cell types in which that protein was detected. Green arrows indicate proteins that were observed in less than half of the surveyed cell types. Stars (*) indicate proteins that were not detected or reported. **(B)** Same as in panel A except the study was that of Lawrence *et al.* (2014) (*14*). In this survey, the iBAQ label-free method was used for protein quantification. This survey also included samples from 4 different tumors.

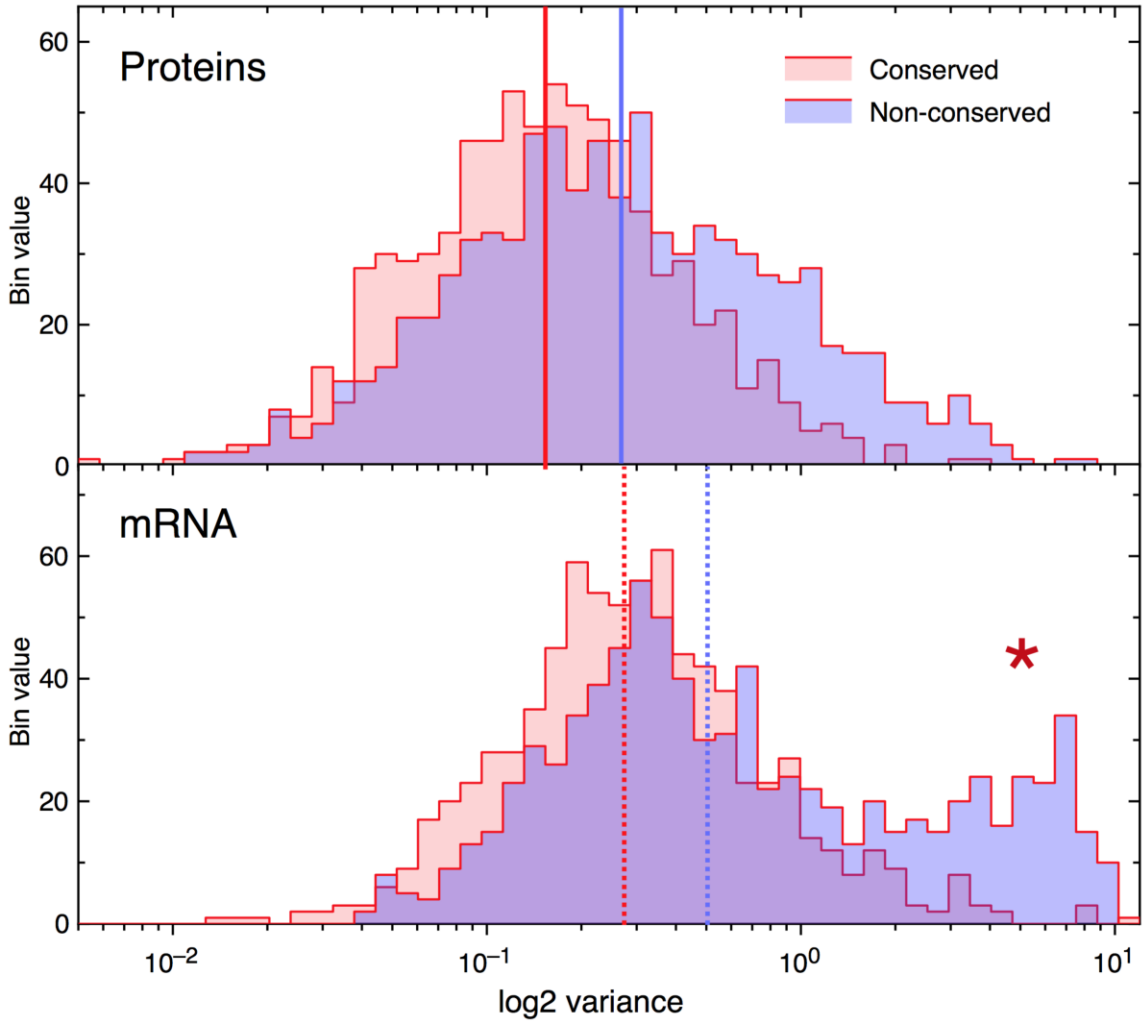

Figure S2: Comparison of $\log_2$ variance distribution of conserved versus nonconserved proteins and mRNA. Proteins and mRNAs measured in this study across seven cell lines corresponding to those showing conserved abundance values across species (Khan et al, (*42*) are shown in pink. Proteins and corresponding mRNAs that were unique to our data set is shown in blue. Calculated variances for each protein and mRNA were binned into 50 groups and displayed as the frequency distribution of the mean $\log_2$ variance values of each bin. There were 1847 protein/mRNA pairs for the conserved group and 834 for the non-conserved group. Vertical lines indicate the median values of each group. The starred (*) mRNA peak corresponds to highly variable mRNA species.

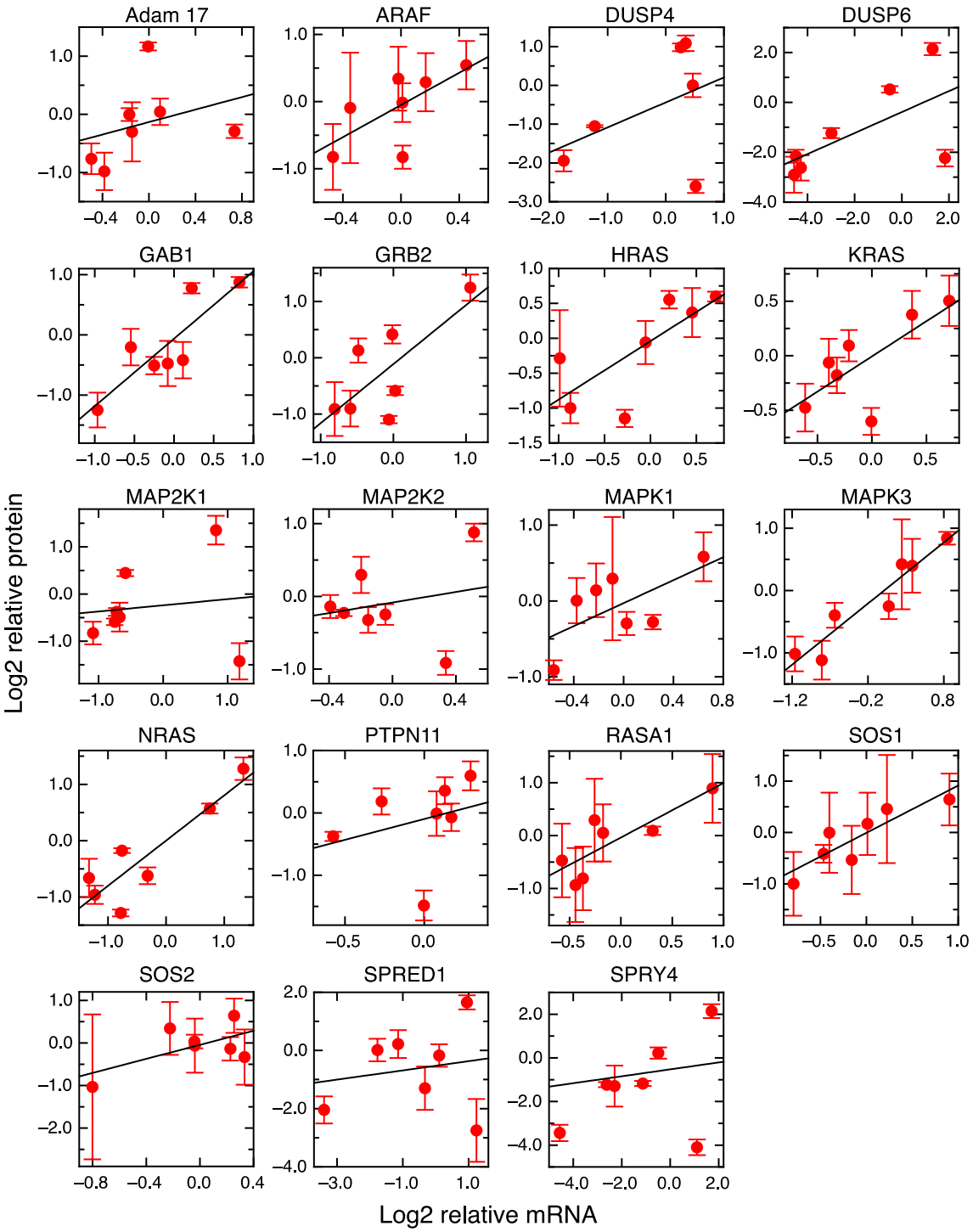

Figure S3: Relationship between mRNA expression and protein abundances of EGFR-MAPK pathway components across all cell lines. The $\log_2$ value of the mRNA of each cell line divided by the average of all lines was plotted against the comparable protein value. Protein and RNA-Seq measurements were made on duplicate samples from the same experiment. Error bars representing the average coefficient of variation from N=4 samples separately processed and analyzed. The lines were calculated by linear regression against the values.

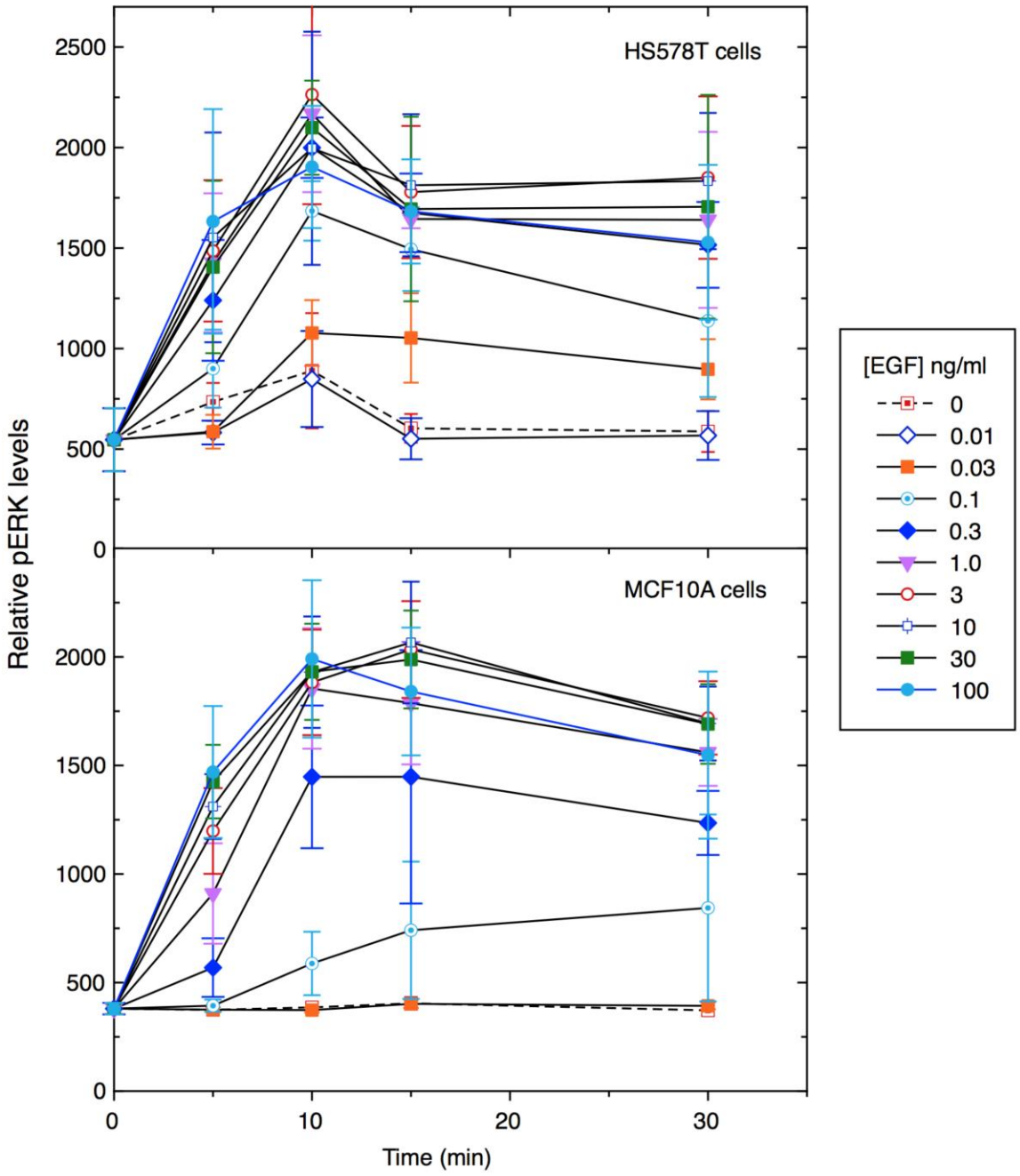

Figure S4: Kinetic response of cells to EGF as a function of ligand dose. HS578T cells (top) and MCF10A cells (bottom) were treated with the indicated concentration of EGF for 5-30 min, fixed, and evaluated for phosphorylated (pERK) abundance by immunofluorescence as described in the Materials and Methods. Data are mean ± S.D. of triplicate technical measurements.

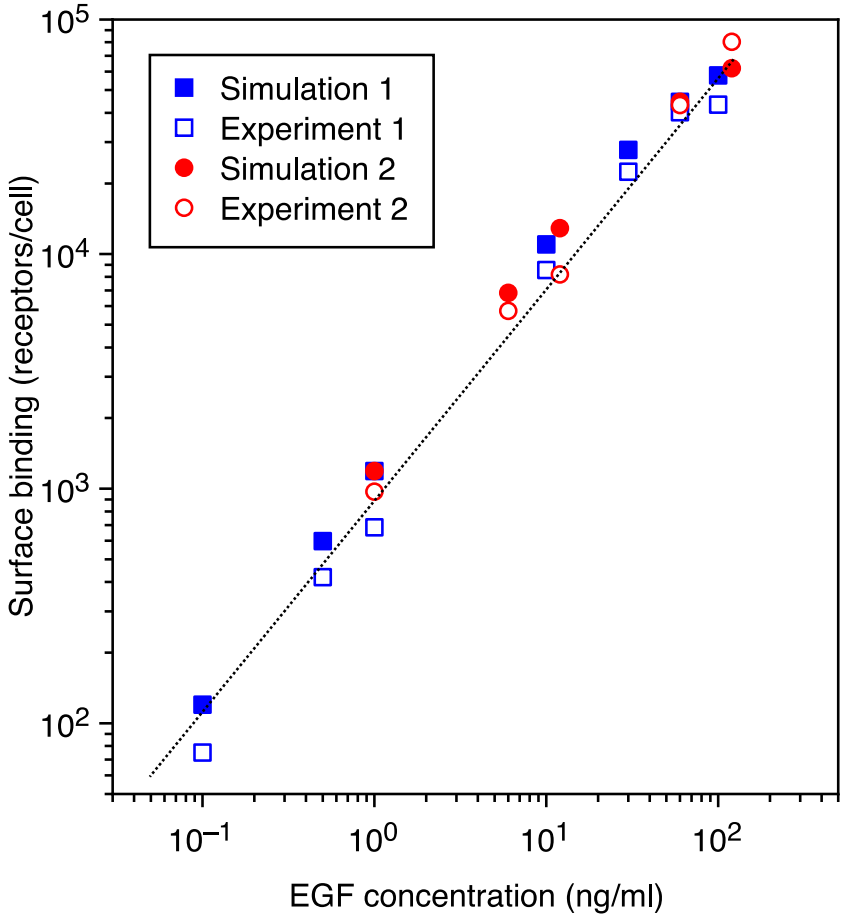

Figure S5: Comparison between simulated and measured EGF binding to cells. Mammary epithelial cells (HMEC 184A1) were incubated with concentrations of ^{125}I-EGF ranging from 1-120 ng/ml (circles) or 0.1-100 ng/ml (squares) for 5 min. The total amount of ligand associated with the cells was quantified and converted to receptor/cell based on specific activity of the ligand and measured cell number (open symbols). Simulation of EGF binding was accomplished using the equations and parameters described in Materials and Methods (solid symbols). Shown are the results of two independent experiments. Some symbols appear missing because of overlap of the symbols.

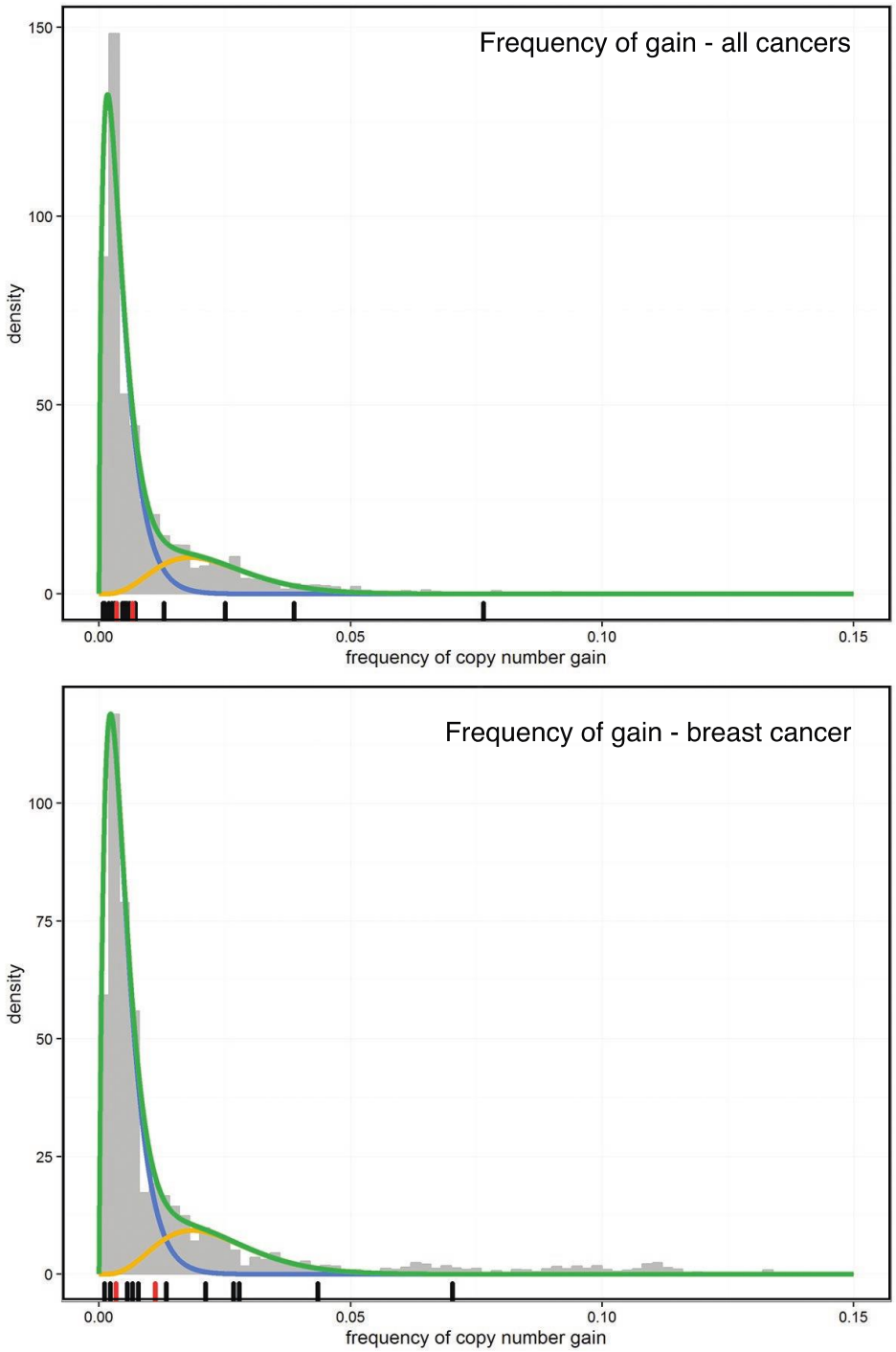

Figure S6. Two-component mixture model of the copy number gain frequencies for breast and all cancers. Blue line is model bulk of the genes (~80%) with the yellow line representing the higher frequency component. For all cancers, N=6293 samples. For breast cancer, N=899 samples. The genes corresponding to the EGFR/ERK pathway shown as a rug below the histogram. The amplification frequency of SOS1 and SOS2 genes (highlighted with red) is insignificant (p-values are 0.58 and 0.057, respectively).

Table S1: EGFR-MAPK pathway–associated genes regulated by pathway perturbation. The table (Excel file) shows fold change in expression, Tyr and Ser/Thr phosphorylation of the "core" and feedback components of the EGFR-MAPK pathway in response to EGF or, where indicated, monoclonal antibody 225mAb (225) or the MEK inhibitor U0126.

Table S2: RNA expression and protein abundance of HMECs ranked by RNA-Seq reads. The table (Excel file) shows the total number of RNA-Seq reads uniquely mapping to the indicated genes as well as peptide and total spectra counts of the corresponding proteins in HMECs as detected by shotgun proteomics. Only genes displaying at least 5 unique reads are listed.

Table S3: iTRAQ and transcriptome analysis of proteins with conserved abundances. The table (Excel file) shows a comparison of mRNA and protein variance in a cell line panel with conserved proteins described by Khan *et al.* (44) in table S4 that corresponded to "models 1 and 4", in which the protein abundance pattern was consistent with directional selection along human lineage and patterns with no significant difference between mean abundance. Ratios are the $\log_2$ ratios of the abundance of each cell type divided by the average of all cell types. Slope, intercept, F statistic and R^2 values are from linear regression of relative mRNA to relative protein levels. Red-labeled columns designate data taken from Khan *et al*. (*48*).

Table S4: Transcriptomics analysis of EGFR-MAPK pathway components across seven cell lines. The table (Excel file) shows results of the RNA-Seq analysis of genes encoding EGFR-MAPK pathway proteins and regulators across the cell line panel. Expression was normalized to reads per kilobase of gene length per million mRNA reads (RPKM).

Table S5: SRM analysis of all of the EGFR-MAPK pathway proteins. The table (Excel file) shows the SRM data from the panel of cell lines. PRISM_IS is the intensity of the calibration peptide using the PRISM approach, Scheduled_IS is the intensity from the Scheduled SRM approach and IS_usedforcal is the value used in the actual abundance measurement. Coefficient of variation (CV) is calculated from 4 independent runs. Data has also been normalized to copies per cell. Color of the cells indicates relative intracellular abundances.

Table S6: Variance analysis of signaling proteins in the EGFR-MAPK pathway. The table (Excel file) shows the relative $\log_2$ gene expression and protein abundance values for the EGFR-MAPK pathway proteins and well as results from the regression and variance analyses. Also included are the comparable $\log_2$ variance values from the study by Khan *et al. (48)* as well as the iTRAQ and SRM studies described in the main text.

Table S7: Copy number gain or loss frequency for genes in the EGFR pathway. The table (Excel file) shows the analysis of copy number variation (CNV) data from the COSMIC database to determine whether amplification of genes in the EGFR-MAPK pathway occurred at a higher frequency than random in cancers. Frequency is calculated for all cancers (N=6293 samples) and for breast cancer (N=899). The frequency distribution modeled a two-component mixture of beta distributions. Colored cells represent significant difference in frequency from random ($p<0.05$) by expectation maximization algorithm.