Figure S1

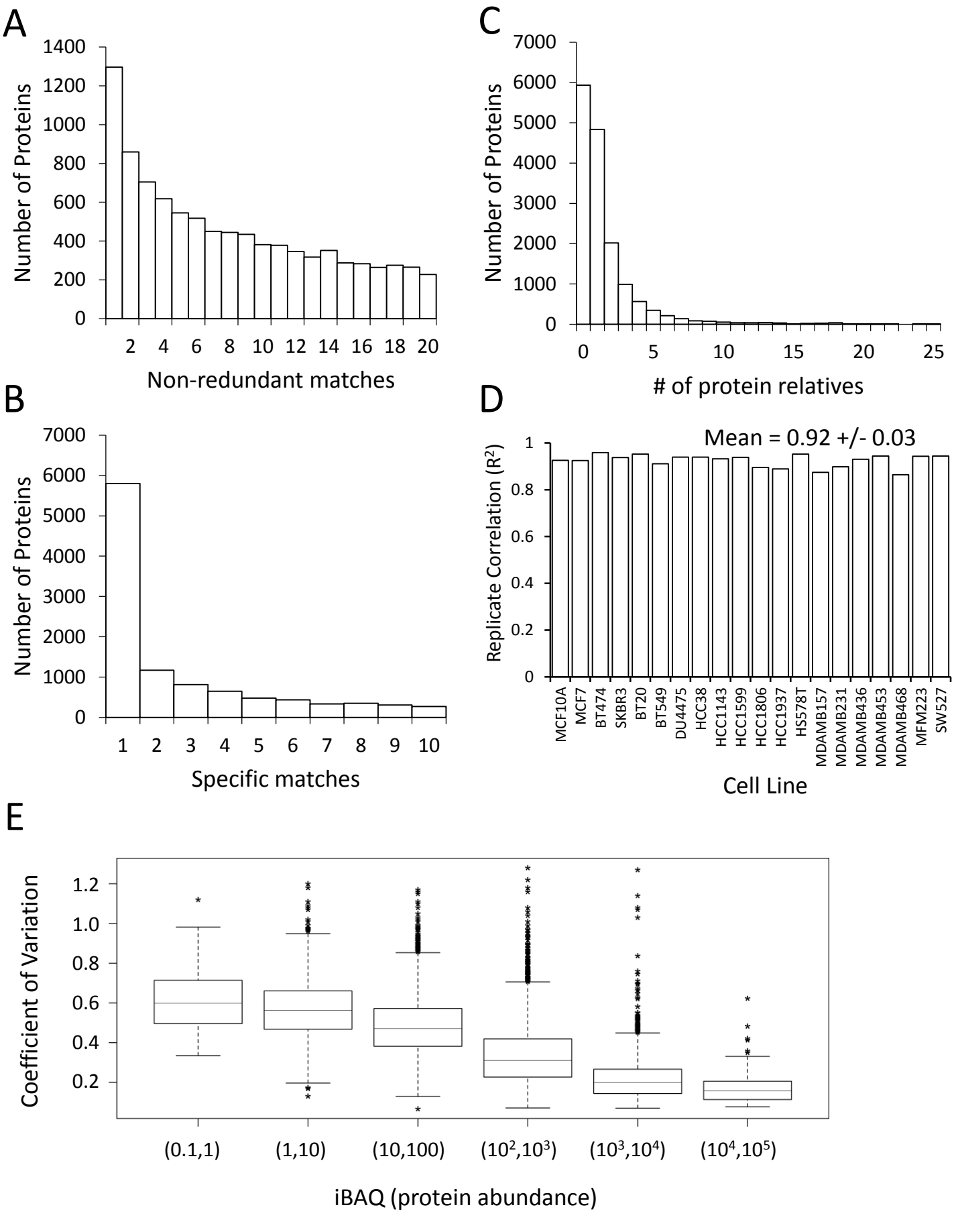

A
Number of Proteins
Non-redundant matches
B
Number of Proteins
Specific matches
C
Number of Proteins
# of protein relatives
D
Mean = 0.92 +/- 0.03
Replicate Correlation (R^2)
MCF10A
MCF7
BT474
SKBR3
BT20
BT549
DU4475
HCC38
HCC1143
HCC1599
HCC1806
HCC1937
HS578T
MDAMB157
MDAMB231
MDAMB436
MDAMB453
MDAMB468
MFM223
SW527
Cell Line
E
Coefficient of Variation
(0.1,1)
(1,10)
(10,100)
(10^2,10^3)
(10^3,10^4)
(10^4,10^5)
iBAQ (protein abundance)

Figure S2

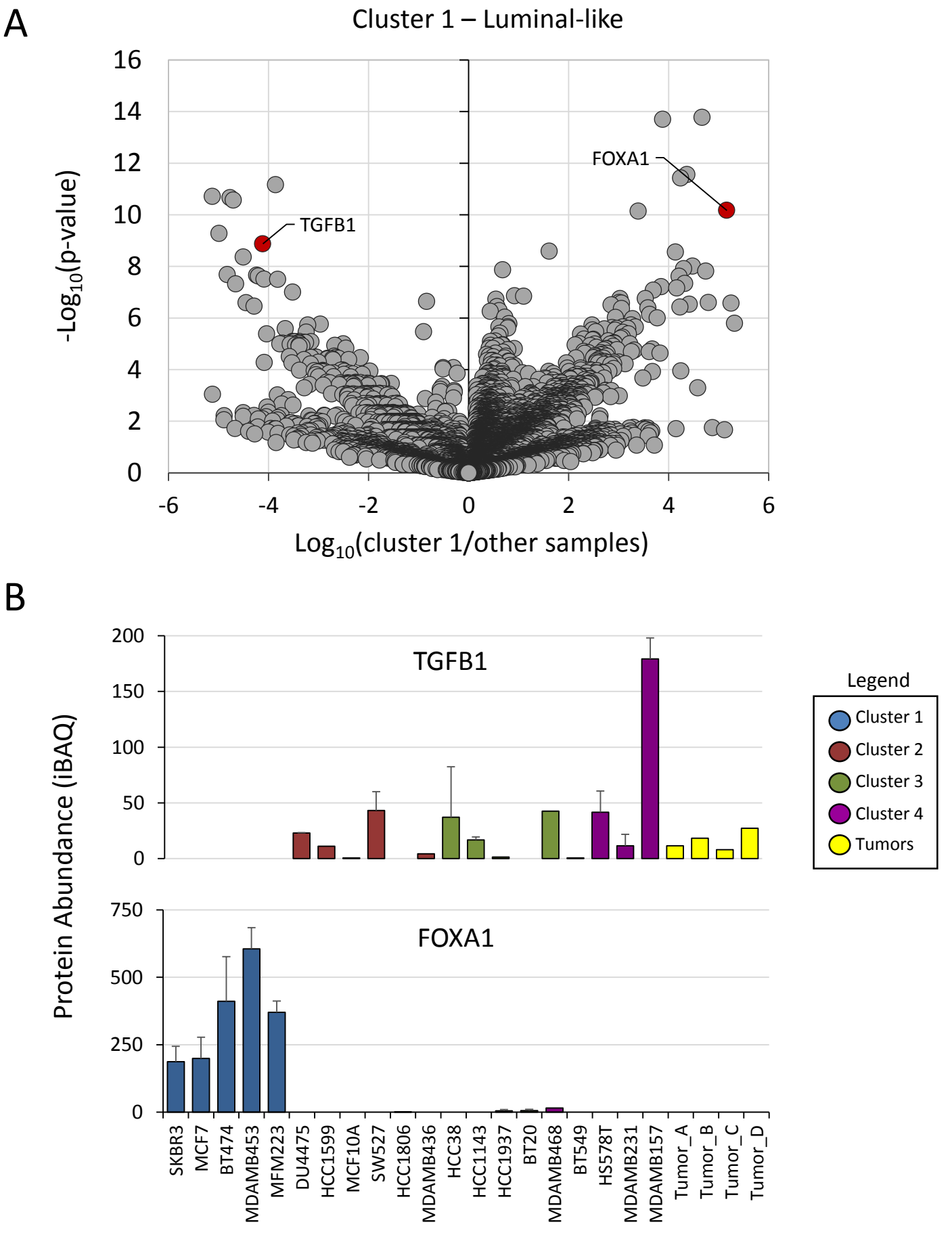
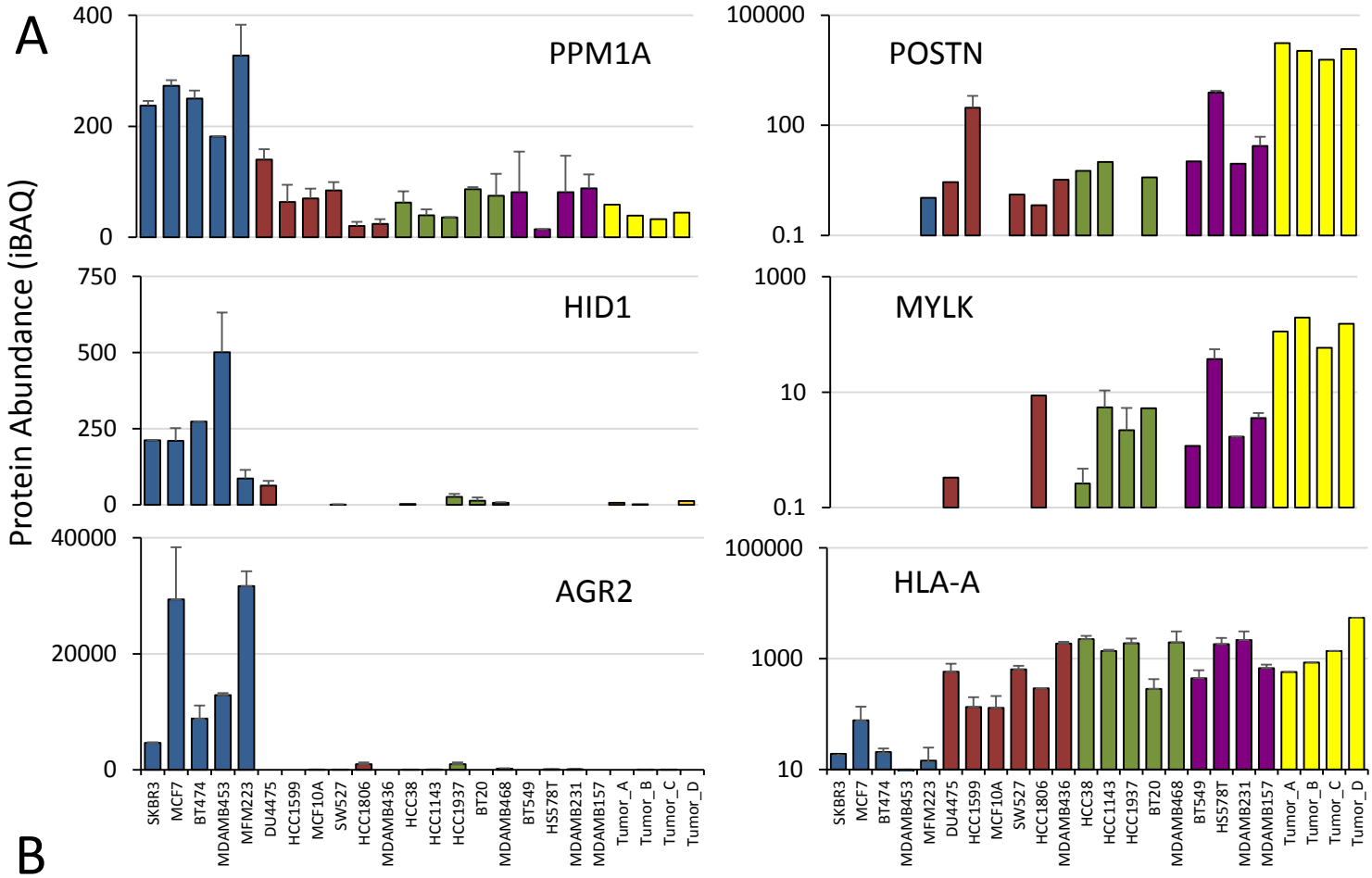

A
Cluster 1 – Luminal-like
-Log10(p-value)
Log10(cluster 1/other samples)
FOXA1
TGFB1
B
Protein Abundance (iBAQ)
TGFB1
FOXA1
Legend
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Tumors
SKBR3
MCF7
BT474
MDAMB453
MFM223
DU4475
HCC1599
MCF10A
SW527
HCC1806
MDAMB436
HCC38
HCC1143
HCC1937
BT20
MDAMB468
BT549
HS578T
MDAMB231
MDAMB157
Tumor_A
Tumor_B
Tumor_C
Tumor_D

Figure S3


A
Protein Abundance (iBAQ)
PPM1A
POSTN
HID1
MYLK
AGR2
HLA-A
SKBR3
MCF7
BT474
MDAMB453
MFM223
DU4475
HCC1599
MCF10A
SW527
HCC1806
MDAMB436
HCC38
HCC1143
HCC1937
BT20
MDAMB468
BT549
HS578T
MDAMB231
MDAMB157
Tumor_A
Tumor_B
Tumor_C
Tumor_D

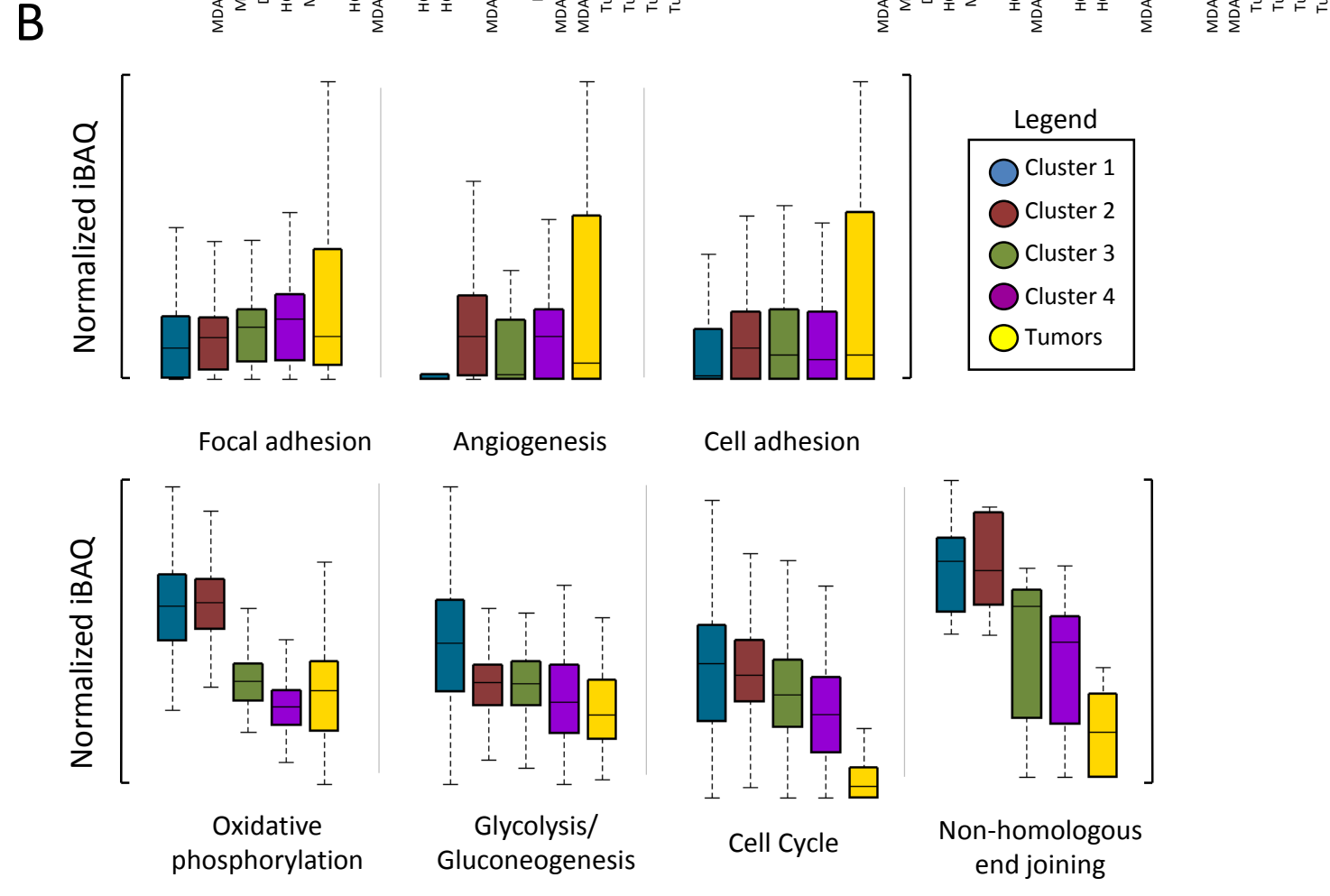

B
Normalized iBAQ
Legend
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Tumors
Focal adhesion
Angiogenesis
Cell adhesion
Normalized iBAQ
Oxidative phosphorylation
Glycolysis/ Gluconeogenesis
Cell Cycle
Non-homologous end joining

Figure S4

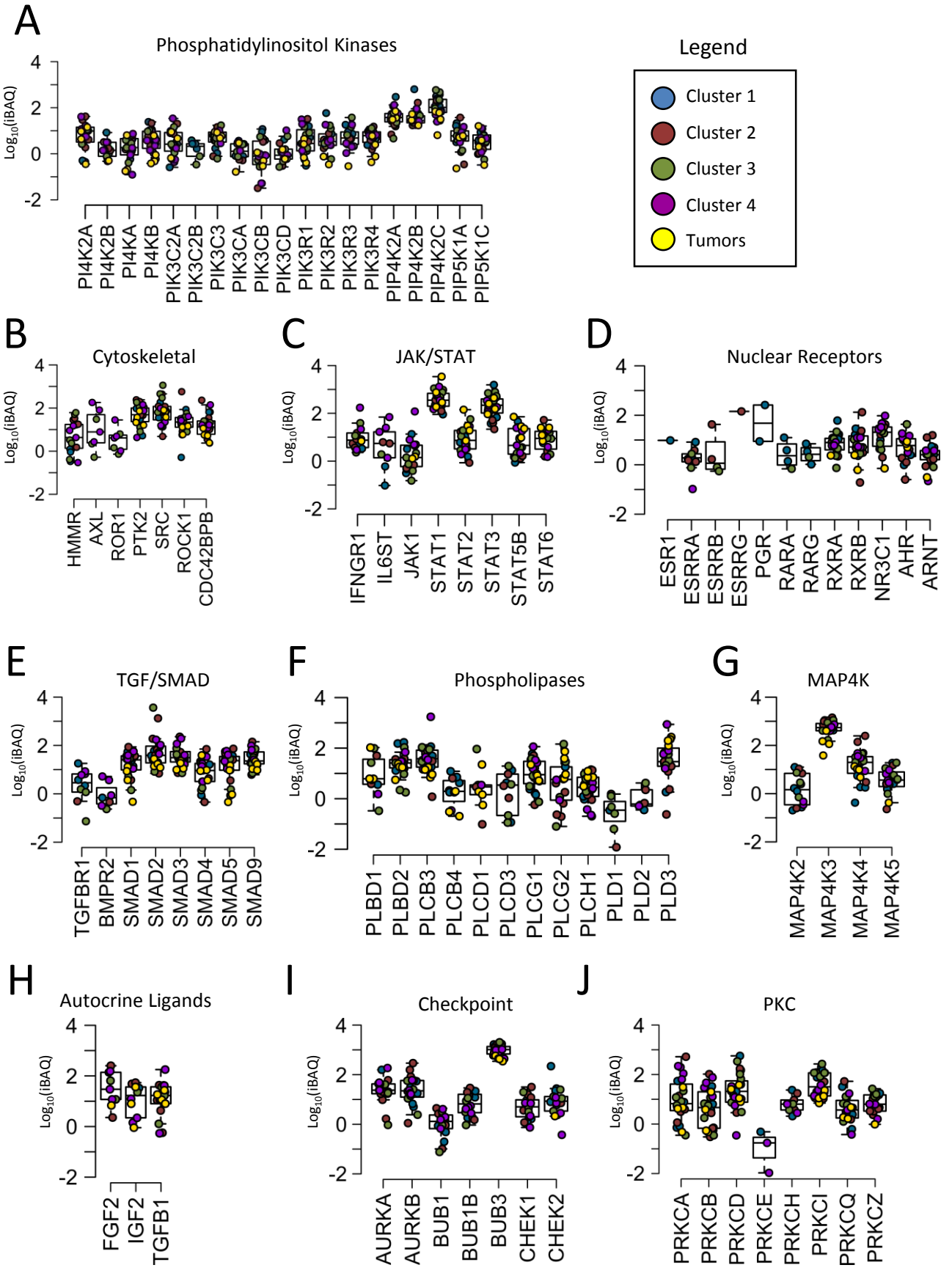

A
Phosphatidylinositol Kinases
Log10(iBAQ)
PI4K2A
PI4K2B
PI4KA
PI4KB
PIK3C2A
PIK3C2B
PIK3C3
PIK3CA
PIK3CB
PIK3CD
PIK3R1
PIK3R2
PIK3R3
PIK3R4
PIP4K2A
PIP4K2B
PIP4K2C
PIP5K1A
PIP5K1C
Legend
Cluster 1
Cluster 2
Cluster 3
Cluster 4
Tumors
B
Cytoskeletal
HMMR
AXL
ROR1
PTK2
SRC
ROCK1
CDC42BPB
C
JAK/STAT
IFNGR1
IL6ST
JAK1
STAT1
STAT2
STAT3
STAT5B
STAT6
D
Nuclear Receptors
ESR1
ESRRA
ESRRB
ESRRG
PGR
RARA
RARG
RXRA
RXRB
NR3C1
AHR
ARNT
E
TGF/SMAD
TGFBR1
BMPR2
SMAD1
SMAD2
SMAD3
SMAD4
SMAD5
SMAD9
F
Phospholipases
PLBD1
PLBD2
PLCB3
PLCB4
PLCD1
PLCD3
PLCG1
PLCG2
PLCH1
PLD1
PLD2
PLD3
G
MAP4K
MAP4K2
MAP4K3
MAP4K4
MAP4K5
H
Autocrine Ligands
FGF2
IGF2
TGFB1
I
Checkpoint
AURKA
AURKB
BUB1
BUB1B
BUB3
CHEK1
CHEK2
J
PKC
PRKCA
PRKCB
PRKCD
PRKCE
PRKCH
PRKCI
PRKCQ
PRKCZ

Figure S5

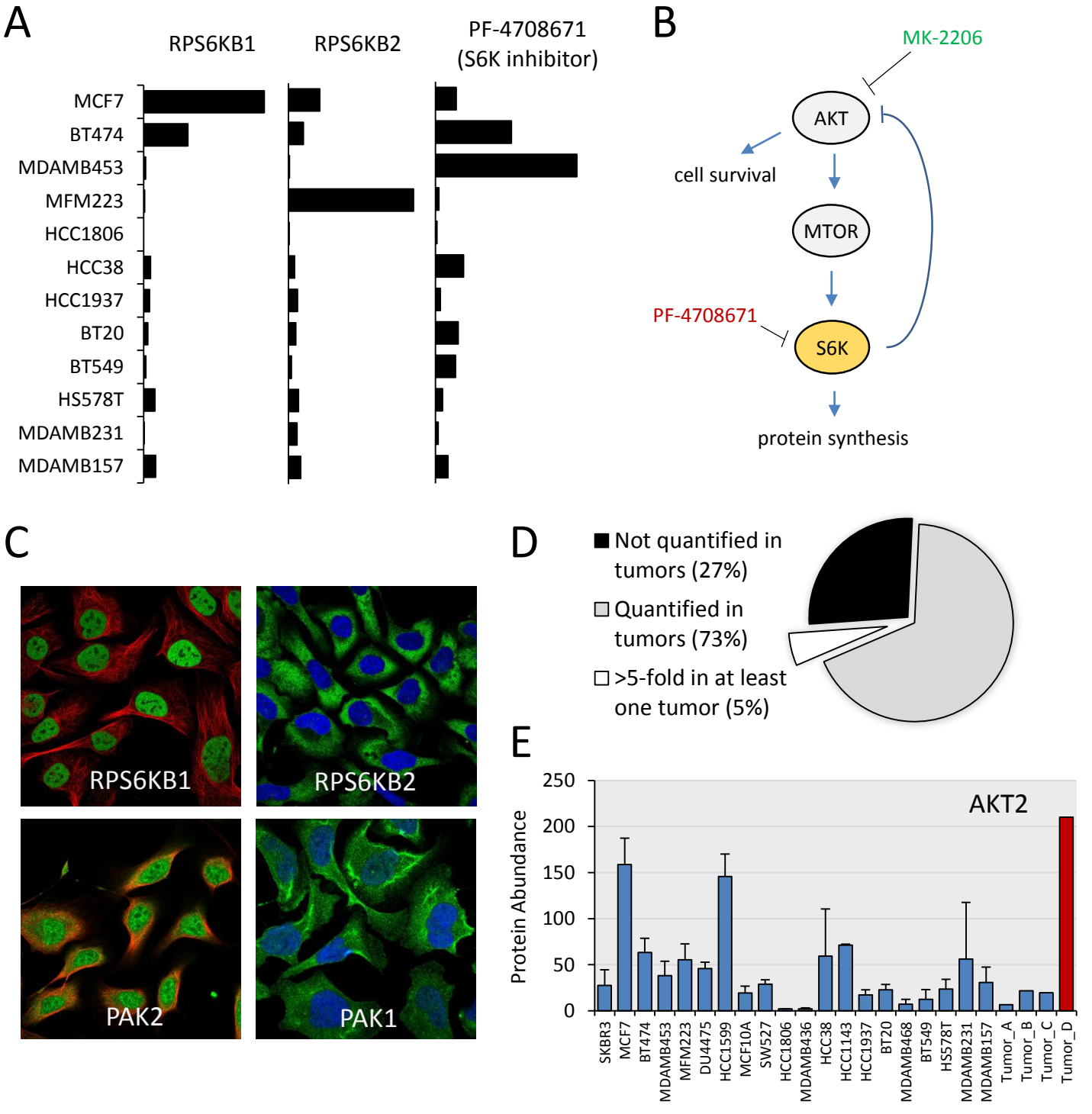

A
RPS6KB1
RPS6KB2
PF-4708671
(S6K inhibitor)
MCF7
BT474
MDAMB453
MFM223
HCC1806
HCC38
HCC1937
BT20
BT549
HS578T
MDAMB231
MDAMB157
B
MK-2206
AKT
cell survival
MTOR
PF-4708671
S6K
protein synthesis
C
RPS6KB1
RPS6KB2
PAK2
PAK1
D
Not quantified in tumors (27%)
Quantified in tumors (73%)
>5-fold in at least one tumor (5%)
E
AKT2
Protein Abundance
250
200
150
100
50
0
SKBR3
MCF7
BT474
MDAMB453
MFM223
DU4475
HCC1599
MCF10A
SW527
HCC1806
MDAMB436
HCC38
HCC1143
HCC1937
BT20
MDAMB468
BT549
HS578T
MDAMB231
MDAMB157
Tumor_A
Tumor_B
Tumor_C
Tumor_D

Figure S1 | Statistics of protein identification and quantification, related to Figure 1
(A) Distribution of the number of non-redundant peptide matches (peptides with a unique amino acid sequence) per protein. (B) Distribution of the number of specific peptide matches (peptides matching to no other protein in the dataset) per protein. (C) Distribution of the number of protein relatives (proteins sharing at least one peptide with one other protein in the dataset) per protein. (D) Correlation coefficient between replicates for each cell line. (E) Coefficient of variation *versus* absolute protein abundance. Protein abundances were binned as indicated in the x-axis.

Figure S2 | Identification of subtype-associated proteins using volcano plots, related to Figure 3
(A) Volcano plot enables selection of proteins significantly underexpressed (top left, e.g. TGFB1) or overexpressed (top right, e.g. FOXA1) in the samples from subtype cluster 1 *versus* other samples. (B) Protein expression of transforming growth factor beta 1 was only expressed in TNBC samples. Expression of Forkhead box A1 (FOXA), also known as hepatocyte nuclear factor 3-alpha, was almost exclusive to samples in cluster 1. Sample labels are shown in the bottom panel. Color corresponds to cluster assignment from Figure 3A. Error bars represent S.D.

Figure S3 | Additional subtype-associated proteins and pathways, related to Figure 3
(A) Proteins associated with cluster 1, representing luminal-like breast cancer *versus* triple-negative breast cancer. Protein phosphatase 1A (PPM1A), down-regulated in multiple cancers 1 (HID1), and anterior gradient protein 2 (AGR2) expression were significantly decreased in triple-negative breast cancer. Periostin (POSTN), myosin light chain kinase (MYLK), and MHC class I antigen A (HLA-A) were increased in TNBC. Error bars represent S.D. (B) Boxplot distribution of protein abundance for each cluster within gene ontology pathways as indicated. Color corresponds to cluster assignment from Figure 3A.

Figure S4 | Absolute abundance distribution of additional signaling sub-networks and components, related to Figure 4
Distribution of absolute abundance across samples for each protein in the indicated signaling networks. Each dot represents a sample, color coded according to cluster assignment from Figure 4A.

Figure S5 | Correlation of signaling components and drug sensitivity, related to Figure 7
(A) S6K expression was inversely correlated with sensitivity to an S6K inhibitor. Left two panels: protein abundance (iBAQ) across cell lines of RPS6KB1 and RPS6KB2. Right panel: drug sensitivity (inverse IC50, M^{-1}) across the same cell lines of S6K inhibitor (PF-4708671). (B) Schematic of negative feedback signaling between S6K and AKT. S6K expression is associated with sensitivity to AKT inhibitor (MK-2206) but inversely associated with S6K inhibitor (PF-4708671). (C) Immunofluorescent images of U2OS cells from the Human Protein Atlas (http://www.proteinatlas.org/) (Uhlen et al., 2010) showing the subcellular localization for RPS6KB1 (nuclear), RPS6KB2 (cytoplasmic), PAK2 (nuclear), and PAK1 (cytoplasmic). Green: protein, Blue: nucleus, Red: microtubules. (D) Percentage of proteins used in the drug sensitivity analysis that were also quantifiable in tumors. (E) Protein kinase AKT2 was most highly expressed in a tumor sample.

Table S1 | TNBC proteome
Intensity-based protein abundance (iBAQ) profile of each sample. Values represent the summed peptide intensity normalized by total proteome intensity for that sample and by the number of theoretically observable peptides for the protein.

Table S2 | TNBC drug screen IC50 data, related to Figure 7
Half maximal inhibitory concentrations (IC50) for each of the TNBC cell lines. Values are reported as $\log_{10}$(molar concentration)

Table S3 | TNBC drug screen full dose response data, related to Figure 7
Raw dose response data for each of the TNBC cell lines

Supplemental Experimental Procedures

Cell Culture

Triple negative breast cancer cell lines were purchased from ATCC (American Type Culture Collection, Manassus, VA). MCF7, SKBR3, BT474, and MCF10A cells were obtained from Dr. Hanna Irie (Mt. Sinai Hospital). MCF10A were grown in DMEM-F12 (Gibco) with addition of 5% horse serum (Gibco), 20ng/mL EGF (Peprotech), 0.5mg/mL hydrocortisone (Sigma), 100ng/mL cholera toxin (Sigma), 10µg/mL insulin (Sigma). All other cell lines were grown in RPMI-1640 (Gibco) with addition of 10% fetal bovine serum (Gibco) and penicillin-streptomycin-glutamine (Gibco). Patient tumor specimens were purchased from Indivumed GmbH.

Sample preparation

Cultured cells were washed 3 times quickly with ice cold phosphate buffered saline and flash frozen on liquid nitrogen. They were scraped directly into chilled denaturing buffer containing 50mM Tris pH 8.2, 75mM NaCl, 9M urea, complete EDTA-free protease inhibitor cocktail (Roche), and phosphatase inhibitors (50mM sodium fluoride, 1mM sodium orthovanadate, 10mM sodium pyrophosphate, 50mM β-glycerophosphate) and sonicated on ice for two cycles of 30s each. Tumor tissues were dounce homogenized on ice in the same lysis buffer above prior to sonication. All lysates were centrifuged at 12,000 g for 10 min to pellet insoluble material, the supernatant assayed for protein content using the bicinchoninic acid method and saved for analysis at -80°C. Protein extracts were reduced with 5mM DTT at 55°C and alkylated with 15mM iodoacetamide at room temperature in the dark. Extracts from each sample (25µg) were diluted and digested in solution overnight with either lysyl-endopeptidase (Lys-C) (Wako) or sequencing grade trypsin (Promega). Digestion products were acidified to pH ~2 and loaded directly onto pre-equilibrated stop-and-go-extraction tips constructed in-house from SDB-XC Empore wafers (3M)(Rappsilber et al., 2007). Peptides were desalted and fractionated on the tips by basic reverse-phase using a step-wise gradient of increasing acetonitrile (5%, 10%, 15%, 25%, 80%) in 0.1% NH_4OH. Finally, fractions were dried by vacuum centrifugation and resuspended in 3% MeCN, 4% formic acid for analysis by LC-MS/MS.

LC-MS/MS

Peptide fractions were injected onto a 40cm x 100µm column packed in-house with 1.9µm Reprosil C18 reverse phase material (Dr. Maisch GmbH), separated by liquid chromatography gradient on an EASY-nLC-1000 (Thermo) equipped with column oven set to 50°C, and analyzed online by tandem mass spectrometry in a hybrid quadrupole-orbitrap Q-Exactive mass spectrometer (Thermo). Mass spectra were acquired in centroid mode using a data dependent acquisition strategy where the twenty most intense precursors were selected for fragmentation, and fragmented ions were excluded from further selection during 40s. Full MS scans were acquired from 300 to 2000 m/z at 70,000 FWHM resolution with a maximum injection time of 100ms and fill target of 3e6 ions. MS/MS fragmentation spectra were collected at 17,500 FWHM with maximum injection time of 50ms using a 2.0 m/z precursor isolation window and fill target of 5e4 ions. Acquisition time for each fraction was 90min, and included column wash and equilibration.

Data processing

Raw spectra were converted to the mzXML open data format and searched using Sequest (release 2012.01.0 of UW Sequest) against a concatenated forward and reverse version of the Uniprot human protein sequence database (v11/29/2012), allowing for up to two missed cleavages, methionine oxidation (+15.9949 Da), and protein N-terminal acetylation (+42.0105 Da). Cysteine carbamidomethylation (+57.0214 Da) was set as a fixed modification. Precursor mass tolerance was set to 50ppm and fragment ion tolerance set to 0.01 Da. Peptide spectral matches for all fractions corresponding to the same sample were filtered to reach a protein identification false discovery rate of less than 1%, resulting in an aggregate peptide-level FDR of less than 0.1%. Peptides were assembled into proteins using parsimony principles (Nesvizhskii and Aebersold, 2005). Integrated MS1 intensity over time peak areas for identified peptides were calculated using an in-house peptide quantification algorithm. Protein quantifications were calculated using the intensity-based absolute quantitation (iBAQ) approach (Schwanhäusser et al., 2011). For each sample (including all fractions), summed peak areas for all peptides matching to the same protein were divided by the maximum number of observable peptides, and normalized to the total sum intensity of the observed proteome. Common contaminants (e.g. keratins, serum proteins) were excluded prior to normalization. Mass spectrometry data was also analyzed using the standalone MaxQuant platform (Cox and Mann, 2008), for which we obtained similar number of peptide and protein identifications as well as quantification values (data not shown).

Drug screen and curve fitting

Cells were added to 384-well plates at a density of 2,000 cells per well in 50µL of RPMI 1640 containing penicillin-streptomycin using a Matrix WellMate liquid handler (Thermo Scientific), and incubated overnight to allow attachment. Compounds were added (50nL ranging from 5pM to 100µM) to cells using the CyBi-Well Vario Workstation (CyBio) and incubated at 37°C, 5% CO_2 for 96 hours. The final solvent (DMSO) concentration in the assay was 0.1%. Cell viability was measured by luminescence using quantitation of ATP as an indicator of metabolically active cells. CellTiter-Glo reagent (Promega) was dispensed into individual wells with the WellMate following the manufacturer's recommended procedures and, following 20 minutes incubation on an orbital shaker, luminescence was measured on an EnVision Multi-label plate reader (Perkin Elmer). Measurements were corrected for background luminescence and percentage cell viability is reported as relative to the DMSO solvent control. Non-linear curve fitting was performed using MATLAB's 'nlinfit' function. After curve fitting, IC50 values were extracted based on the curve fits, similar to the Cancer Cell Line Encyclopedia (CCLE)(Barretina et al., 2012). External drug sensitivity data (IC50) was downloaded from the "Genomics of Drug Sensitivity in Cancer" resource (Yang et al., 2013), release 2.0 (http://www.cancerrxgene.org).

Bioinformatics

Hierarchical clustering and PCA (principal component analysis) were performed using Cluster 3.0. Average iBAQ values were normalized to a scale from 0 to 1 and filtered for presence in at least 25% of samples and differentially regulated proteins (by S.D.). Protein isoforms were not considered and missing values were not imputed. Unweighted clustering was performed in both dimensions using centered correlation as the similarity metric with centroid linkage. Gene ontology (GO) mapping and enrichment analysis was performed using DAVID 6.7 (Database for Annotation, Visualization and Integrated Discovery)(Huang et al., 2009). The cancer gene

census, copy number, and exome data were downloaded from COSMIC (catalog of somatic mutations in cancer)(Forbes et al., 2011). For the cell line SKBR3, mutational data was acquired from CCLE (Barretina et al., 2012). Associations between census gene mutations and protein expression were assessed by using a heteroscedastic unpaired t-test on $\log_{10}$ transformed protein expression values. To generate a network of common gene-protein relationships, we applied a significance threshold ($P < 0.001$) to differences in protein expression that were associated with cancer census gene mutations. These were plotted using Cytoscape version 3.1.0 (Shannon et al., 2003) with a spring-embedded layout. Drug sensitivity associations were assessed by pairwise Pearson's correlation of protein abundance versus inverted IC50 using the 'cor' function in R. Correlation significance was assessed using default settings of the 'cor.test' function in R (a Fisher's Z transformation), and corrected for multiple hypothesis testing using the Benjamini and Hochberg method.

Statistical analysis

Significance tests and correlation analysis were performed using built-in functions within Microsoft Office Excel 2013 or R statistical computing environment version 3.1.0. Gene enrichment significance testing was performed in DAVID version 6.7 using the EASE metric, a modified Fisher's exact test (Huang et al., 2009). All error bars represent standard deviation unless otherwise noted.

Supplemental References

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372.

Forbes, S.A., Bindal, N., Bamford, S., Cole, C., Kok, C.Y., Beare, D., Jia, M., Shepherd, R., Leung, K., Menzies, A., et al. (2011). COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. *39*, D945–D950.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Nesvizhskii, A.I., and Aebersold, R. (2005). Interpretation of shotgun proteomic data: the protein inference problem. Mol. Cell. Proteomics MCP *4*, 1419–1440.

Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906.

Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., et al. (2010). Towards a knowledge-based Human Protein Atlas. Nat. Biotechnol. *28*, 1248–1250.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955–D961.