Supplemental Figure Legends

Figure S1 – Comparison of diGly site alterations in response to Btz or epoxomicin and examination of diGly-site identification overlap between 293T and HCT116 cells related to figure 1.

A) Known ubiquitylation substrates examined in this study. Shown are the E3s for the specific targets as well as primary references

B) The overlap of all quantified sites from four biological replicates of 8hr Btz treatment.

C) The log2 ratio of each ubiquitin diGly containing peptide in response to 8 hr Btz (blue bar) or epoxomicin (red bar) treatment. The diGly modified lysine position is indicated below the y-axis.

D) The log2 ratio of a subset of known ubiquitin modified proteins in response to 8 hr Btz (blue bar) or epoxomicin (red bar) treatment. Error bars represent the SEM of multiple MS1 quantifications for the indicated peptide.

E) Overlap of identified sites from all quantified sites from HCT116 cells with sites identified sites from HEK 293T cells treated with Btz for 8hrs.

All error bars represent the SEM of multiple MS1 quantifications for the indicated site.

Figure S2 – Estimation of ISG15 contribution to observed diGly sites related to figure 2.

A) Whole cell lysates from HCT116 cells treated with Btz for 0, 2, 4, or 8 hours were immunoblotted with the indicated antibody. HCT116 cells were treated with interferon-β for 24 hrs to stimulate ISG15 expression as a positive control.

B) Total number of observed diGly-modified peptides that correspond to ISG15, NEDD8, or Ubiquitin from all datasets in this study.

C) Histogram of the fraction of quantified diGly sites decreased greater than the indicated percentage after treatment with USP2cc.

Figure S3 – Known unstable proteins increase diGly-modified peptide abundance over a time course of Btz treatment related to figure 3.

A) Cells were either untreated (light media) or treated with Btz for 0, 2, 4, or 8 hours (heavy media). Shown is the distribution of log2 ratios for all quantified peptides with 0 (blue bars), 2 (red bars), 4 (green bars), and 8 (purple bars) hours of bortezomib treatment.

B) The log2 ratio (top) and total spectral counts (TSCs) (bottom) of each ubiquitin diGly peptide in untreated cells (blue bars) or cells treated with Btz for 2 (red bars), 4 (green bars), or 8 (purple bars) hours. Error bars represent the SEM of multiple MS1 quantifications for the indicated peptide.
C) The distribution of proteins with a diGly-modified peptide in one or more groups as indicated in Figure 2. The number above each bar indicates the total number of proteins with sites that cluster with a single group or multiple clusters.
D) The log2 ratio (top) and total spectral counts (bottom) of known unstable proteins in untreated cells (blue bars) or cells treated with Btz for 2 (red bars), 4 (green bars), or 8 (purple bars) hours. The protein corresponding to each peptide and the position of the modified lysine is indicated. Error bars represent the SEM of multiple MS1 quantifications for the indicated peptide.

Figure S4 – Western blot validation of protein level alterations related to figure 4.
A) Cells were treated as with Btz or MLN4924 for 0,2,4,or 8 hrs, or treated with both Btz and MLN4924 for 8 hrs. Whole cell extracts were immunoblotted with the indicated antibody.
B) Scatter plot of normalized TSCs for diGly vs protein comparing diGly modifications mapped to the protein of comparison.

Figure S5 – Comparison of log2 ratios from cells treated with Btz and CHX or with CHX alone related to figure 5.
A) Distribution of log2 abundance ratios from K8 labeled cells treated with Btz and CHX for 8hrs (green line) or with CHX alone (red line).
B) Log2 ratios from unlabeled, Btz treated cells mixed with K8 labeled cells treated with Btz and CHX (blue line) compared to 8hr CHX treatment alone (red line).

Figure S6 – Schematic of multi-classifer criteria and demonstration of Ub E1 dependent neddylation related to figure 6.
A) Schematic of muliti-classifer criteria used to categorize diGly-modified sites.
B) HCT116 cells were treated with Btz, MLN4924, and a pan-E1 inhibitor as indicated. Whole cell lysates were probed with α-NEDD8 antibodies.

Figure S7 – Comparison of diGly sites identified within this study and previous studies related to figure 7.

A) Comparison of diGly-modified peptides observed in three previous studies (Danielsen et al., 2011; Meierhofer et al., 2008; Xu et al., 2010) to our dataset with sites in common (blue) and sites we did not observe (red). The percentage overlap is indicated.
B) Total number of previously identified diGly-modified lysines from the studies listed in A, compared to the sites identified in this study (sites identified in 293T cells are excluded). Boxes are drawn to scale
C) Distribution of log2 abundance ratios from K8 labeled cells treated with Btz and CHX for 8hrs (green line) or with CHX alone (red line). Log2 ratios from unlabeled, Btz treated cells mixed with K8 labeled cells treated with Btz and CHX (blue line) compared to 8hr CHX treatment alone (red line).

Supplemental Methods

Antibodies

The following antibodies were utilized in this study. Cdc2 (#9116), CDT1 (#3386), CK1 (#2655), DCBL2 (#4804), FASN (#3180), HIF1α (#3716), Lamin A/C (#4777), NQO1 (#3187), GADD45 (#4632), SMC1 (#4802), Rpb1CTD (#2629), ASCT2 (#5100), Aurora B (#3094), 14-3-3 θ (#9638), 14-3-3 τ/γ (#9639), Ku80 (#2180), β-actin (#4967), Ub-H2B (#5546), PP2A (#2041) were obtained from Cell Signaling Technology (catalog number). NRF2 (2178-1), GLUL (S2813), ACL (#1699-1), ISG15 (#2625-1) from Epitomics, H2B (61037) from Active Motif, CCT5 (#LS-C81945) from Lifespan Biosciences, CSN3 (#A300-012A), SMC3 (#A302-068A), SETX (#A301-104A), CSN7b (#A300-240A) from Bethyl Laboratories, Ubiquitin (MAB1510) from Millipore. Antibodies to NEDD8 were a gift from Millennium Pharmaceuticals.

Production of diGly Antibody

A rabbit polyclonal antibody that recognizes the diGly remnant was produced following the general motif antibody immunization strategy as described previously (Zhang et al., 2002). The peptide library immunogen had the sequence CXXXXXXK(GG)XXXXXX, where X is all common amino acid residues except cysteine and tryptophan and K(GG) is glycine-glycine dipeptide attached to the epsilon amino group of the lysine residue through an isopeptide bond, synthesized as described previously (Gerber et al., 2007). Rabbits were immunized with the peptide library immunogen crosslinked to KLH carrier protein through its amino-terminal cysteine residue. Since the antibody was designed to recognize tryptic digested

proteins, it could not be evaluated by Western blots of undigested cellular proteins; instead antisera from the rabbits were characterized by peptide ELISA against immunogen peptide and unmodified control lysine peptide. After producing the polyclonal antibody we derived a monoclonal antibody from rabbit splenocytes using CST's proprietary XMT technology (http://www.cellsignal.com/technologies/xmt/index.html).

Sample Preparation

HCT116 cells were grown in lysine and arginine free DMEM supplemented with 10% dialyzed FBS (Gibco), 2mM L-glutamine, pen/strep, and light (K0) lysine (50μg/mL) and arginine (85μg/ml). Heavy media was the same except the light lysine was replaced with K8-lysine (Cambridge Isotopes) at the same concentration. Cells were treated with bortezomib, MLN4942, and cycloheximide at 1μM, 1μM, and 1μg/ml, respectively for the times indicated. Bortezomib, MLN4924, and the pan-E1 inhibitor, (Brownell et al., 2010)) were obtained as gifts from Millennium Pharmaceuticals. After the indicated treatments, heavy and light grown cells were mixed 1:1 by cell number and lysed in 6 mL of denaturing lysis buffer (8M Urea, 50mM Tris pH8, 50mM NaCl, protease inhibitors (EDTA-free, roche), and phosphatase inhibitors. Samples were incubated on ice for 10 min and then sonicated with 3x 10s pulses. We typically obtained 25-35 mg of total protein. Lysates were digested with trypsin as described previously (Villen and Gygi, 2008) with one modification. Prior to trypsin digestion, lysates were diluted 1:1 with 50mM Tris pH8 to lower the urea concentration to 4M and were digested with 5ng/μl lys-C (Wako) for 1 hr at room temperature. Four multiple myeloma cell lines AMO-1, L363, RPMI8226, and U266 were grown in medium RPMI-1640 medium (Gibco) free of lysine and arginine supplemented with 10% dialyzed FBS (Gibco), pen/strep, and light (K0) lysine (Sigma, 50 ug/ml) and arginine (Sigma, 100 ug/ml) or heavy lysine (U-13C6, 50 ug/ml, Cambridge Isotopes), and heavy arginine (U-13C6, 100 ug/ml. Cambridge Isotopes). DMSO was added to light media as mock treatment and heavy cells were treated with bortezomib (4 nM, or 10 nM) for 24 hours.

Immunoprecipitation of diGly containing peptides

Lyophilized peptides from 25-35mg of digested proteins were resuspended in 1.5ml of IAP buffer (50mM MOPS pH7.4, 10mM Na_2HPO_4, 50mM NaCl) and centrifuged at max speed for 5 minutes to remove any insoluble material. The supernatant was incubated with α-diGly antibody (32μg/IP) coupled to protein A agarose beads for 1hr at 4˚C and washed with IAP buffer 3x and once with PBS. The eluted peptides were filtered with 0.2μm filters to minimize remove trace agarose. Peptides were desalted using C18 stage-tip method (Rappsilber et al., 2003) and resuspended in 5% formic acid prior to mass spec analysis.

Each lysate was immunoprecipitated sequentially four times and analyzed separately by mass spectrometry.

USP2cc treatment

For samples utilized for USP2cc treatment, heavy and light cells were harvested separately after 8 hr bortezomib treatment and lysed in 2mL non-denaturing buffer (0.5% NP40, 150mM NaCl, 50mM Tris pH8, protease inhibitors, phosphatase inhibitors, and 5mM 1-10 orthophenathroline). Samples were sonicated and 10ug of purified USP2cc was added to the heavy lysate and both the heavy and light lysates were incubated at room temperature for 1 hr. After USP2cc incubation, lysates were denatured by addition of urea to a concentration of 8M. Lysates were mixed 1:1 based on total protein as determined by BCA (Pierce-Thermo) and prepared for digestion as detailed above.

SILAC Switching

HCT116 cells grown in media containing light lysine were switched to media containing heavy lysine (K8) for 4 hours. Following the 4 hr K8 labeling, cells were either untreated, treated with bortezomib, or both bortezomib and cycloheximide for 8 hrs in K8 media. Cells were harvested, lysed, and digested as before.

Mass Spectrometry Analysis

Desalted and lyophilized diGly-captured peptides were resuspended in 5µl formic acid and 4µl of the sample was used for mass spec analysis. The peptides were separated on 100µm x 20cm C18 reversed phase (Maccel C18 3µ 200Å, The Nest Group, Inc.) with a gradient of 6% to 27% acetonitrile in 0.125% formic acid over 165min (Haas et al., 2006; Tolonen et al., 2011). The twenty most intense peaks from each full MS scan acquired in the Orbitrap were selected for MS/MS in the linear ion trap. Specific settings for the mass spectrometer can be found in the RAW files provided.

Identification of Peptides and Proteins

Following data acquisition, RAW files corresponding to all LC-MS/MS analyses were processed using software to detect and correct instrument-borne errors in precursor monoisotopic peak assignment while refining precursor mass measurements. MS/MS spectra were then exported as DTA files and searched using Sequest (Eng et al., 1994) against a composite protein database containing all protein sequences from the human IPI database (Kersey et al., 2004) in forward as well as reverse orientation, as well as common contaminating protein sequences. To accommodate the possibility of two variable modifications on Lysine residues in these experiments, each LC-MS/MS analysis was searched twice: while the mass of

Lys was assigned to its natural abundance value during the first search, during the second search its mass was increased to reflect SILAC labeling (+8.014199); diGly modification of Lys (+114.042927) were set as a variable modification in both searches. The heavy- and light-labeled peptides from each search were subsequently combined using custom scripts. Other parameters used for database searching include: 50 ppm precursor mass tolerance; 1.0 Da product ion mass tolerance; tryptic digestion with up to three missed cleavages; and variable oxidation of Met (+15.994946). For selected experiments search parameters also included variable SILAC labeling of Arg (+10.008269).

Because database search algorithms such as Sequest return a mixture of random spectral matches along with correct peptide identifications, additional filtering is required. Using the target-decoy search strategy as a guide (Elias and Gygi, 2007), we employed a multivariate approach based on linear discriminant analysis to distinguish correct and incorrect peptide identifications using parameters including Xcorr, ΔCn' (defined as the normalized difference in Xcorr between the first-ranked peptide identification and the next ranked identification corresponding to a distinct amino acid sequence, and not simply a rearrangement of post-translational modifications), precursor mass error, numbers of missed cleavages, and peptide length (Huttlin et al., 2010). Peptides with length shorter than six amino acids were discarded prior to classification and the remaining peptides were filtered to an initial peptide-level false discovery rate of 1%, retaining only those peptides containing at least one diGly-modified Lys residue. Peptides were then grouped into proteins and scored using a probabilistic model and filtered to a final protein-level false discovery rate of 1%. Following this subsequent filtering, the final peptide-level false discovery rate was well below 1%. To account for potential redundancy, each peptide was mapped to all matching IPI entries. For purposes of site and protein counting as well as data representation, sites were exclusively assigned to the minimal set of proteins required to account for all observed peptides.

Localization of diGly sites

To score the localization of individual diGly modifications, an algorithm derived from the AScore was used, with minor modifications (Beausoleil et al., 2006). Rather than iteratively considering multiple peak depths during analysis, ranging from 1 to 10 peaks per 100 m/z window, a constant peak depth of 5 peaks per 100 m/z window was assumed. Given each peptide sequence and number of diGly modifications, all possible permutations were enumerated and scored using a binomial model to evaluate the similarity of their theoretical fragmentation patterns to the observed MS/MS spectrum. Site-determining ions were then determined for each site of diGly modification by comparing the predicted MS/MS fragmentation pattern for the best-scoring peptide sequence with the next best scoring alternative site assignment for each diGly modification and calculating the cumulative binomial probability

associated with observing the number of site-determining ions that were found in the MS/MS spectrum after filtering to the top 5 peaks per 100 m/z window. This was then converted to a human-readable score by calculating $-10*\text{Log}_{10}(p)$ where p is the cumulative binomial probability associated with the observed number of matched site-determining ions. Sites scoring above 13 ($p < 0.05$) were considered localized; peptides whose sites failed to meet this minimum threshold were discarded. Peptides whose sequences only allow a single arrangement of diGly modifications are unambiguously localized and were assigned scores of 1000. Due to previous reports that ubiquitylation blocks tryptic cleavage immediately following modified residues (Seyfried et al., 2008; Shi et al., 2011) we did not allow diGly sites to occur on C-terminal Lysine residues. The positions of all modified lysines are measured with respect to the initiator methionine, regardless of any N-terminal processing.

Typical results for a single LC-MS Analysis

In a single LTQ-Orbitrap Velos LC-MS analysis used in this study, we typically collected 34-35 thousand MS/MS spectra, resulting in 7,000-9,000 peptide-spectral matches at a peptide-level false discovery rate of 1%. Typically about 70% of these identified peptides contained one or more diGly modifications. After additional filtering to control the protein-level false discovery rate at a level of 1%, we typically accepted 6000-8000 peptide-spectral matches from around 2000 proteins. Finally, peptides containing GG sites that could not be confidently localized were removed. After all filtering steps were complete, typically less than 0.5% of peptides contained localized diGly modifications on c-terminal Lys residues. Such numbers are consistent with the peptide and protein-level false discovery rates that were observed. These peptides containing diGly modifications on c-terminal Lys residues were discarded prior to assembling the final dataset.

Protein and Site Quantification

SILAC ratios (Heavy: Light) were determined for each peptide in automated fashion by generating extracted ion chromatograms within 10 ppm of the observed m/z for the monoisotopic and +1 isotopes for both heavy and light isotopomers. Extracted ion chromatograms were then integrated and the sum of heavy EIC's was divided by the sum of the light EIC's to determine the SILAC ratio. Ratios for each peptide then underwent Log_2 transformation. Signal-to-noise ratios were then determined for both heavy and light isotopic envelopes by comparing their observed signal intensities with the median signal intensity observed in nearby m/z ranges within several minutes of the peptide's elution. In cases where only a single isotopomer (heavy or light) was detectable, the signal-to-noise ratio was used for

quantitation instead. For an individual peptide to be used for quantification of sites and proteins, it had to meet one of two conditions: i) Both heavy and light isotopic envelopes must be detected with signal-to-noise ratios above 5.0; ii) One isotopic envelope (heavy or light) must have a signal-to-noise ratio above 10.0.

Proteins and sites of diGly modification were quantified by grouping all matching peptides and filtering as described above. The median Log_2 ratio for all matching quantified peptides was taken as an estimate of the Log_2 ratio for each protein or site and the standard deviation across all quantified peptides was calculated as an indication of variability. A subset of sites and proteins were identified based on peptides whose identifications were confident, but whose quantification failed to meet our minimum requirements. The identities of these are included with quantification represented as "NA".

When peptides were assigned to quantify individual diGly sites, only peptides containing a single diGly site were used for quantification. Peptides containing multiple diGly sites represent a distinct modification state for the protein and thus could show different ratios than would be observed for any of the individual sites of modification. Thus, peptides containing multiple diGly sites were grouped based on the combinations of sites they contained and each combination of sites was quantified separately. Thus, ratios reported for each individual site truly reflect the fold change for peptides observed strictly in the presence of that single PTM.

Comparison with Previous Datasets

Three previous papers have published sets of ubiquitylation sites identified by mass spectrometry following affinity purification (Danielsen et al., 2011; Meierhofer et al., 2008; Xu et al., 2010). To compare sites from these studies with those identified in our work, lists of peptides were extracted from supplementary tables accompanying each publication. These peptides were then mapped to the IPI human database (Version 3.60) (Kersey et al., 2004) and compared against a master list of all sites observed in the present study.

Another recent publication surveyed Lysine acetylation in human cells (Choudhary et al., 2009). Similarly, to determine whether acetylation influenced an individual Lysine residue's likelihood of also undergoing ubiquitylation, the 3600 acetylation sites reported in the literature were downloaded from supplementary tables and compared against our database of diGly sites. The frequency with which acetylated Lysines were also observed to bear diGly modifications was compared first against the overall frequency with which diGly sites were observed when compared with the total number of Lysines in the IPI human proteome. The frequency of observing sites that bore both acetylation and ubiquitylation was also

compared against the background frequency with which Lysines were found to be ubiquitylated within only those proteins found to contain at least one acetylation site. Both comparisons indicated a highly significant enrichment in diGly modifications among sites that were also acetylated, as measured by a Binomial test.

Analysis of Sequence surrounding diGly Sites

To identify amino acids that were over- or under-represented in protein sequences surrounding observed diGly sites, we made a list of the amino acid sequences within +/- 6 residues of each Lysine in those proteins that were observed to contain at least one diGly site. (Only the minimal set of proteins needed to account for all sites was used for this analysis.) In cases where a Lysine fell within six residues of the protein N- or C-terminus, the character 'x' was inserted to fill any empty positions. The frequency with which each amino acid was found in each position was then calculated. Separately, amino acid sequences within +/- 6 residues of each observed diGly site were extracted and the numbers of occurrences of each amino acid in each position were counted. A binomial test was then used to evaluate the significance of enrichment or depletion by comparing the number of observations of each amino acid to its underlying background frequency in that position, given the number of diGly sites considered. The resulting p-values were then Log_{10}-transformed and plotted as a heat map, with enrichment represented by red and depletion represented by blue.

Pfam Domain Analysis

To examine the distribution of diGly modifications on proteins containing various Pfam domains (Finn et al., 2010) and within those domains themselves, Pfam domains were mapped to all protein sequences within the human IPI database (version 3.60) (Kersey et al., 2004). Subsequently, those proteins containing each domain were identified. The number of proteins containing each domain was noted, as well as the number of those proteins that also bore at least one diGly modification. Enrichment or depletion of diGly modifications among proteins bearing each domain was evaluated via a hypergeometric test. In addition, to evaluate levels of diGly modification within the boundaries of each domain, the number of Lysines falling inside the domain and the total number of Lysines overall were determined across all protein sequences containing each domain. The enrichment or depletion of diGly sites within each domain was then evaluated by comparing the numbers of diGly sites seen inside the domain with the underlying distribution of Lysine residues across matching proteins, again using the hypergeometric distribution. P-values were then adjusted for multiple hypothesis testing using the Benjamini-Hochberg approach (Benjamini and Hochberg, 1995) and a 5% FDR threshold was selected for identifying significant enrichment or depletion.

Data Dissemination

The RAW files will be deposited into Proteome Commons (http://www.proteomecommons.org) or provided upon request. A website with all the site and protein data will be available upon acceptance of the paper https://gygi.med.harvard.edu/ggbase/.

Supplementary References

Beausoleil, S. A., Villen, J., Gerber, S. A., Rush, J., and Gygi, S. P. (2006). A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol *24*, 1285-1292.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B *57*, 125-133.

Brownell, J. E., Sintchak, M. D., Gavin, J. M., Liao, H., Bruzzese, F. J., Bump, N. J., Soucy, T. A., Milhollen, M. A., Yang, X., Burkhardt, A. L*., et al.* (2010). Substrate-assisted inhibition of ubiquitin-like protein-activating enzymes: the NEDD8 E1 inhibitor MLN4924 forms a NEDD8-AMP mimetic in situ. Mol Cell *37*, 102-111.

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Walther, T. C., Olsen, J. V., and Mann, M. (2009). Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science *325*, 834-840.

Danielsen, J. M., Sylvestersen, K. B., Bekker-Jensen, S., Szklarczyk, D., Poulsen, J. W., Horn, H., Jensen, L. J., Mailand, N., and Nielsen, M. L. (2011). Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. Mol Cell Proteomics *10*, M110 003590.

Elias, J. E., and Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods *4*, 207-214.

Eng, J. K., McCormack, A. L., and Yates, J. R., 3rd (1994). An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom *5*, 976-989.

Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K*., et al.* (2010). The Pfam protein families database. Nucleic Acids Res *38*, D211-222.

Gerber, S. A., Kettenbach, A. N., Rush, J., and Gygi, S. P. (2007). The absolute quantification strategy: application to phosphorylation profiling of human separase serine 1126. Methods Mol Biol *359*, 71-86.

Haas, W., Faherty, B. K., Gerber, S. A., Elias, J. E., Beausoleil, S. A., Bakalarski, C. E., Li, X., Villen, J., and Gygi, S. P. (2006). Optimization and use of peptide mass measurement accuracy in shotgun proteomics. Mol Cell Proteomics *5*, 1326-1337.

Huttlin, E. L., Jedrychowski, M. P., Elias, J. E., Goswami, T., Rad, R., Beausoleil, S. A., Villen, J., Haas, W., Sowa, M. E., and Gygi, S. P. (2010). A tissue-specific atlas of mouse protein phosphorylation and expression. Cell *143*, 1174-1189.

Kersey, P. J., Duarte, J., Williams, A., Karavidopoulou, Y., Birney, E., and Apweiler, R. (2004). The International Protein Index: an integrated database for proteomics experiments. Proteomics *4*, 1985-1988.

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

Rappsilber, J., Ishihama, Y., and Mann, M. (2003). Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. Anal Chem *75*, 663-670.

Seyfried, N. T., Xu, P., Duong, D. M., Cheng, D., Hanfelt, J., and Peng, J. (2008). Systematic approach for validating the ubiquitinated proteome. Anal Chem *80*, 4161-4169.

Shi, Y., Xu, P., and Qin, J. (2011). Ubiquitinated proteome: ready for global? Mol Cell Proteomics *10*, R110 006882.
Tolonen, A. C., Haas, W., Chilaka, A. C., Aach, J., Gygi, S. P., and Church, G. M. (2011). Proteome-wide systems analysis of a cellulosic biofuel-producing microbe. Mol Syst Biol *7*, 461.
Villen, J., and Gygi, S. P. (2008). The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat Protoc *3*, 1630-1638.
Xu, G., Paige, J. S., and Jaffrey, S. R. (2010). Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. Nat Biotechnol *28*, 868-873.
Zhang, H., Zha, X., Tan, Y., Hornbeck, P. V., Mastrangelo, A. J., Alessi, D. R., Polakiewicz, R. D., and Comb, M. J. (2002). Phosphoprotein analysis using antibodies broadly reactive against phosphorylated motifs. J Biol Chem *277*, 39379-39387.

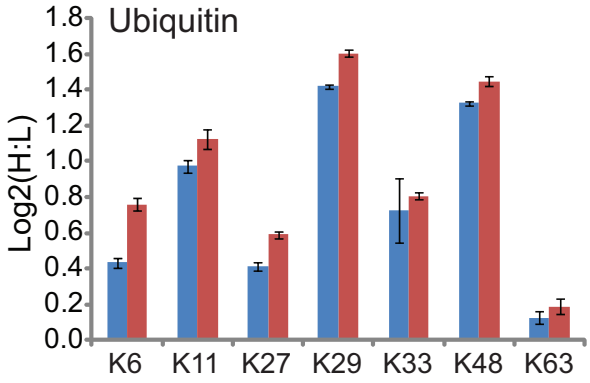
Figure S1

A

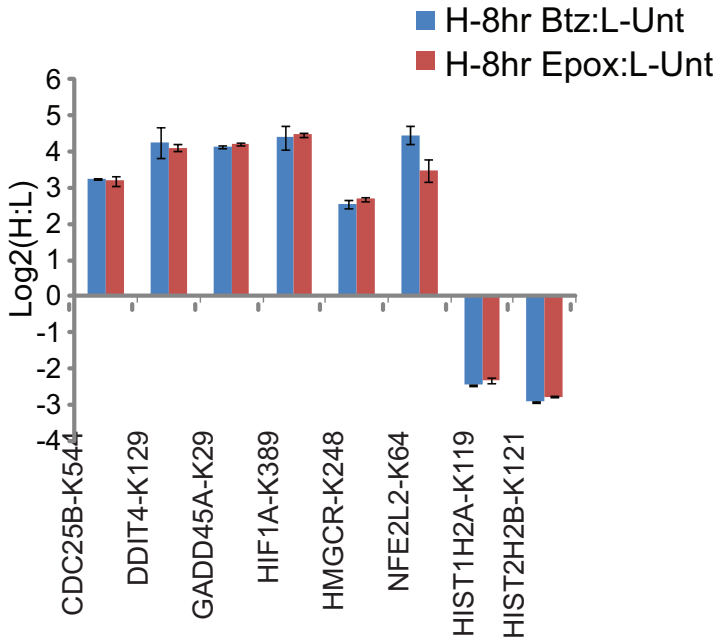
Substrate	E3	Reference
NFE2L2 (NRF2)	CRL3-KEAP1	reviewed in: Taguchi et al (2011) Genes Cells 16, 123-140.
CCND1 (Cyclin D1)	SCF-FBX4	Lin et al (2006) Mol Cell 24, 355-366.
HIF1⍰	CRL2-VHL	Huang et al (1998) Proc Natl Acad Sci USA 95, 7987-7992.
SET8	CRL4-CDT2	Abbas et al (2010) Mol Cell 40, 9-21. Centore et al (2010) Mol Cell 40, 22-33.
RNA Pol II	CRL3 Rsp5	Verma et al (2011) Mol Cell 41, 82-92. Somesh et al (2005) Cell 121, 913-923.
ATF4	SCF-⍰-TRCP	Lassot et al (2001) Mol Cell Biol. 21, 2192-202.

B

C



D



E

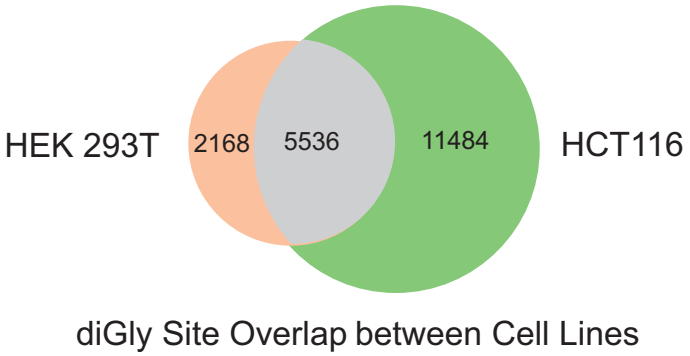


Figure S1 – Comparison of diGly site alterations in response to Btz or epoxomicin and examination of diGly-site identification overlap between 293T and HCT116 cells related to figure 1.

A) Known ubiquitylation substrates examined in this study. Shown are the E3s for the specific targets as well as primary references. B) The overlap of all quantified sites from four biological replicates of 8hr Btz treatment. C) The log2 ratio of each ubiquitin diGly containing peptide in response to 8 hr Btz (blue bar) or epoxomicin (red bar) treatment. The diGly modified lysine position is indicated below the y-axis. D) The log2 ratio of a subset of known ubiquitin modified proteins in response to 8 hr Btz (blue bar) or epoxomicin (red bar) treatment. Error bars represent the SEM of multiple MS1 quantifications for the indicated peptide.
E) Overlap of identified sites from all quantified sites from HCT116 cells with sites identified sites from HEK 293T cells treated with Btz for 8hrs.
All error bars represent the SEM of multiple MS1 quantifications for the indicated site.

Figure S2

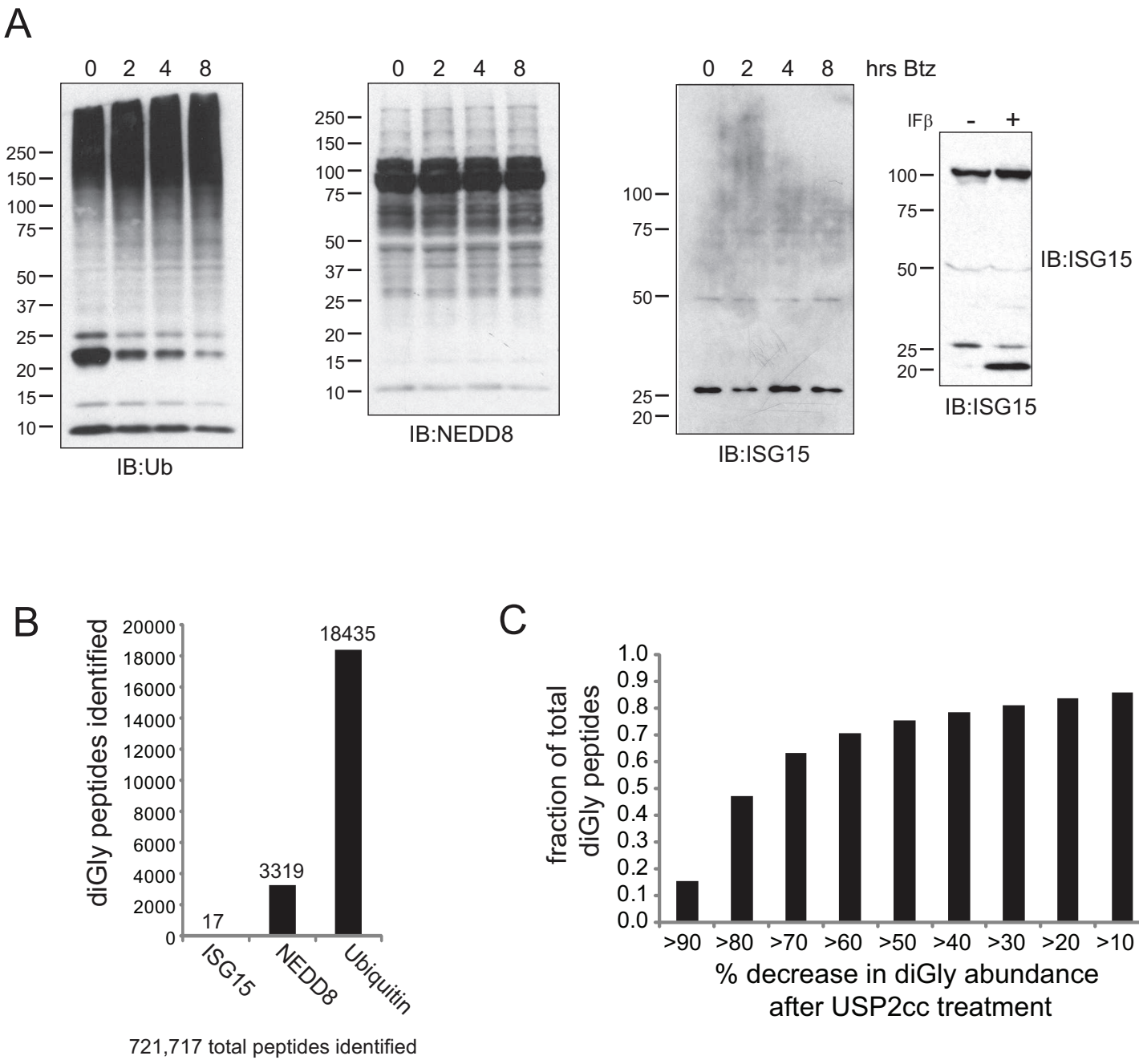


Figure S2 – Estimation of ISG15 contribution to observed diGly sites related to figure 2.
A) Whole cell lysates from HCT116 cells treated with Btz for 0, 2, 4, or 8 hours were immunoblotted with the indicated antibody. HCT116 cells were treated with interferon-β for 24 hrs to stimulate ISG15 expression as a positive control. B) Total number of observed diGly-modified peptides that correspond to ISG15, NEDD8, or Ubiquitin from all datasets in this study.C) Histogram of the fraction of quantified diGly sites decreased greater than the indicated percentage after treatment with USP2cc.

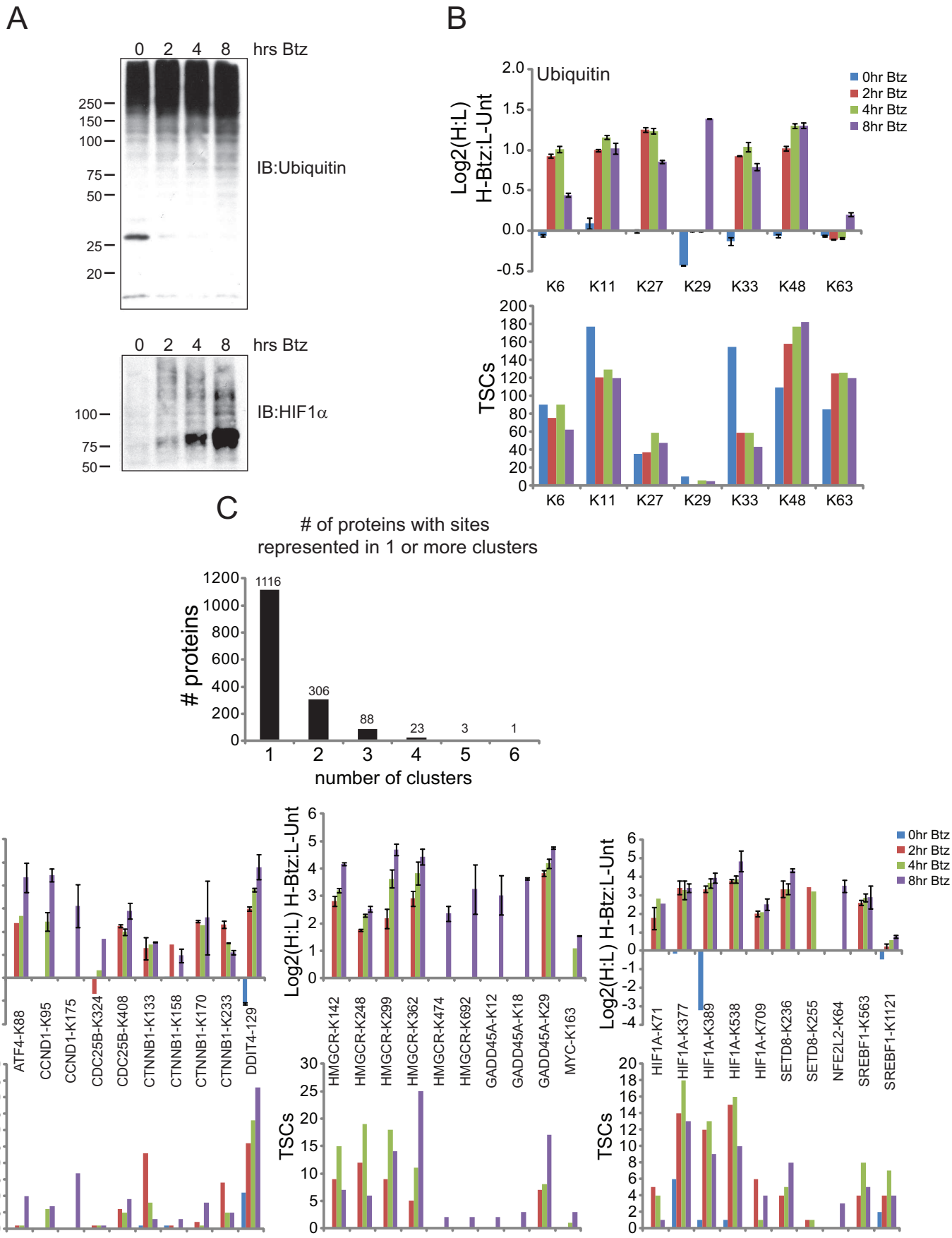
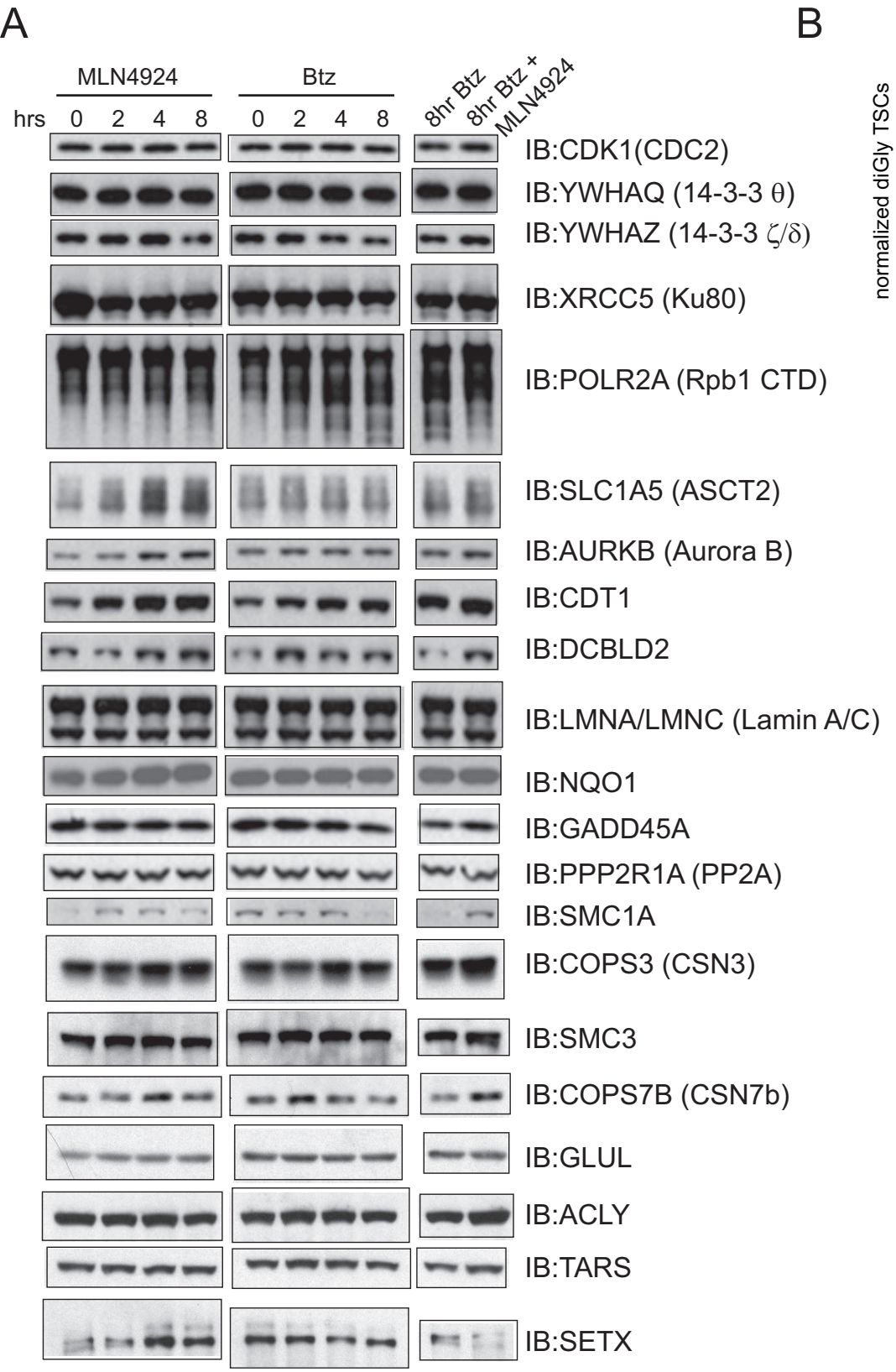


Figure S3 – Known unstable proteins increase diGly-modified peptide abundance over a time course of Btz treatment related to figure 3.
A) Cells were either untreated (light media) or treated with Btz for 0, 2, 4, or 8 hours (heavy media). Shown is the distribution of log2 ratios for all quantified peptides with 0 (blue bars), 2 (red bars), 4 (green bars), and 8 (purple bars) hours of bortezomib treatment. B) The log2 ratio (top) and total spectral counts (TSCs) (bottom) of each ubiquitin diGly peptide in untreated cells (blue bars) or cells treated with Btz for 2 (red bars), 4 (green bars), or 8 (purple bars) hours. Error bars represent the SEM of multiple MS1 quantifications for the indicated peptide. C) The distribution of proteins with a diGly-modified peptide in one or more groups as indicated in Figure 2. The number above each bar indicates the total number of proteins with sites that cluster with a single group or multiple clusters.D) The log2 ratio (top) and total spectral counts (bottom) of known unstable proteins in untreated cells (blue bars) or cells treated with Btz for 2 (red bars), 4 (green bars), or 8 (purple bars) hours. The protein corresponding to each peptide and the position of the modified lysine is indicated. Error bars represent the SEM of multiple MS1 quantifications for the indicated peptide.



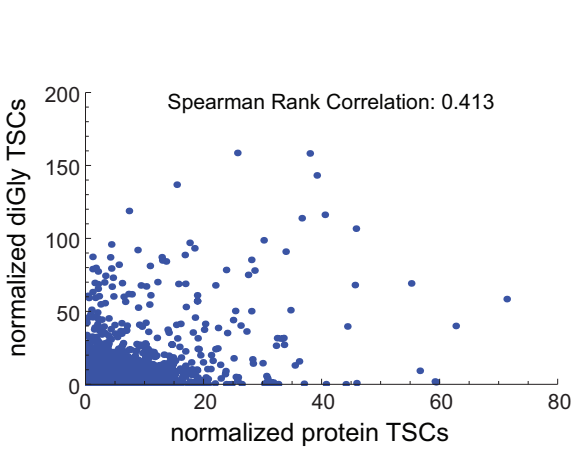


Figure S4 – Western blot validation of protein level alterations.
A) Cells were treated as with Btz or MLN4924 for 0,2,4,or 8 hrs, or treated with both Btz and MLN4924 for 8 hrs. Whole cell extracts were immunoblotted with the indicated antibody.
B) Scatter plot of normalized TSCs for diGly vs protein comparing diGly modifications mapped to the protein of comparison.

Figure S5

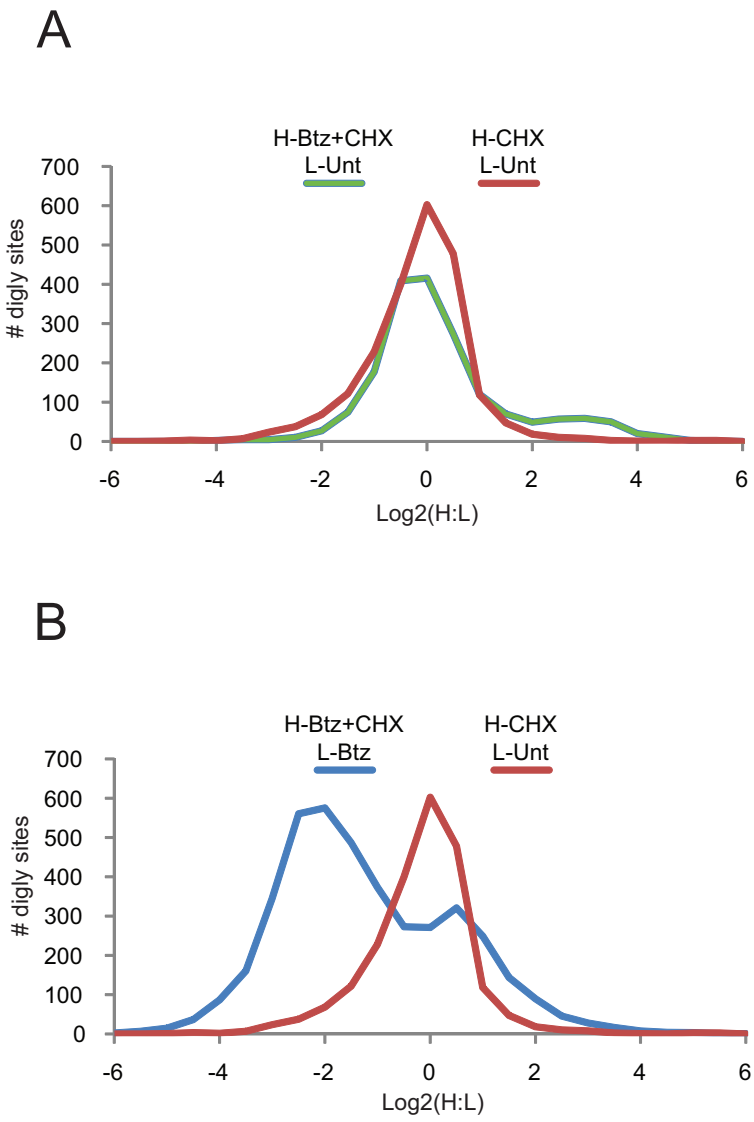
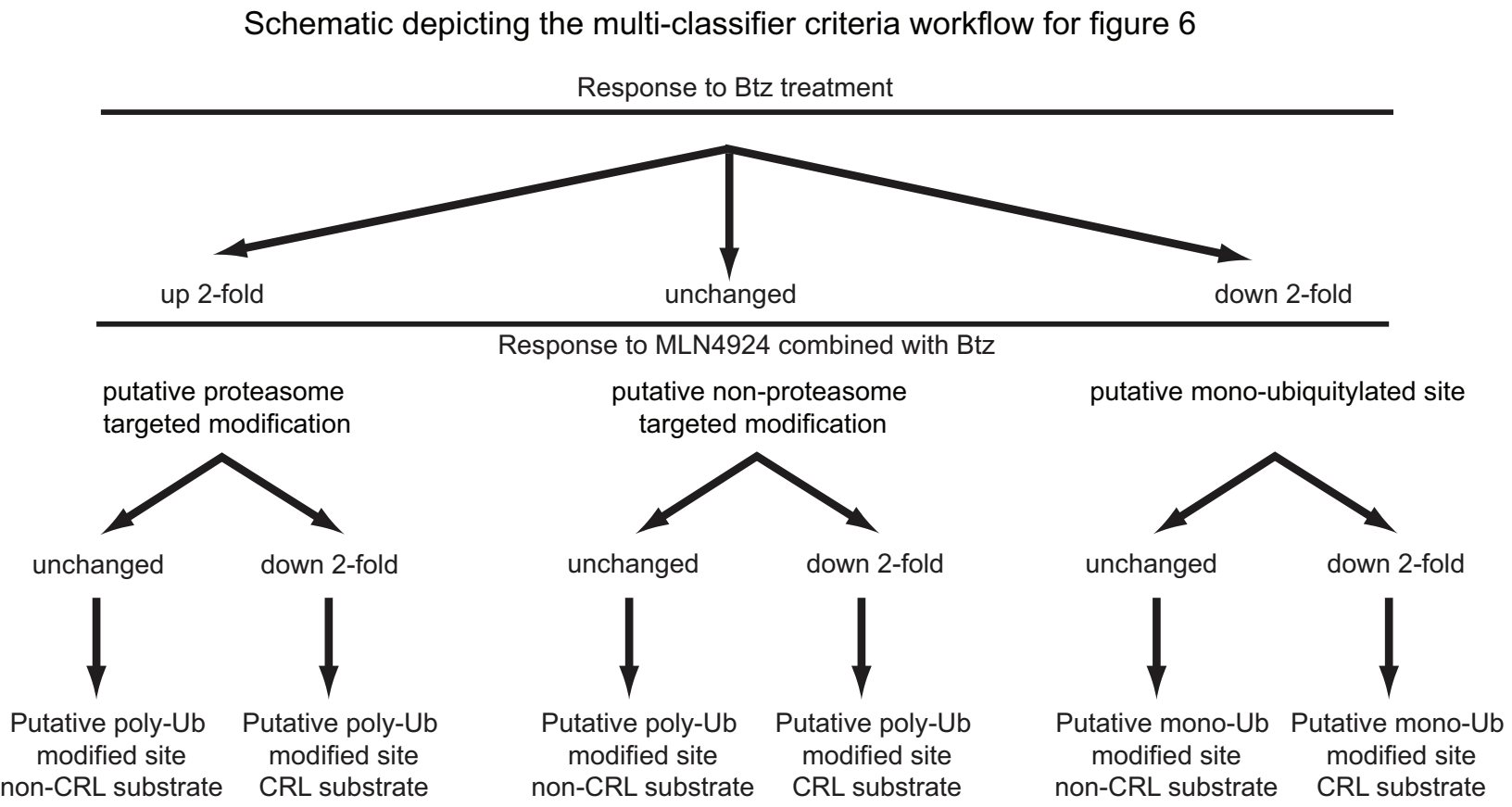


Figure S5 – Comparison of log2 ratios from cells treated with Btz and CHX or with CHX alone related to figure 5.
A) Distribution of log2 abundance ratios from K8 labeled cells treated with Btz and CHX for 8hrs (green line) or with CHX alone (red line).
B) Log2 ratios from unlabeled, Btz treated cells mixed with K8 labeled cells treated with Btz and CHX (blue line) compared to 8hr CHX treatment alone (red line).

Figure S6

A



B

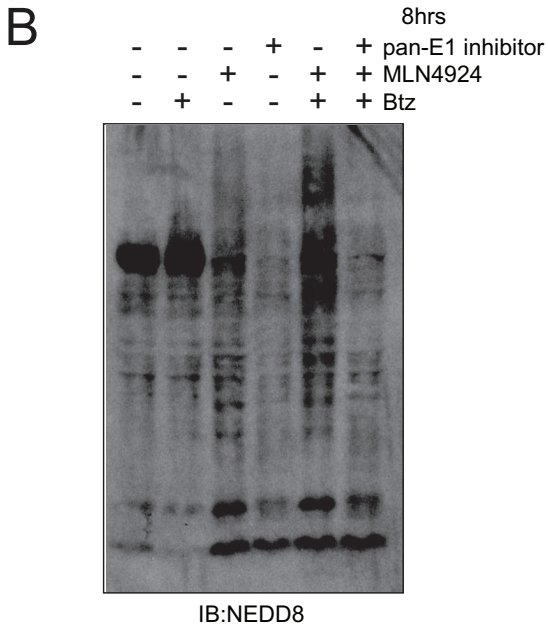


Figure S6 – Schematic of multi-classifer criteria and demonstration of Ub E1 dependent neddylation related to figure 6.
A) Schematic of muliti-classifer criteria used to categorize diGly-modified sites.
B) HCT116 cells were treated with Btz, MLN4924, and a pan-E1 inhibitor as indicated. Whole cell lysates were probed with α-NEDD8 antibodies

Figure S7

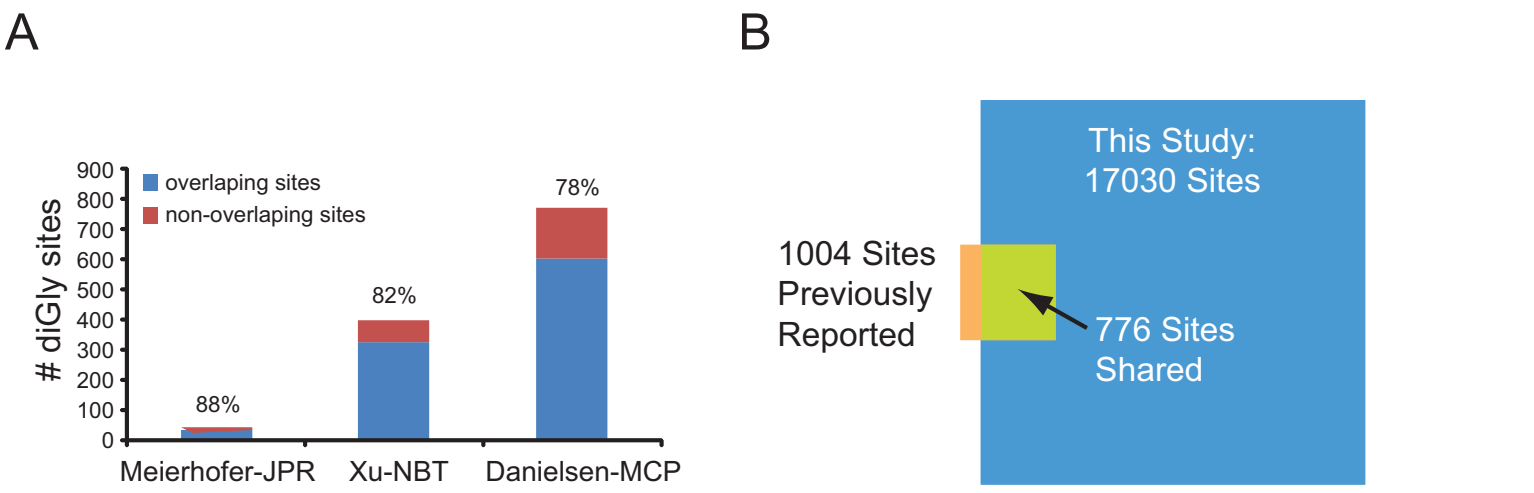


Figure S7 – Comparison of diGly sites identified within this study and previous studies related to figure 7.
A) Comparison of diGly-modified peptides observed in three previous studies (Danielsen et al., 2011; Meierhofer et al., 2008; Xu et al., 2010) to our dataset with sites in common (blue) and sites we did not observe (red). The percentage overlap is indicated.
B) Total number of previously identified diGly-modified lysines from the studies listed in A, compared to the sites identified in this study (sites identified in 293T cells are excluded). Boxes are drawn to scale
C) Distribution of log2 abundance ratios from K8 labeled cells treated with Btz and CHX for 8hrs (green line) or with CHX alone (red line). Log2 ratios from unlabeled, Btz treated cells mixed with K8 labeled cells treated with Btz and CHX (blue line) compared to 8hr CHX treatment alone (red line).