SUPPLEMENTARY INFORMATION

PROTEOMIC SNAPSHOT OF THE EGF-INDUCED UBIQUITIN NETWORK

Elisabetta Argenzio[1]∗, Tanja Bange[2]∗, Barbara Oldrini[1]∗, Fabrizio Bianchi[1,3], Raghunath Peesari[1], Sara Mari[1], Pier Paolo Di Fiore[1,3,4], Matthias Mann[2], Simona Polo[1,3]

TABLE OF CONTENTS

- Sheet “Endogenous 1175 proteins”. Sheet “TAP 582 proteins”.
- Sheet “Common 284 proteins”.
- Sheet “NR 1472 proteins”.
- Sheet “Ub sites”.
- Sheet “Control purification”.
- Sheet “I-DIRT”.

Table S2. Supplemental data to the identification and analysis of EGF-regulated Ubiproteomes.

- Sheet “endogenous”.
- Sheet “TAP”.
- Sheet “NR”.

Table S3. Functional analysis.

- Sheet “Ub signature”.
- Sheet “Hubs”.
- Sheet “overlap pY”.

REFERENCES TO SUPPLEMENTARY MATERIALS

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Constructs

FLAG-6His-Ubiquitin was engineered by recombinant PCR. The human cDNA encoding for RPS27A (fusion protein of S27a ribosomal protein and ubiquitin) was used as a template for the PCR reaction. Primers were designed with the BAMHI site and 6His sequence at the N-terminus, and the EcoRV site at the C-terminus, respectively. Digested product was cloned in pcDNA3.1 FLAG vector using BAMHI and EcoRV enzymes. Initial experiments performed with this construct showed a constitutive monoubiquitination of Eps15, possibly caused by the unregulated overexpression of the tagged Ub (data not shown). The cDNA of FLAG-6His-Ubiquitin was therefore subcloned into the TET-on inducible vector psG213 using HindIII and EcoRV. HA-Ub and FLAG-tagged Eps15 were previously described (Woelk *et al*, 2006).

GST-tagged S5a was engineered by recombinant PCR. The primers were designed based on the human S5a sequence, and contained the restriction sites for BamH1 and XhoI. Digested PCR product was cloned into the pGEX-6P1 vector (Pharmacia). All constructs were sequence verified.

Immunofluorescence

B82L-EGFR cells were grown on fibronectin-coated coverslips, with or without 4 μg/ml doxycycline, for 24, 48 and 72 h and fixed with 4% paraformaldehyde for 10 min at room temperature. Fixed cells were permeabilized with 0.1% TritonX-100, 0.2% BSA in PBS for 10 min and subsequently blocked with 2% BSA in PBS, for 1 h at room temperature. Coverslips were incubated for 30 min with anti-FLAG antibody,

washed twice with PBS, and then incubated for 30 min with Cy3-conjugated secondary antibodies (Alexa) and for 5 min with DAPI, prior to mounting with Mowiol

Internalization assays with rhodamine-EGF were performed with HeLa (Fig. 4) or MCF10A cells (Fig. S9). Briefly, cells were grown on gelatin-coated coverslips, serum starved for 4 h and then stimulated with rhodamine-EGF (0.5 μg/ml) for the indicated times. After fixation and permeabilization cells were then subjected to anti-EphA2 staining (Fig. S9) or anti-Ub staining (Fig. 4) as described previously (Newton *et al*, 2008). EphA antibody (Upstate) was used at 4 μg/ml, 2FK2 antibody (ENZO) at 5 μg/ml, and K63 or K48 Ub-chain specific antibodies (Genentech) at 1μg/ml. Primary antibodies were detected using Alexa-488 conjugated secondary antibody (Molecular Probes).

Functional ablation of EphA2

Silencing of the EphA2 in MCF10A cells was achieved by transient transfection of siRNA oligos (from Invitrogen). Cells were plated in complete growth medium without antibiotics at 30-50% confluence. For each 10 mm plate, 6 μl of RNAi duplex (10 μM) were diluted in 1 ml Opti-MEM medium containing 20 μl Lipofectamine™ RNAiMAX (Invitrogen). The mixture was incubated for 10-20 min at room temperature and added to cells (10 nM final RNAi concentration). Cells were then incubated for 48 h at 37°C and plated for the subsequent assays. Two different RNAi oligos were used: Oligo 1, matching the UTR region: CCCTGTCCCTCTAGTGCCTTCTTT, Oligo 2, matching the CDS region: CCGGAGGACGTTTACTTCTCCAAGT. In all of the reported assays, the two targeting oligos yielded comparable results. Results obtained with Oligo 1 are shown.

For Oligo 1, a mismatched control was designed by introducing four random mutations in the sequence, and was used in all experiments (labeled "control" in the Figures).

BrdU incorporation assay

MCF10A cells were plated on coverslips at 50% confluency and serum starved for 24 h. Cells were then incubated with EGF (20 ng/ml) for 8 h at 37°C, followed by the addition of BrdU (33.3 μM, SIGMA) for 30 min. After fixation in 4% paraformadehyde, cells were permeabilized in 0.2% Triton-X100, 2% BSA in PBS, blocked in goat serum for 30 min, and incubated with anti-BrdU antibody (Becton Dickinson) in the presence of 3 mM $MgCl_2$, DNAse (Promega), 2% BSA. Primary antibody was detected using Alexa-488 conjugated secondary antibody (Molecular Probes) and nuclei were stained with DAPI. Comparative immunofluorescence analyses were performed in parallel with identical acquisition parameters. Approximately 3000 cells, in duplicate, were screened for each condition. Data are expressed as the mean +/- s.e.m. of three independent experiments.

Cell migration assay

Cell migration assays were performed using a BD Boyden Chamber (BD Biosciences) with 8 μm pores. Both chambers were filled with growth medium (1:1 DMEM/HAM F12, 5% horse serum, 0.5% glutammine, 50 ng/ml cholera toxin, 0.01 mg/ml insulin, 500 ng/ml hydrocortisone). The lower chamber additionally contained EGF (1.5 ng/ml). Serum starved MCF10A cells (4×10^4 cells/well) were seeded into the upper chamber of the transwell and allowed to migrate for 3 h at 37°C. Three replicates for each condition were performed. After the incubation period, cells remaining in the

upper chamber were washed away with PBS and removed by a cotton swab. Fixed cells were then stained with DAPI. Cells were counted in three randomly chosen fields using an inverted fluorescence microscope (10X magnificence). Data are expressed as the mean +/- s.e.m. of three independent experiments.

SILAC labeling of HeLa and B82L-EGFR cells

HeLa and B82L-EGFR cells were grown in DMEM medium supplemented with 10% fetal bovine dialyzed serum (Gibco Invitrogen) and lacking L-Arg and L-Lys (customized product from Gibco Invitrogen). This medium was further supplemented with either 28 mg/L L-Arg $^{12}C_6$, $^{14}N_4$-HCl and 48 mg/L L-Lys $^{12}C_6$, $^{14}N_2$-HCl ("light" medium) or 28 mg/L L-Arg $^{13}C_6$, $^{15}N_4$-HCl and 48 mg/L L-Lys $^{13}C_6$, $^{15}N_2$-HCl ("heavy" medium). Cells were grown in these media for at least 5 replication cycles. The incorporation of labeled amino acids was detected by mass spectrometry analysis of total cell lysates.

Affinity ("endogenous") purification with FK2 antibody

SILAC labeled or unlabeled HeLa cells were serum-starved and treated with 100 ng/ml EGF for 10 min or left untreated. Cells were lysed in JS buffer (50 mM HEPES pH 7.4, 150 mM NaCl, 10% glycerol, 1% Triton X-100, 1 mM EDTA, 1 mM EGTA, 1 mM NEM) with 2% SDS, and TCLs were clarified and assayed for protein content as described above for the TAP protocol (see flow chart, Fig. S1A).

Thirty milligrams of 1:1 mixed labeled/unlabeled lysates were diluted 1:10 with JS buffer and loaded onto two aminolinked-S5a columns at 4°C for 2 h. The S5a flow-through was then loaded onto a carbolinked-FK2 affinity column. FK2-conjugated beads were extensively washed with urea buffer (1M urea, 10 mM Tris pH 8, 250 mM

NaCl, 0.2% Triton X-100, 5% glycerol). Finally, S5a- and FK2-interacting proteins were directly eluted in Laemmli buffer and resolved on 4-12% gradient NuPAGE Novex Bis-Tris gels (Invitrogen, Carlsbad, CA). Gels were stained with a Colloidal Blue Staining Kit (Invitrogen), and lanes were excised in 8 to 10 slices and subjected to a standard in-gel trypsin digestion protocol, as previously described (Shevchenko *et al*, 2006). In the second and third replicate, chloroacetamide was used instead iodoacetamide for the carboxymethylation of cysteines (Nielsen *et al*, 2008). Peptides produced by in-gel tryptic digest were concentrated in a speedvac, acidified to pH <2 with 10% TFA and desalted on C18 STAGE tips, as previously described (Rappsilber *et al*, 2007). The yield (% of ubiquitinated proteins recovered at the end of the purification procedure, with respect to the initial amount of Ub-containing proteins) was determined by quantifying discrete bands present in the blots using Photoshop (data not shown).

Tandem Affinity Purification (TAP)

SILAC labeled or unlabeled B82L-EGFR cells, induced with 4 µg/ml doxycycline for 48 h, were serum-starved and treated with 100 ng/ml EGF for 10 min or left untreated. Cells were lysed in non-denaturing lysis buffer (10 mM Tris/HCl pH 8, 100 mM NaH_2PO_4, 500 mM NaCl, 5% glycerol, 0.2% Triton X-100, 20 mM imidazole, 1 mM NEM, 40 mM NaF, 20 mM Na_3VO_4, EDTA-free protease inhibitor cocktail). Total cell lysates (TCLs) were then clarified by sonication on ice (5 cycles of 20 sec) and centrifugation at 12000 rpm for 10 min at 4ºC, followed by ultracentrifugation at 45000 rpm for 45 min at 4ºC. Protein concentration was measured by the BCA assay (Pierce) following manufacturer's instructions, and mixtures of labeled and unlabeled lysates were combined in protein concentration ratios 1:1 (see flow chart, Fig. S1B).

To preserve protein ubiquitination and fully expose the FLAG-6His tags, extracts were denaturated with 8 M urea. Fifty milligrams of mixed extract at a final protein concentration of 1 mg/ml were incubated with 3 ml of Ni-NTA agarose beads (50% slurry) on a circular rocker for 3 h at 4°C. Protein bound beads were then collected by slow speed centrifugation and washed using a urea gradient (8 – 1M) in lysis buffer for 60 bead volumes. The pH of the washing buffer was also adjusted during the washing step, gradually decreasing from 8 to 6.3 and then rising again to 8 before the elution. Bound proteins were eluted in 4 bead volumes of elution buffer (500 mM imidazole, 1 M urea, 10 mM Tris/HCl pH 8, 250 mM NaCl, 5% glycerol, EDTA-free protease inhibitor cocktail) at room temperature and eluates were dialyzed overnight at 4°C. The dialyzed proteins were incubated with FLAG-agarose beads on a circular rocker for 1 h at 4°C. Anti-FLAG beads were then washed with 50 bead volumes of elution buffer without imidazole. TAP purified proteins were finally eluted with 4 bead volumes of FLAG elution buffer (300 µg/ml FLAG peptide, 1 M urea, 10 mM Tris/HCl pH 8, 250 mM NaCl, 5% glycerol, EDTA-free protease inhibitor cocktail), precipitated with ethanol and dissolved in 100 µl denaturation solution (6 M urea, 2 M thiourea in 20 mM ammonium hydrogen carbonate). Samples were divided in two and processed independently as technical replicates. Proteins were reduced with 1 mM DTT for 30 min. Thiols were carboxymethylated with 5.5 mM chloroacetamide (Nielsen *et al*, 2008) for 20 min in the dark, and proteins were digested with 0.5 µg LysC (Waco) for 3 h. After a 4-fold dilution with 50 mM ammonium bi-carbonate, 0.5 µg of trypsin (Promega) was added and the sample was incubated overnight. All steps were performed at room temperature. The digestion was terminated by acidifying the sample to pH < 2 with 10% TFA. The peptide mixture was then

desalted and purified on C18 STAGE tips, as previously described (Rappsilber *et al*, 2007). The yield was determined as described above for the "endogenous" protocol.

I-DIRT experiment

We evaluated possible contaminants in the TAP approach by performing an adapted I-DIRT experiment (Tackett *et al*, 2005). Briefly, B82L-EGFR cells were grown in light isotopic medium and induced with doxycycline for 48 h (induces FLAG-6His-Ubiquitin expression), or were grown in heavy isotopic medium and left untreated. Lysates were then mixed 1:1 and purified as described above for the TAP protocol. A total of 775 proteins with at least one unique peptide were identified. The overall distributions of the H vs. L ratios for the I-DIRT and TAP experiments are depicted in Fig. S2B. We applied the same filtering criteria as for the "high confidence data sets" (see below). The resulting 587 proteins are listed in Table S1 (sheet I-DIRT). Since contaminants and co-purified proteins are expected to have an H/L ratio of 1, we evaluated proteins with a ratio > 0.75, which have the higher probability of being a contaminant. A total of 32 proteins were present in the "TAP" data set, and were therefore considered as contaminants and thus removed from the "high confidence data set" (5.4 % of the whole data set).

Filtering procedure to define potential regulated proteins

SILAC peptide and protein quantitation was performed automatically with MaxQuant (Cox *et al*, 2008). Potentially up- or down-regulated outliers were identified by a significance score (Significance B), calculated with MaxQuant. This strategy avoids the use of a cutoff based only on empirical decisions. Instead, it allows the identification of proteins that are hyper- or hypo-ubiquitinated using a statistical

measure of how distant the identified proteins are (in terms of H/L ratio) from all other proteins present in the Ubiproteome. Candidates lying comparatively far from the bulk of the distribution were identified by first estimating the variance of the distribution of all protein ratios in a non-parametric way, and then reporting the error function for the z-score corresponding to the given ratio. Significance B takes into account the fact that abundant proteins can be measured more accurately than low abundant proteins in MS. Therefore, for high abundant proteins even small changes (down to 10-15%) might be distinct from the bulk population and represent confident candidates. Considering our "high confident data sets" (see main text) as a starting list, proteins with a p-value ≤ 0.1 (≥ 90% probability that the normalized ratio reported for that considered protein is significant) were considered as outliers in the bulk distribution and were retained for further analyses. We next calculated the coefficient of variability (CV) of normalized ratios across the experimental replicates, which is an estimate of the accuracy of a protein ratio considering replicates. The CV cutoff was arbitrarily set at 10, and all proteins with CV > 10 (i.e., proteins with high variability in experimental replicates) were excluded. Finally, by manual inspection of the final list, we eliminated proteins that present opposite regulation in the replicates. Table S1 shows the relative sizes of the regulated candidates lists obtained following this filtering procedure.

Quantitation of the "signature peptides"

All possible tryptic peptides of Ub can be involved or influenced by chain formation. Thus, with our experimental set up, it was not possible to quantify changes, before and after EGF treatment, in the various chains normalized to the total protein amount. Therefore, we evaluated the ratios of the "signature peptides" for the various chains,

which allows a relative quantitation of the level of a specific chain before and after EGF stimulation. These ratios were assigned by the MaxQuant software (version 12.35) using an FDR for the GlyGly site of 0,1 and the option “match between runs”. This option matches, on the basis of their elution time and mass, sequenced SILAC pairs from previous runs/fractions with SILAC pairs that have not been picked for sequencing but can still be taken into consideration for quantitation. We have also provided standard errors and statistical testing (p-values and multiple testing correction) that determine if the enrichments are indeed significant.

Identification of ubiquitination sites

Ubiquitination sites, including information about number of modified sites within a peptide, PTM score and score difference, were assigned automatically by the MaxQuant software (version 12.35) using the settings described previously (Cox *et al*, 2008). The generated list of potential ubiquitination sites was further validated manually with the following criteria. Since the Ub modification leads to a missed cleavage, all sites assigned to a C-terminal lysine, as well as contaminants and reverse hits, were discarded immediately. Next, the remaining sites were subdivided in three different confident groups. In the group with the highest confidence (group I), the PEP of the modified peptide was < 0.1 and the localization probability score was > 90%. The majority of the modified peptides were sequenced many times and/or the corresponding unmodified peptides without the missed cleavage site were detected with a PEP < 0.1. In group II, the fragmentation spectra did not contain sufficient information to assign the ubiquitination site to a specific lysine residue within the peptide sequence. Usually one lysine was internal and one was at the C-terminus. Thus the former has the highest probability of being the modified site. In group III,

potentially modified peptides were kept as low scoring versions (PEP > 0.1), if the unmodified counterpart was sequenced various times and identified with a PEP < 0.1. The identified sequences are listed with their best identification values in Table S1.

Clustering and Functional analysis

Orthologs in the human and mouse experiments were matched based on the mammalian orthology tables downloaded from the MGI website (http://www.informatics.jax.org/).

GeneSpring GX 7.3.1 (Agilent technologies) or Cluster 3.0 (Michael Eisen, Stanford University) softwares were used to perform the hierarchical clustering analysis. Unless indicated otherwise, $\log_2$ of the normalized protein ratio was used as input data, standard correlation was used as the similarity measure, and the average linkage was used as the clustering algorithm.

The online web tool DAVID (www.david.abcc.ncifcrf.gov. (Dennis *et al*, 2003)) classification system was used to perform ontology analysis with the PANTHER biological processes terms. PANTHER/X ontology (http://www.pantherdb.org) is a controlled vocabulary of molecular function and biological process terms, arranged as directed acyclic graphs (DAGs) similar to Gene Ontology™ (GO), but greatly abbreviated and simplified to facilitate high-throughput analyses (Mi *et al*, 2007; Thomas *et al*, 2003). Briefly, HUGO official genes symbols of the SILAC candidates were uploaded on DAVID and ontology terms enriched with $p < 0.05$ (Fisher's exact test) were retained and further analyzed using Excel.

The Ingenuity pathway analysis software (Ingenuity® Systems, www.ingenuity.com) was used for canonical pathway enrichment and network analyses. Briefly, the list of identified proteins was uploaded in the software and the HUGO official gene symbols

together with IPI accession numbers were used to match to the IPA database. Trends of regulation refer to the average normalized ratios across replicates. The human and mouse genomes were used as background for the analysis of the human- and mouse-regulated identified proteins, respectively. Both the Fisher's exact test, and the Benjamin and Hochberg modified t-test were used to compute statistical significance of the enriched pathways. Interaction networks were built based on the mouse and human IPA interaction databases (Ingenuity® Systems, www.ingenuity.com). Every network was limited to a maximum of 35 proteins (IPA, default settings). The Network p-value is the likelihood that the Network Eligible Molecules that are part of a network are found therein by random chance alone. Functional classification of networks was performed using the Molecular Signature Database (MSigDB) [MSigDB, www.broad.mit.edu/gsea/msigdb. (Subramanian *et al*, 2007)] that contains molecular signatures altered by genetic and chemical perturbations, in addition to canonical pathways and classical gene ontology. A list of 167 network-eligible proteins was derived from the 265 proteins of the NR EGF-Ubiproteome. P-values were calculated in MSigDB with the hypergeometric distribution of overlapping genes over all genes in the gene set.

The JMP IN 5.0 software (SAS) was used to annotate the EGF Ubiproteome (265 proteins; 153 proteins with at least 1 interactor) to the BioGRID database. R software (http://www.r-project.org/) was used to generate 5000 random lists of proteins (with the same size as the BioGRID annotated EGF-Ubiproteome, 153 proteins) present in the BioGRID human database, and to calculate relative random distributions and significance p-values.

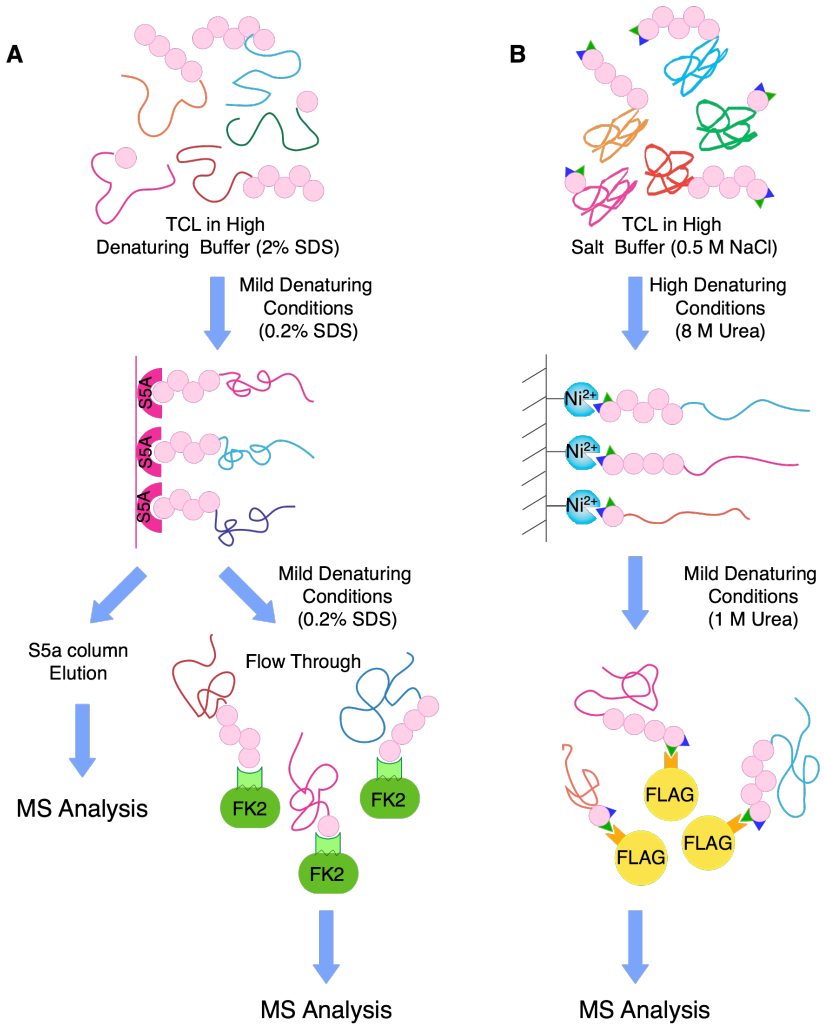

Figure S1

Figure S1. Flow charts of the "endogenous" and "TAP" purification approaches. The extraction of ubiquitinated proteins from cells was carefully set up for both approaches. Briefly, to obtain a snapshot of the ubiquitin network and protect ubiquitinated proteins from the action of deubiquitinases, NEM and denaturing agents

were added to the lysis buffer. Lysates were sonicated to improve the extraction, and then clarified by ultracentrifugation (see Supplemental Experimental Procedures for details). **A.** "Endogenous" approach. Although the presence of K48-linked proteins cannot be excluded *a priori,* it is reasonable to hypothesize that the EGF-Ubiproteome is highly enriched in proteins that are monoubiquitinated or linked to non-degradative polyUb chains (i.e., K63-polyUb chains). Since these species are poorly represented in the steady-state Ubiproteome, we introduced a depletion step with an S5a-based column. S5a is the polyUb-binding subunit of the 19S cap, which should bind to polyubiquitinated proteins that interact with the proteasome. Total cell lysates (TCLs) from HeLa cells were obtained in denaturing buffer containing 2% SDS. Samples were diluted 1:10 and loaded onto a GST-S5a column. The flow-through from the S5a column was then loaded onto an anti-FK2 agarose-conjugated affinity column. After several washes with urea and high salt, anti-FK2 and S5a bound proteins were analyzed by mass spectrometry (Table S1). The control purification was performed following the same procedure, except that avidin-agarose beads (Pierce) were used instead of the anti-FK2 agarose beads (Table S1). **B.** "TAP" approach. FLAG-6His-Ub expression was induced in B82L-EGFR cells for 48 h. TCLs were prepared in high-salt buffer and subsequently diluted and denaturated with a buffer containing 8 M urea. Ubiquitinated proteins were fractionated by Ni^{2+} chromatography, partially renatured on beads and then eluted with imidazole. The eluate was then incubated with anti-FLAG conjugated agarose beads. Anti-FLAG-bound proteins were analyzed by mass spectrometry (Table S1). As a control for the TAP purification, we set-up an I-DIRT experiment in which the light-labeled cells were treated with doxycycline, while the heavy-labeled cells were left untreated (see Supplemental Experimental Procedures).

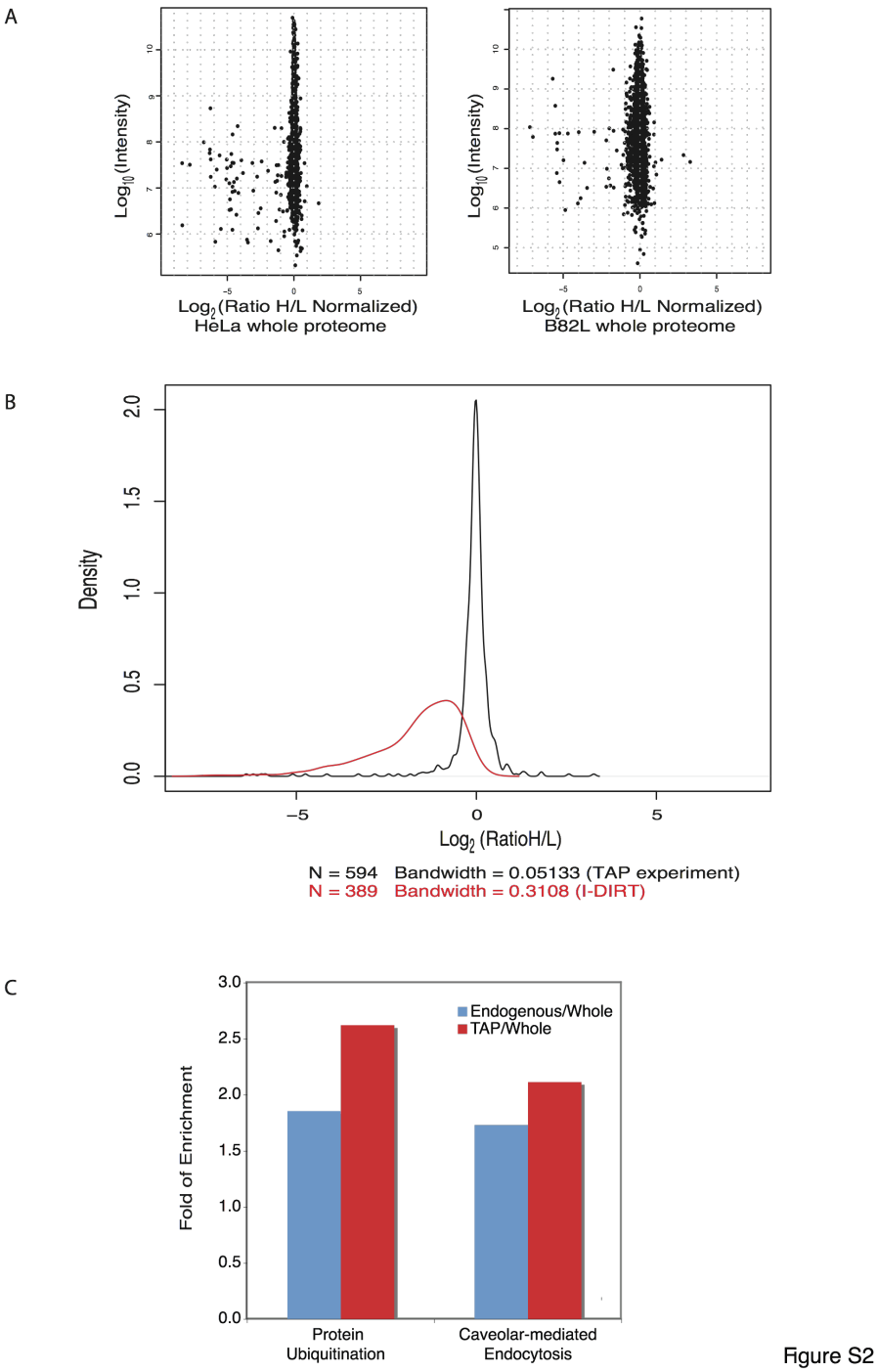

Figure S2. Supplemental data on the analysis of the whole proteomes and Ubiproteomes A. Scatterplot representing H/L ratio vs. intensity of the human (left) and mouse (right) whole proteomes. Whole cell lysates were resolved on a 4-12% gradient NuPAGE Novex Bis-Tris gel and lanes were subjected to a standard in-gel

trypsin digestion protocol, as described previously (REF). MS analysis and quantification were performed as described for the Ubiproteomes (see Experimental Procedures), and 4528 (human) and 4173 (mouse) proteins were unambiguously identified. No major change in protein amounts following 10 min of EGF treatment were observed. X-axis: $\log_2$ transformed normalized protein group ratio H[EGF treated] vs. L[untreated]; Y-axis: $\log_{10}$ of the sum of peptide intensities of the corresponding protein group. **B.** Comparison of the distribution of the H/L ratios for the I-DIRT and TAP experiments. Protein lists were filtered as described for the Ubiproteomes, and the kernel density estimation of the H/L ratios [performed with the R software package (www.r-project.org)] were overlaid. X-axis: $\log_2$ transformed H/L ratios from the two experiments; Y-axis: density. The TAP experiment is shown in black, while the I-DIRT experiment is shown in red. The number of proteins (N) and the calculated bandwidths are indicated. In the I-DIRT experiment, contaminating proteins are expected to have a ratio of 1. However, we observed that the vast majority of peptides were detected in the light version, as demonstrated by a clear shift and broadening towards the left. This result indicates that very few contaminants were isolated in the TAP purification. **C.** Histograms representing the enrichment of representative canonical pathways (identified by Ingenuity) in the total Ubiproteomes over the whole proteomes. Light blue, endogenous approach, red, TAP approach. Notably, a similar enrichment was observed for other pathways, indicating selective enrichment during the purification procedures.

A

	TYPE OF PROTEOME					
	Ubiquitin					Nedd8
	This paper	Matsumoto	Matsumoto (denaturing)	Vasilescu	Meierhofer	Jones
Total Proteins	1472	343	165	71	669	508
Common Proteins (%)		204 (59)	110 (67)	54 (76)	468 (70)	301 (59)

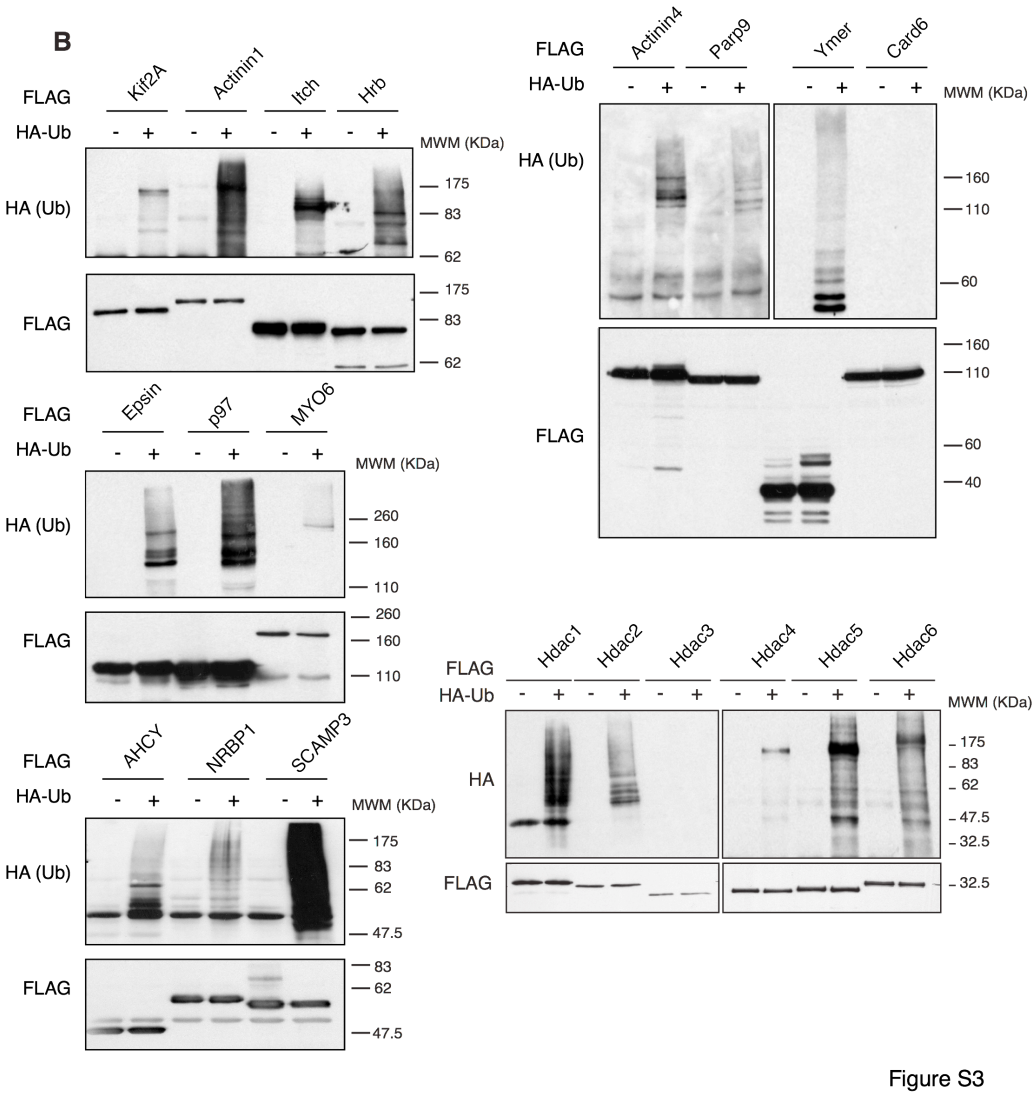

Figure S3

Figure S3. The Ubiproteome is mainly composed of Ub-modified proteins. A. Overlap between our Ubiproteome and published Ubi- (Matsumoto *et al*, 2005; Meierhofer *et al*, 2008; Vasilescu *et al*, 2005) and Nedd8- (Jones *et al*, 2008) proteomes. Total number of proteins identified in our and published Ubi-/Nedd8-

proteomes, and the number and relative percentage of proteins overlapping with our Ubiproteome are reported. HUGO official gene symbols together with IPI accession numbers were used to match databases. Note that the overlap between our Ubiproteome and those of Matsumoto *et al*., was 59% and 67%, for proteomes identified in non-denaturing and denaturing conditions, respectively. This evidence is in agreement with the notion that our Ubiproteome is essentially composed of Ub-modified proteins. **B.** Validation of candidates. 293T cells were transfected with the indicated constructs and lysed after 24 h. Lysates were IP and IB as indicated. To evaluate the false positive-rates of the human and mouse Ubiproteomes, 38 and 30 proteins, respectively, were selected on the basis of availability of tagged constructs. Three proteins scored negative: AP2A1 and CAV1 for the human Ubiproteome; CARD6 (shown in the figure) for the mouse Ubiproteome. Twenty tested candidates are reported as representative examples. Bottom-right panel shows the validation of the entire HDAC family since three members (1, 2 and 6) were in the Ubiproteomes and other two (4 and 5) were identified, but eliminated by the filtering process.

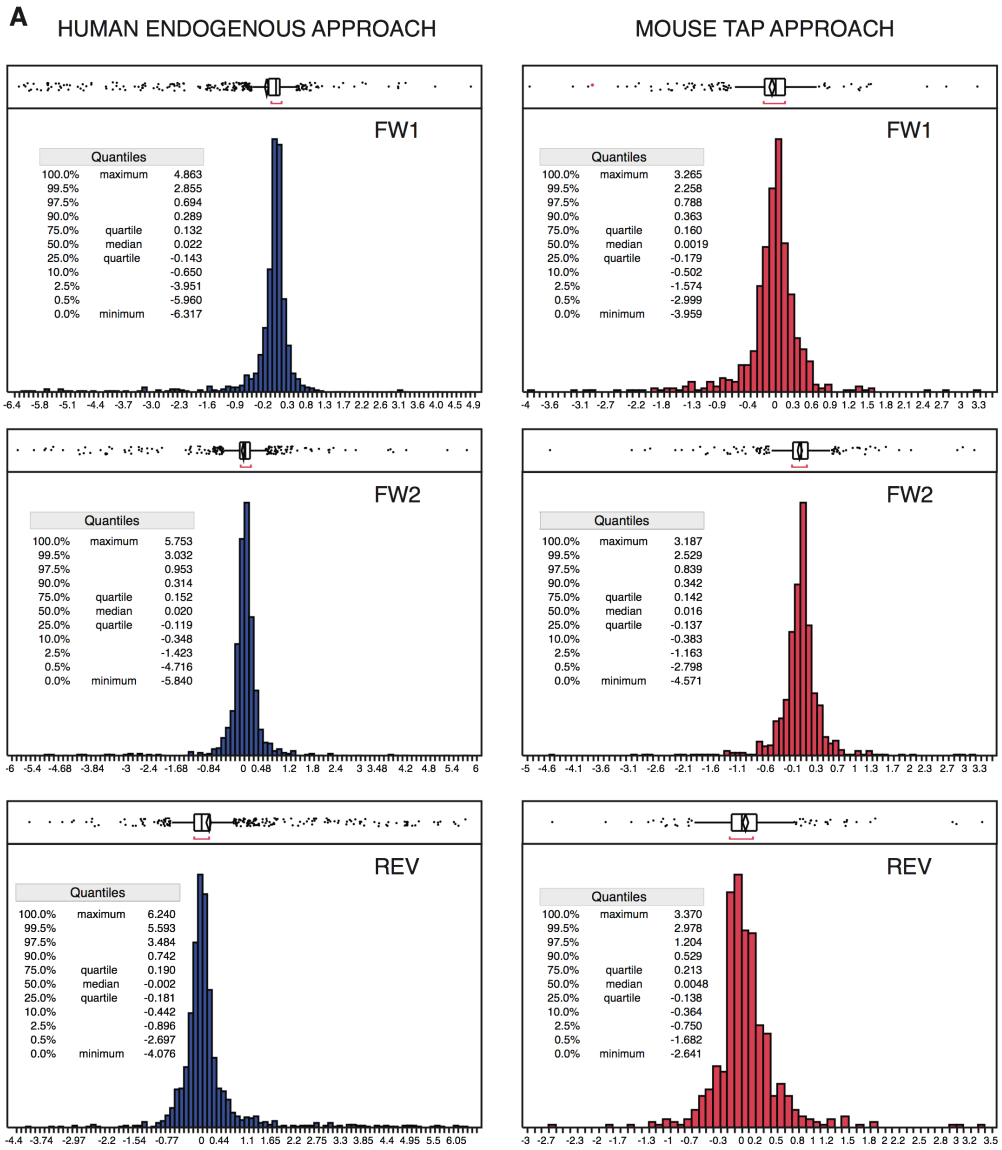

Figure S4

Figure S4. Frequency distributions of log ratios. Bar graphs and box plots of $\log_2$ transformed protein group ratio (H[EGF treated] vs. L[untreated]; values are displayed on x-axis) in the three experimental replicates of the endogenous/human

(left) and TAP/mouse (right) experiments. The reciprocal of the original ratio values was used for the reverse experiment. The quantiles report for each bar graph is shown.

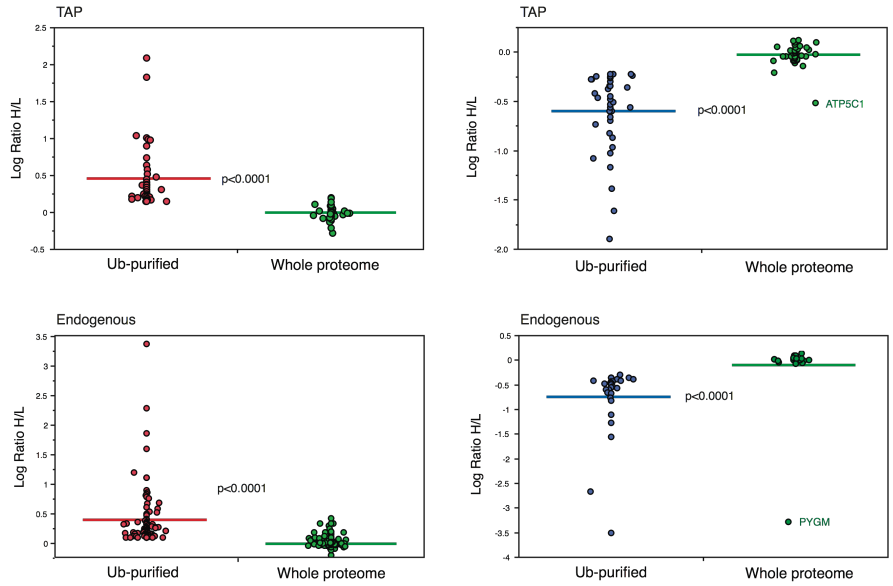

Figure S5

Figure S5. Log ratios of EGF-regulated ubiquitinated proteins. To verify that changes detected in the EGF-Ubiproteomes do not simply reflect changes in protein abundance after EGF (endogenous) or EGF plus doxycycline (TAP) treatment, we compared the proteins mean ratios of the EGF-Ubiproteomes with those of the whole proteomes. Only two proteins, PYGM and Atp5c1, were identified as outliers in the whole proteomes and, therefore, might represent proteins rapidly degraded upon EGF stimulation. Color filled circles represent detected proteins. Mean $\log_N$ ratios for each group are plotted by colored lines. P-values were calculated using the parametric t-test.

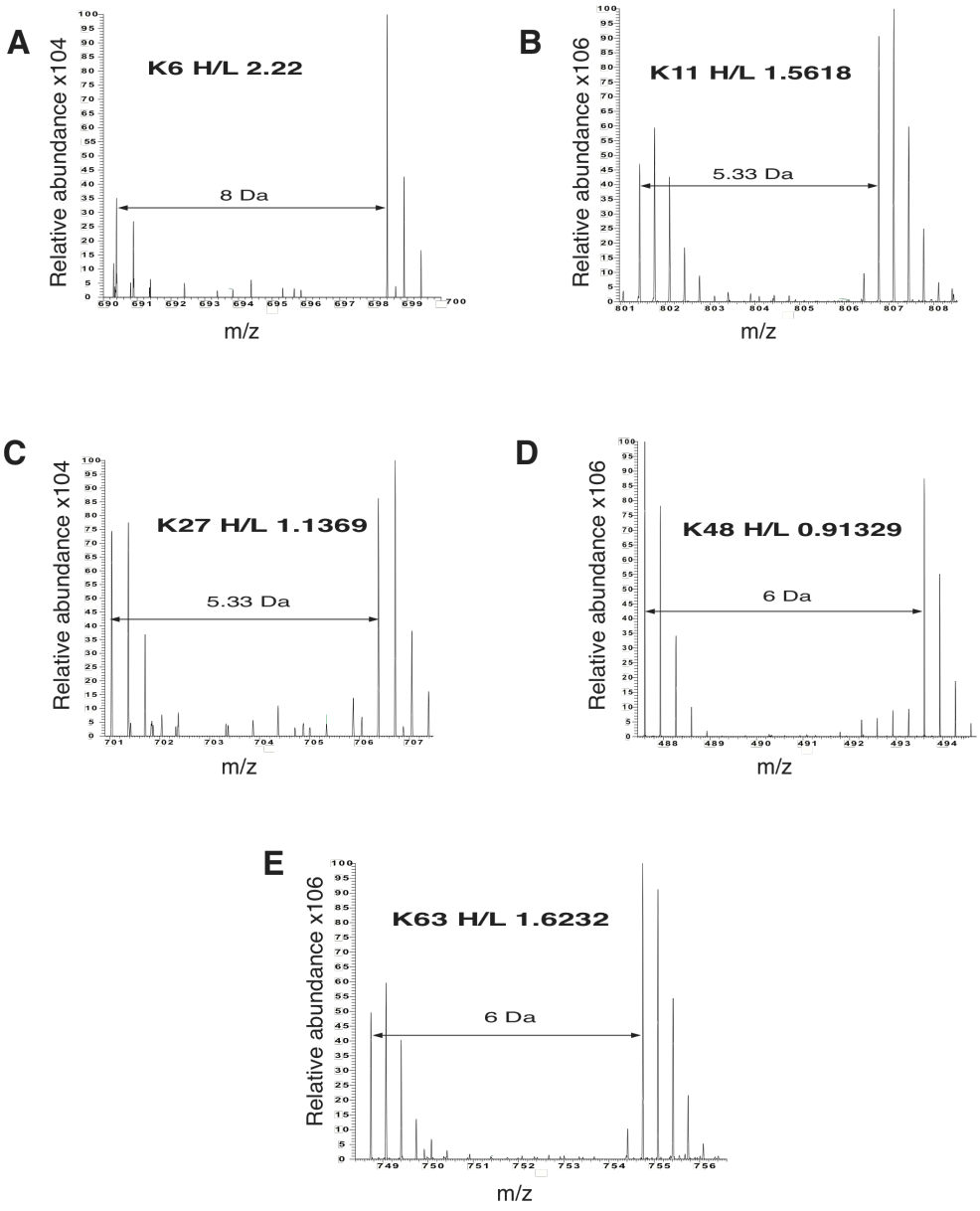

Figure S6

Figure S6 Analysis of Ub chain topology. A. Examples of MS spectra of characteristic "signature" peptide pairs for the following Ub chains are shown. **A.** K6: MH_2^{2+} 690.3894; MQIFVK(+GlyGly)TLTGK. **B.** K11: MH_3^{3+} 801.42688; TLTGK(GlyGly)TITLEVEPSDTIENVK. **C.** K27: MH_3^{3+} 701.03895;

TITLEVEPSDTIENVK(GlyGly)AK. **D.** K48: MH_3^{3+} 487.60005; LIFAGK(GlyGly)QLEDGR. **E.** K63: MH_3^{3+} 748.73761; TLSDYNIQK(GlyGly)ESTLHLVLR.

A

Peptides	Modified Position	Ratio (H/L)	Ratio counts
MQIFVK(GlyGly)TLTGK	K6	1.036	3
TLTGK(GlyGly)TITLEVEPSDTIENVK	K11	1.0496	4
TITLEVEPSDTIENVK(GlyGly)AK	K27	1.076	3
LIFAGK(GlyGly)QLEDGR	K48	1.038	11
TLSDYNIQK(GlyGly)ESTLHLVLR	K63	1.0223	7

B

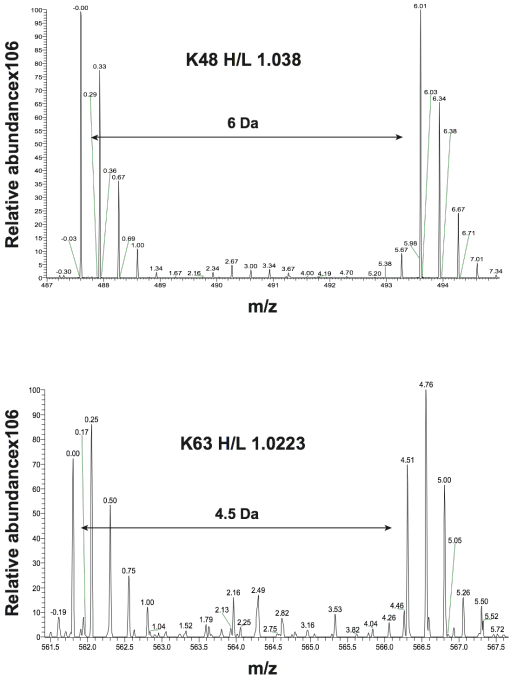

Figure S7

Figure S7. The tagged Ub *per se* does not change the overall level of specific Ub chains. A. To verify that changes detected for the various Ub chains in the TAP approach do not simply reflect problems in chain formation (by E3s) or chain disruption (by DUBs), we analyzed H/L ratios of the "signature" peptides in the whole proteomes of B82L-EGFR cells treated or not with doxycycline (4 μg/ml) for 48 h. No significant changes were observed. **B.** Representative examples of MS spectra of K48 and K63 "signature" peptide pairs are shown.

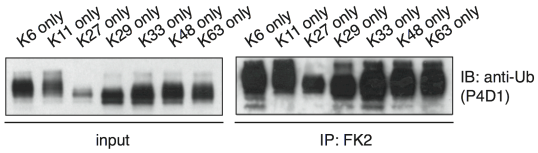

Figure S8

Figure S8. FK2 does not show any preference for linkage-specific chains. Identical amounts (1 μg) of various K-only Ub chains (ENZO) were IP with the FK2 antibody. IB was performed with the P4D1 anti-Ub antibody. Inputs were 100 ng of chains directly loaded on the gel. Note that in all cases a similar stoichiometry of immunoprecipitation was achieved.

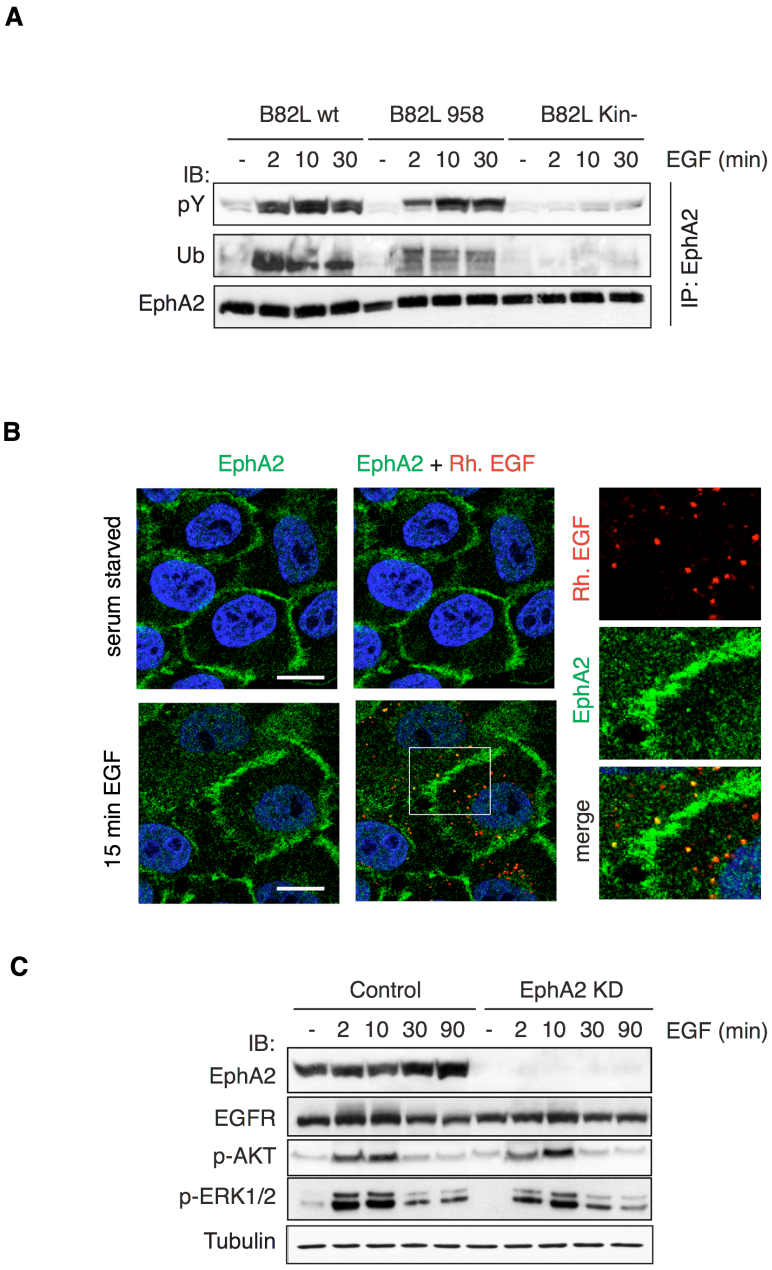

Figure S9

Figure S9. Crosstalk between EGFR and EphA2. A. B82L EGFR wt, 958 and Kin$^-$ cells were serum starved for 24 h and then stimulated with EGF (100 ng/ml) for the indicated times. Lysates (1 mg) were subjected to IP and IB with the indicated antibodies. **B.** MCF10A cells were serum starved for 24 h (upper panels) and then

treated with rhodaminated-EGF (EGF, red) for 15 min at 0.5 μg/ml (bottom panels) and stained by IF using the antibody anti-EpHA2 (green). Blue, DAPI. Rightmost panels, magnification of the RhEGF-treated cells. Bar, 9 μm. **C**. MCF10A cells were subjected to EpHA2-KD, serum starved for 24 h, and then treated for 2 min with EGF (20 ng/ml) for the indicated times. Lysates were subjected to IB with the indicated antibodies.

SUPPLEMENTARY TABLES

Table S1. Supplemental data to the identification and analysis of total Ubiproteomes.

Sheet "EXP size & features". The raw data files were analyzed with the quantitative proteomics software MaxQuant (version 1.0.11.5), developed in-house, which was used for peak list generation, identification and quantitation of SILAC pairs, and filtering (see (Cox and Mann, 2008) for details). -CON/-REV, filtering out of common contaminants, e.g., human keratins and proteases used. FDR, false discovery rates. See Experimental procedures for details.

Sheet "Endogenous 1175 proteins". The human steady-state Ubiproteome (1175 proteins), obtained with the "endogenous" approach, is shown. For each protein the following information is provided:

Cross referencing of output tables	
ID	Unique identifier within the list of protein groups found in the experiment
Annotation data for proteins within the protein group	
Protein IDs	IPI identifier(s) of protein(s) contained within the group
Protein Names	Name(s) of protein(s) contained within the group
Gene Names	Description(s) of gene(s) contained within the group
PEP	Posterior error probability assigned to the protein group
Unique Peptides seq (merged)	Number of different peptide sequences uniquely associated with the protein group. Sum of all three experiments
Ratio H/L Count (merged)	Number of peptides used for quantitation; sum of the quantifiable peptides of the corresponding triplicates
RatioH/L Significance (B) (merged)	Significance is calculated by first estimating the variance of the distribution of all protein ratios in a non-parametric way, and then reporting the error

	function for the z-score corresponding to the given ratio. It also takes into account the dependency of the distribution on the summed protein intensity
Ratio H/LNormalized 1	As above for single experiment 1
RatioH/L Significance (B) 1	As above for single experiment 1
Ratio H/LNormalized 2	As above for single experiment 2
RatioH/L Significance (B) 2	As above for single experiment 2
Ratio H/LNormalized 3	As above for single experiment 3
RatioH/L Significance (B) 3	As above for single experiment 3
Peptides (seq) 1	Specific number of peptide sequences of experiment 1 associated with the protein group
Peptides (seq) 2	Specific number of peptide sequences of experiment 2 associated with the protein group
Peptides (seq) 3	Specific number of peptide sequences of experiment 3 associated with the protein group

Sheet "TAP 582 proteins". The mouse steady-state Ubiproteome (582 proteins), obtained with the "TAP" approach, is shown. For each protein the information provided is listed as described in the "Endogenous 1175 proteins" sheet.

Sheet "Common 284 proteins". The common proteins identified in the endogenous and TAP approaches are listed. Normalized protein group ratios (H [EGF treated] vs. L [untreated]) are reported together with significance B (p-value). See also Experimental Procedures.

Sheet "NR 1472 proteins". The non-redundant proteins identified in the endogenous and/or the TAP approach are listed. Normalized protein group ratios (H [EGF treated] vs. L [untreated]) are reported together with significance B (p-value). See also Experimental Procedures**.** We would like to comment on the fact that we also identified Ub-like (UBL) modifiers, such as Nedd8, SUMO1 and SUMO2 in our Ubiproteome. This is compatible with the possibility that SUMOs and Nedd8 are ubiquitination substrates. Alternatively, SUMOs and Nedd8 might be appended to different Lys residues on the same substrate (combinatorial PTMs). Indeed, growing

evidence suggests that crosstalk exists between sumoylation and ubiquitination (e.g., PCNA (Hoege *et al*, 2002), HDAC1 (Kirsh *et al*, 2002) and RNF4 (Sun and Ballard, 1999). The possibility that UBL modifiers might be themselves ubiquitinated deserves further attention.

Sheet "Ub sites". Ubiquitination sites detected in the "endogenous" and "TAP" purification approaches. Three different classes are defined in the Experimental Procedures. The following information is provided:

Protein ID	IPI identifier(s) of the corresponding protein(s)
Position	Aa position of the modification site within the protein
Protein Names	Name(s) of protein(s) contained within the group
Gene Names	Description(s) of gene(s) contained within the group
Localization Prob	Probability, expressed as a p-value, with which the lysine(s) within a peptide is occupied (p-value = 1 means 100% probability); the values within a peptide sum up to 1. The site with the highest probability is shown as GlyGly modified in the modified sequence column
Score Diff	Difference of the PTM score of the site with the highest score and the site with the next score within the same modified peptide
PEP	Posterior error probability assigned to the peptide
PTM Score	Post-translational modification score, which is a probability based scoring system for site assignment within a peptide (Olsen et al. 2006)
Modified Sequence	Peptide sequence of the modified peptide
Charge	Charge of the tryptic peptide
m/z	Mass to charge value

We identified 31 ubiquitination sites, based on the 114 Da mass shift of a double glycine that remains attached to the modified lysine after trypsin digestion. Among these sites, we confirmed known ubiquitination sites such as: K119 and K121 in histone H2A and H2B, respectively (Weake and Workman, 2008); K164 in PCNA

(Hoege *et al*, 2002); some of the previously identified sites in the EGFR (Huang *et al*, 2006). We also identified novel sites in proteins not previously known to be ubiquitinated. One representative instance of this, referring to the E2 conjugating enzyme Ubc13, is reported in Fig. 2C-D.

Interestingly, we were also able to identify Nedd8 modified at K6. Trypsin digestion of Nedd8 conjugates leaves, similarly to Ub digestion, a double Gly residue signature. Although we cannot exclude that K6 is neddylated (Jones *et al*, 2008), the specificity of our purification procedure allows us to hypothesize that Nedd8 is ubiquitinated at this site. If confirmed, this finding would support the notion of crosstalk between Nedd8 and Ub, as further suggested by the significant overlap between the proteins identified in our screening and those reported in the recently published dataset of neddylated substrates [(Jones *et al*, 2008) and Fig. S3]. Of note, EGFR itself was reported to be neddylated (Oved *et al*, 2006). Further investigations are required to determine whether Nedd8 can be regulated by ubiquitination, or whether Nedd8-Ub mixed chains exist.

Sheet "Control purification". The proteins identified in the control purification are listed. For each protein the information provided is listed as described in the "Endogenous 1175 proteins" sheet. The control purification was performed following the same purification procedure of the endogenous approach except that avidin-agarose beads (Pierce) were instead of the crosslinked FK2 antibody. Three hundred and forty-four proteins were identified with at least two peptides. Two hundred and thirty-two of these 344 proteins were also present in the "endogenous" data set (18.5% of the whole data set). To identify contaminants, we calculated the ratio between the number of sequenced peptides measured in the FK2 and in the control experiments. Proteins identified with a ratio lower than five were considered as

contaminants and removed from our "high confidence data set" (7% of the whole data set, 89 proteins, shown in yellow). However, we cannot formally exclude that they are real ubiquitinated substrates. Nine proteins had a ratio >1 and, thus, could in principle be induced by EGF.

Sheet "I-DIRT". The proteins identified in the I-DIRT experiment are listed. For each protein, the information provided is listed as described in the "Endogenous 1175 proteins" sheet. For 90 proteins, no ratio could be calculated because of the absence of the heavy peptide. These proteins were therefore considered "no-doubt" ubiquitinated proteins. Since contaminants are expected to have a ratio of 1, we evaluated proteins with ratio > 0.75. We found that 32 proteins were also present in the "TAP" data set (shown in yellow). These proteins were thus considered as contaminants and removed from our "high confidence data set" (5.4 % of the whole data set). However, we cannot formally exclude that they are real ubiquitinated substrates. Two proteins had a ratio > 1 and, thus, could in principle be induced by EGF.

Table S2. Supplemental data to the identification and analysis of EGF-regulated Ubiproteomes.

Sheet "endogenous". List of EGF-regulated ubiquitinated proteins in HeLa cells after 10 min of EGF treatment (100 ng/ml). Normalized protein group ratios (H[EGF treated] vs. L[untreated]) are reported together with Significance B (p, p-value; see Experimental Procedures). Red, EGF-induced hyper-ubiquitinated proteins. Blue, EGF-induced hypo-ubiquitinated proteins.

Sheet "TAP". List of EGF-regulated ubiquitinated proteins in B82L-EGFR cells after 10 min of EGF treatment (100 ng/ml) as described in the "endogenous" sheet.

Sheet "NR". List of non-redundant (NR) EGF-regulated ubiquitinated proteins identified with the two approaches, as described in the "endogenous" sheet. Gray color indicates proteins that were not considered to be EGF-regulated in one approach, but were EGF-regulated in the other approach.

Table S3. Functional analysis.

Sheet "Ub signature". Quantitative comparison of "signature" peptides identifying different Ub chain topologies. Peptide sequence, median ratio (H/L), ratio counts, standard error of the mean (SEM), p-value of significance B calculated based on a modification specific peptide table and over all three experiments, for the "endogenous" and "TAP" samples, and the HeLa whole cell lysate are reported. Significance after multiple test correction (Benjamini-Hochberg FDR 0,05) is indicated (+ significantly different, - not significantly different). With the TAP approach, MS analysis revealed an increase in the K63-, K11- and K6-chain modifications after EGF stimulation, whereas with the endogenous approach only K63-linkages accumulated, although not significantly. A possible explanation for this discrepancy is the different yields of the immunoprecipitation procedures. In the endogenous approach only a relatively small portion of total Ub-containing proteins was immunoprecipitated (~8%). Assuming that all chains bind with the same affinity to the antibody, differences in relative chain abundance might not be detectable due to relative small sample size in respect to the whole Ub proteome. If this hypothesis is correct we should be able to score differences in the whole cell lysate prior to the IP enrichment. Indeed MS analysis of whole HeLa lysates used in SILAC experiments revealed that EGF significantly raised the level of K63 linkages leaving K48 unchanged.

Sheet "Hubs". List of the 65 HUBs (proteins with ≥ 5 interactors) in the EGF-Ubiproteome. The JMP IN 5.0 software (SAS) was used to annotate the EGF-Ubiproteome (265 proteins; 153 proteins with at least 1 interactor) to the BioGRID database. Sixty-five proteins with ≥ 5 interactors are considered HUBs and are reported with their number of interactors (N).

Sheet "overlap pY". Comparison with the phospho.ELM database (Diella *et al*, 2008) and published EGF-PY proteomes (Blagoev *et al*, 2004; Hammond *et al*; Oyama *et al*, 2009). The 55 EGF-Ubiproteome proteins, also found to be tyrosine phosphorylated, are listed with their gene names. Black, present. White, absent.

REFERENCES TO SUPPLEMENTARY MATERIALS

Blagoev B, Ong SE, Kratchmarova I, Mann M (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat Biotechnol* **22:** 1139-1145.

Cox J, 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.

Dennis G, Jr., Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA (2003) DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4:** P3.

Diella F, Gould CM, Chica C, Via A, Gibson TJ (2008) Phospho.ELM: a database of phosphorylation sites--update 2008. *Nucleic Acids Res* **36:** D240-244.

Haglund K, Sigismund S, Polo S, Szymkiewicz I, Di Fiore PP, Dikic I (2003) Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation. *Nat Cell Biol* **5:** 461-466.

Hammond DE, Hyde R, Kratchmarova I, Beynon RJ, Blagoev B, Clague MJ (2010) Quantitative Analysis of HGF and EGF-Dependent Phosphotyrosine Signaling Networks. *J Proteome Res*. **9**: 2734-42.

Hoege C, Pfander B, Moldovan GL, Pyrowolakis G, Jentsch S (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* **419:** 135-141.

Huang F, Kirkpatrick D, Jiang X, Gygi S, Sorkin A (2006) Differential regulation of EGF receptor internalization and degradation by multiubiquitination within the kinase domain. *Mol Cell* **21:** 737-748.

Jones J, Wu K, Yang Y, Guerrero C, Nillegoda N, Pan ZQ, Huang L (2008) A targeted proteomic analysis of the ubiquitin-like modifier nedd8 and associated proteins. *J Proteome Res* **7:** 1274-1287.

Kirsh O, Seeler JS, Pichler A, Gast A, Muller S, Miska E, Mathieu M, Harel-Bellan A, Kouzarides T, Melchior F, Dejean A (2002) The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. *EMBO J* **21:** 2682-2691.

Matsumoto M, Hatakeyama S, Oyamada K, Oda Y, Nishimura T, Nakayama KI (2005) Large-scale analysis of the human ubiquitin-related proteome. *Proteomics* **5:** 4145-4151.

Meierhofer D, Wang X, Huang L, Kaiser P (2008) Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry. *J Proteome Res* **7:** 4566-4576.

Mi H, Guo N, Kejariwal A, Thomas PD (2007) PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res* **35:** D247-252.

Newton K, Matsumoto ML, Wertz IE, Kirkpatrick DS, Lill JR, Tan J, Dugger D, Gordon N, Sidhu SS, Fellouse FA, Komuves L, French DM, Ferrando RE, Lam C, Compaan D, Yu C, Bosanac I, Hymowitz SG, Kelley RF, Dixit VM (2008) Ubiquitin chain editing revealed by polyubiquitin linkage-specific antibodies. *Cell* **134:** 668-678.

Nielsen ML, Vermeulen M, Bonaldi T, Cox J, Moroder L, Mann M (2008) Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry. *Nat Methods* **5:** 459-460.

Oved S, Mosesson Y, Zwang Y, Santonico E, Shtiegman K, Marmor MD, Kochupurakkal BS, Katz M, Lavi S, Cesareni G, Yarden Y (2006) Conjugation to Nedd8 instigates ubiquitylation and down-regulation of activated receptor tyrosine kinases. *J Biol Chem* **281:** 21640-21651.

Oyama M, Kozuka-Hata H, Tasaki S, Semba K, Hattori S, Sugano S, Inoue J, Yamamoto T (2009) Temporal perturbation of tyrosine phosphoproteome dynamics reveals the system-wide regulatory networks. *Mol Cell Proteomics* **8:** 226-231.

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

Shevchenko A, Tomas H, Havlis J, Olsen JV, Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat Protoc* **1:** 2856-2860.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102:** 15545-15550.

Sun SC, Ballard DW (1999) Persistent activation of NF-kappaB by the tax transforming protein of HTLV-1: hijacking cellular IkappaB kinases. *Oncogene* **18:** 6948-6958.

Tackett AJ, DeGrasse JA, Sekedat MD, Oeffinger M, Rout MP, Chait BT (2005) I-DIRT, a general method for distinguishing between specific and nonspecific protein interactions. *J Proteome Res* **4:** 1752-1756.

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13:** 2129-2141.

Vasilescu J, Smith JC, Ethier M, Figeys D (2005) Proteomic analysis of ubiquitinated proteins from human MCF-7 breast cancer cells by immunoaffinity purification and mass spectrometry. *J Proteome Res* **4:** 2192-2200.

Weake VM, Workman JL (2008) Histone ubiquitination: triggering gene activity. *Mol Cell* **29:** 653-663.

Woelk T, Oldrini B, Maspero E, Confalonieri S, Cavallaro E, Di Fiore PP, Polo S (2006) Molecular mechanisms of coupled monoubiquitination. *Nat Cell Biol* **8:** 1246-1254.