proteiNorm – A user-friendly tool for normalization and analysis of TMT and label-free protein quantification

*Stefan Graw [1,2], Jillian Tang [1], Maroof K Zafar [1], Alicia K Byrd [1], Chris Bolden[3], Eric C. Peterson[3], Stephanie D Byrum[1,2,]**

1 Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, 72205

2 Arkansas Children's Research Institute, Little Rock, Arkansas, 72202

3 Department of Pharmacology and Toxicology, University of Arkansas for Medical Sciences, Little Rock, Arkansas, 72205

* Corresponding author

S1. Mouse data set

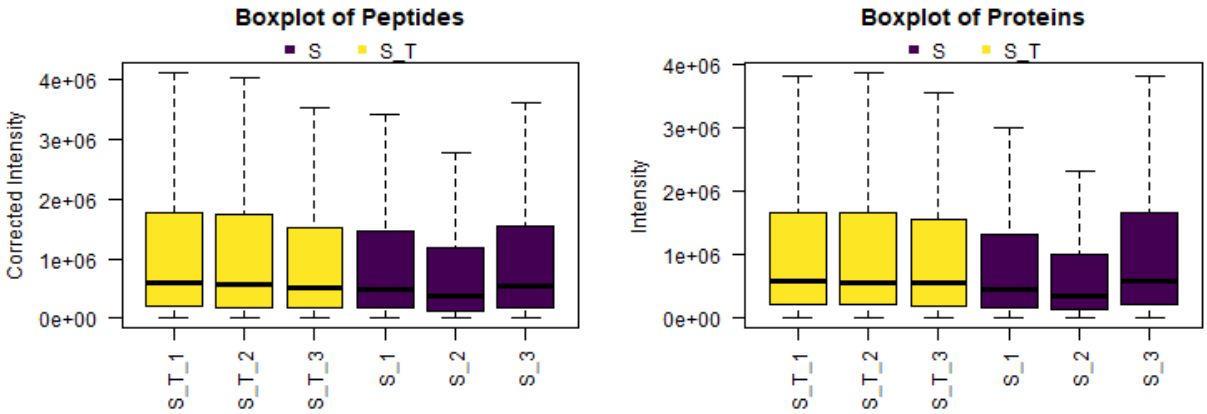

Figure S1 Intensity distribution (left: peptides; right: proteins)

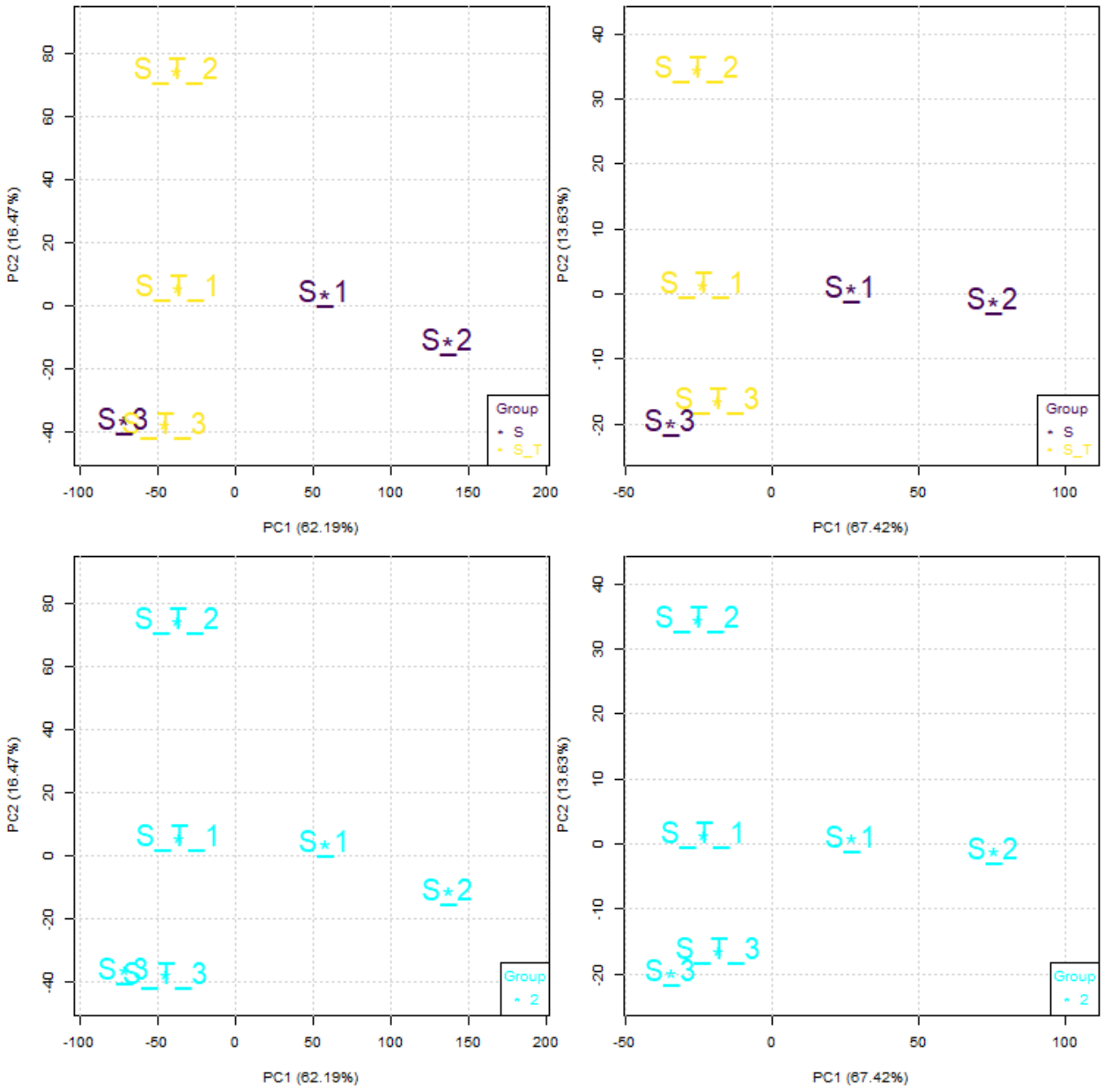

Figure S2 PCA plots before normalization (left: peptides; right: protein; top: colored by treatment; bottom: colored by batch)

Figure S3 Total intensity of remaining normalization methods

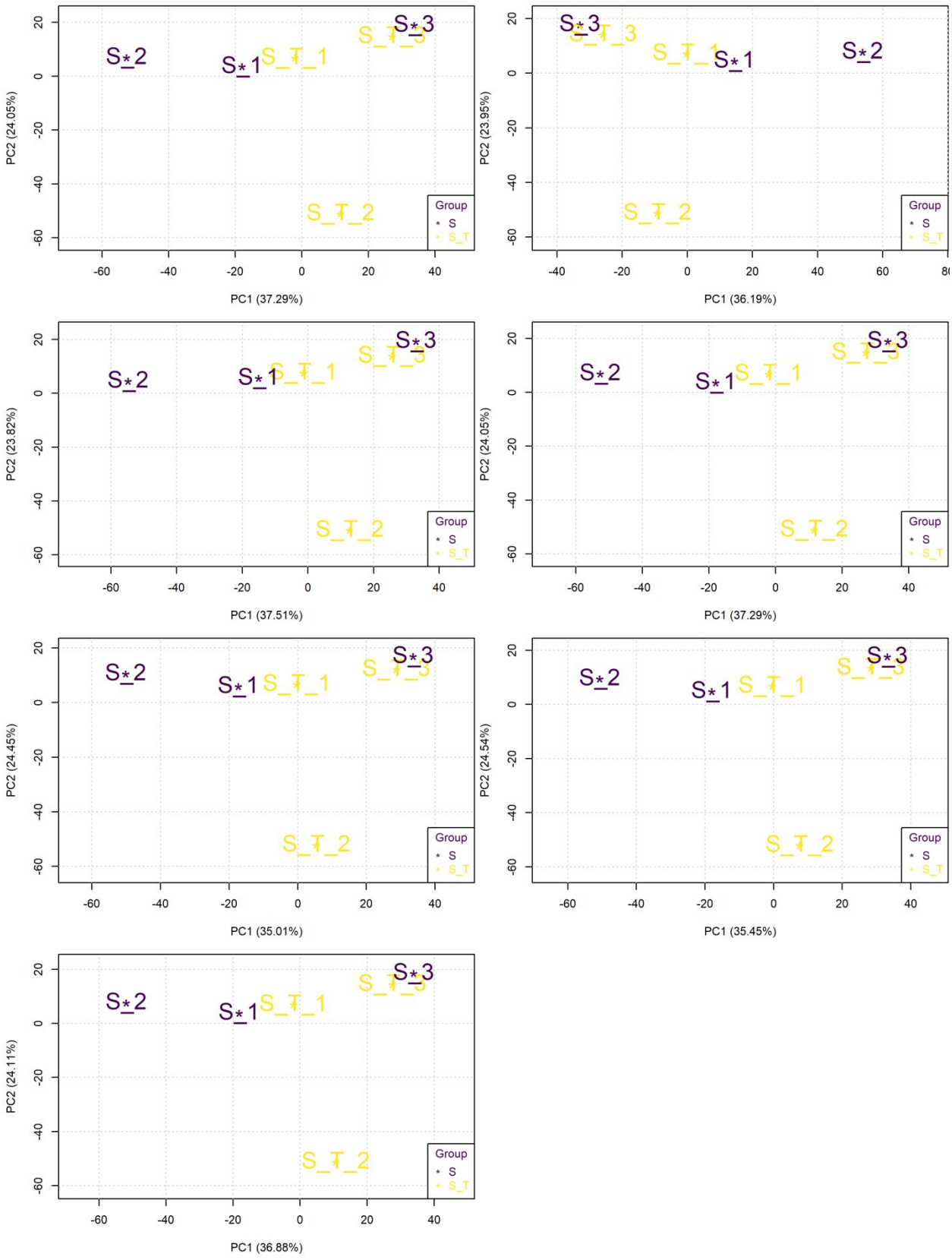

Figure S4 PCA plots of remaining normalization methods

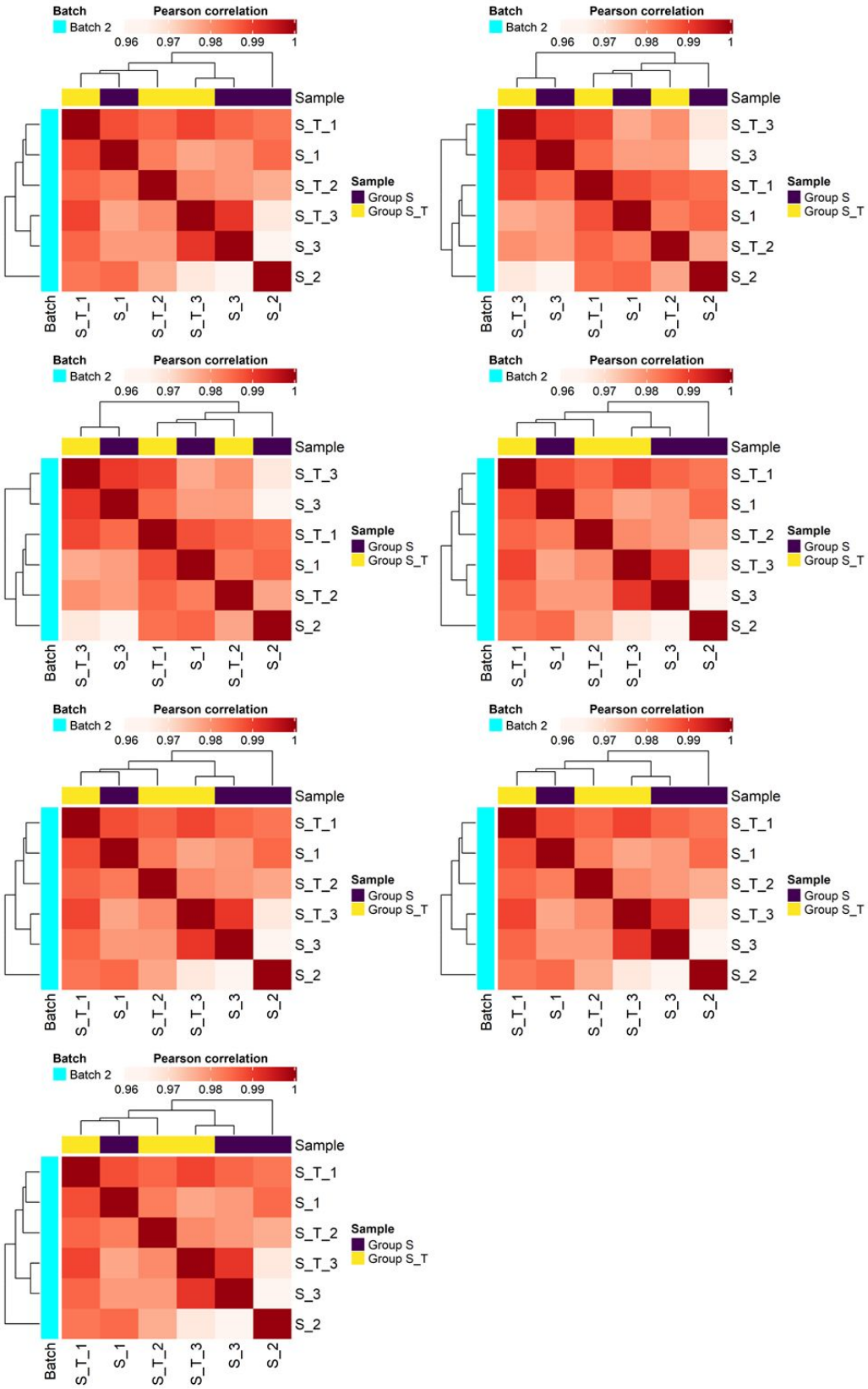

Figure S5 Correlation heatmaps of remaining normalization methods

S2. Breast cancer cell lines

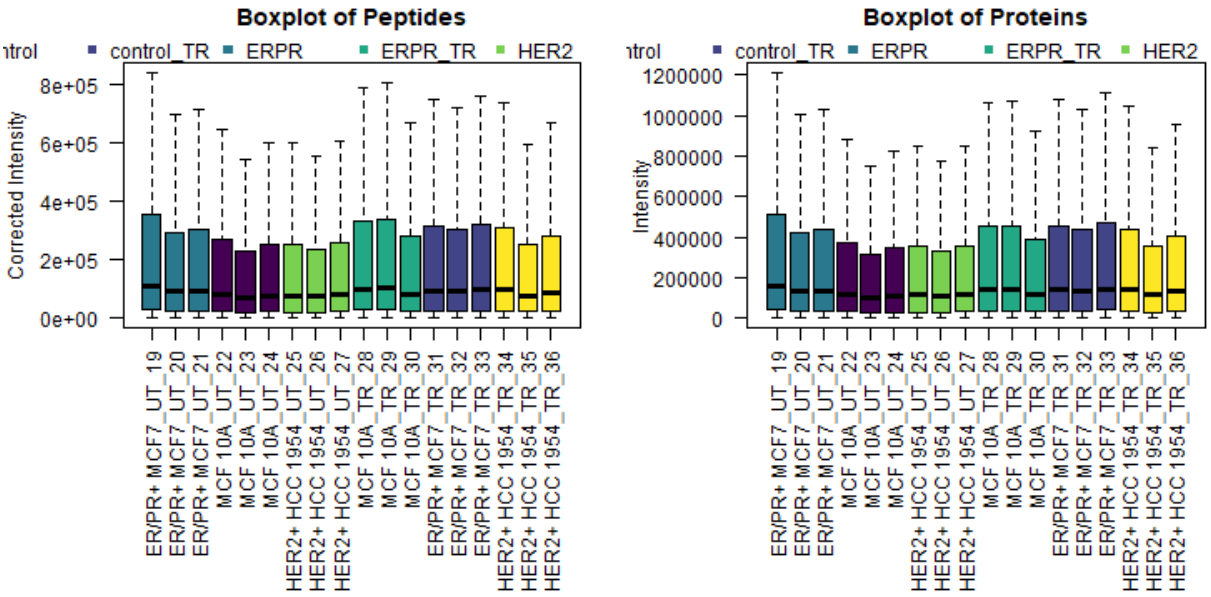

Figure S6 Intensity distribution (left: peptides; right: proteins)

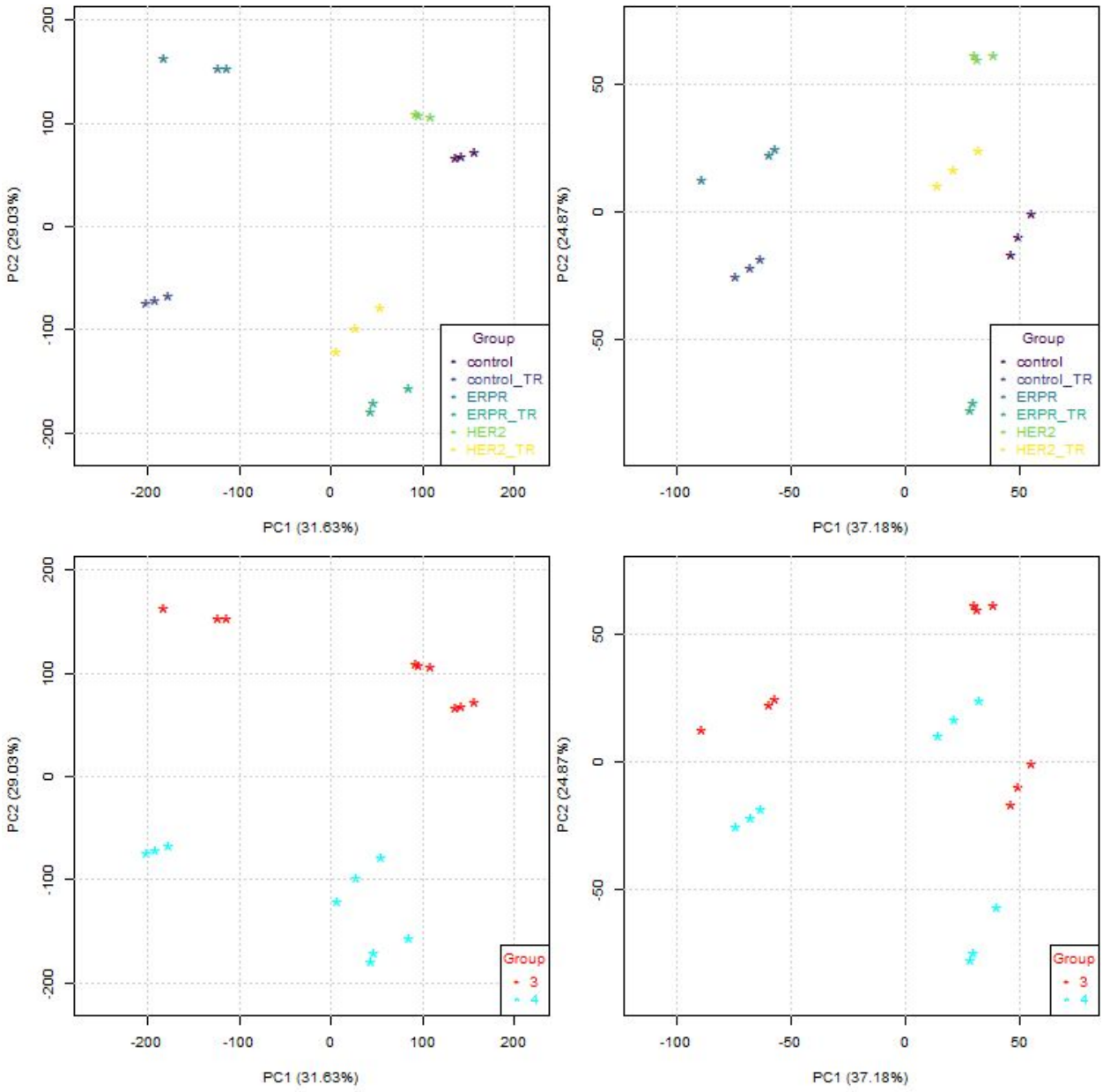

Figure S7 PCA plots before normalization (left: peptides; right: protein; top: colored by treatment; bottom: colored by batch)

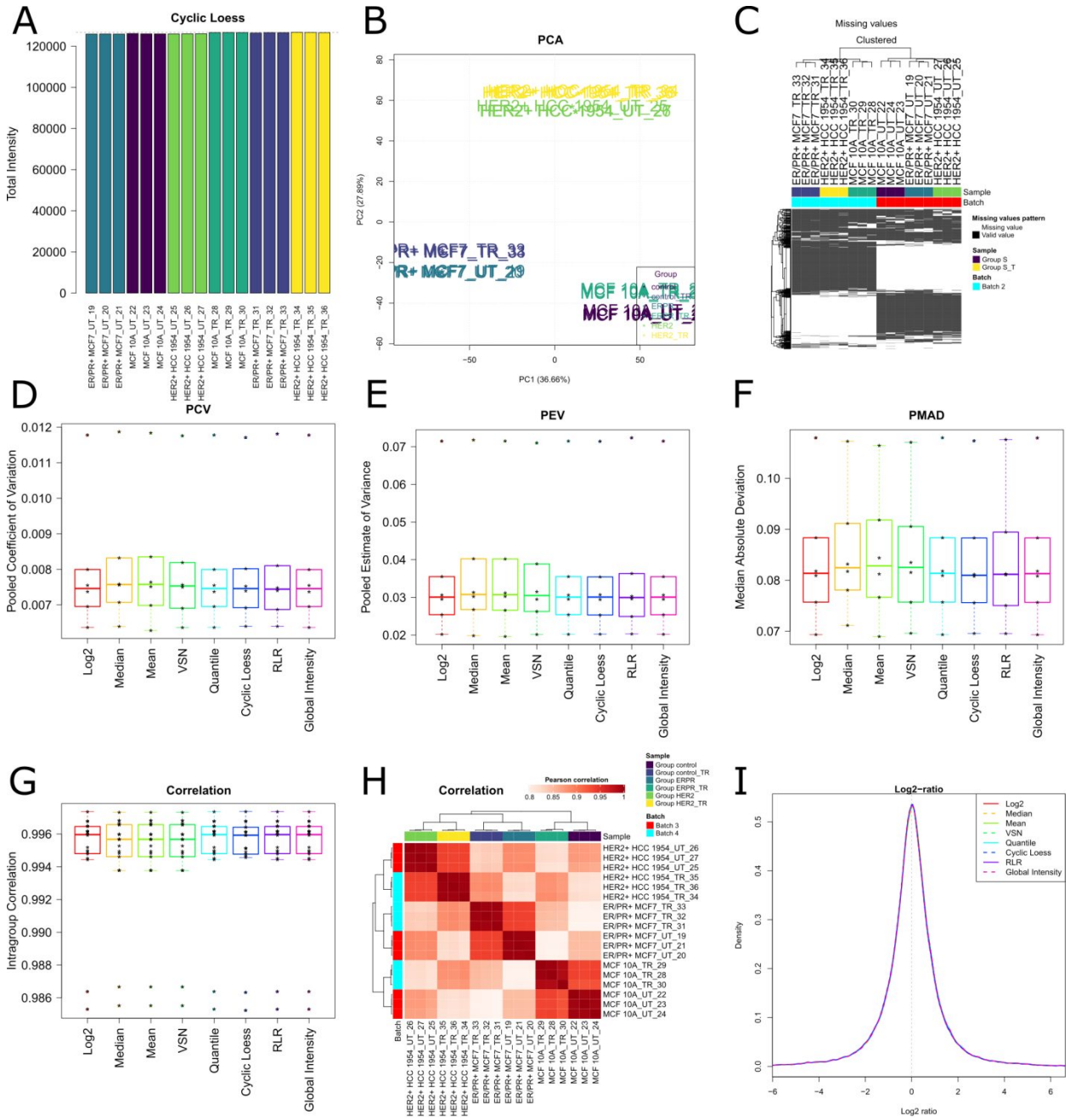

Figure S8 Evaluation of normalization and missing values. (A) Sum of normalized intensities using VSN by sample. (B) Principal component analysis plot based on data normalized by VSN. (C) Sample-clustered heatmap of missing data. (D) Pooled intragroup coefficient of variance comparing different normalization methods. (E) Pooled intragroup estimate of variance comparing different normalization methods. (F) Pooled intragroup median absolute deviation comparing different normalization methods. (G) Pair-wise sample correlations within a group for different normalization methods. (H) Correlation heatmap (all pair-wise samples). (I) Distribution of log2-ratios (all two-group combinations) for different normalization methods.

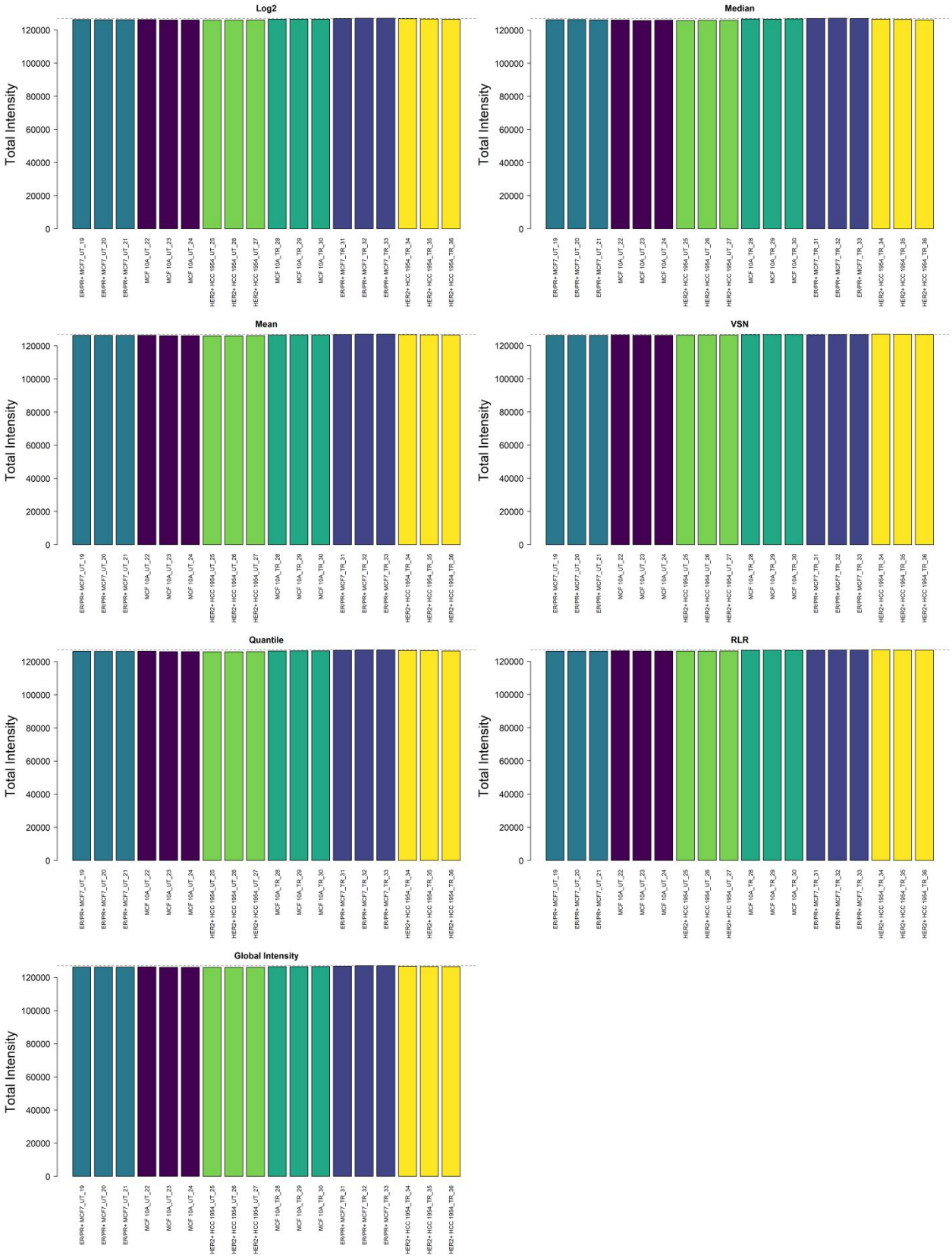

Figure S9 Total intensity of remaining normalization methods

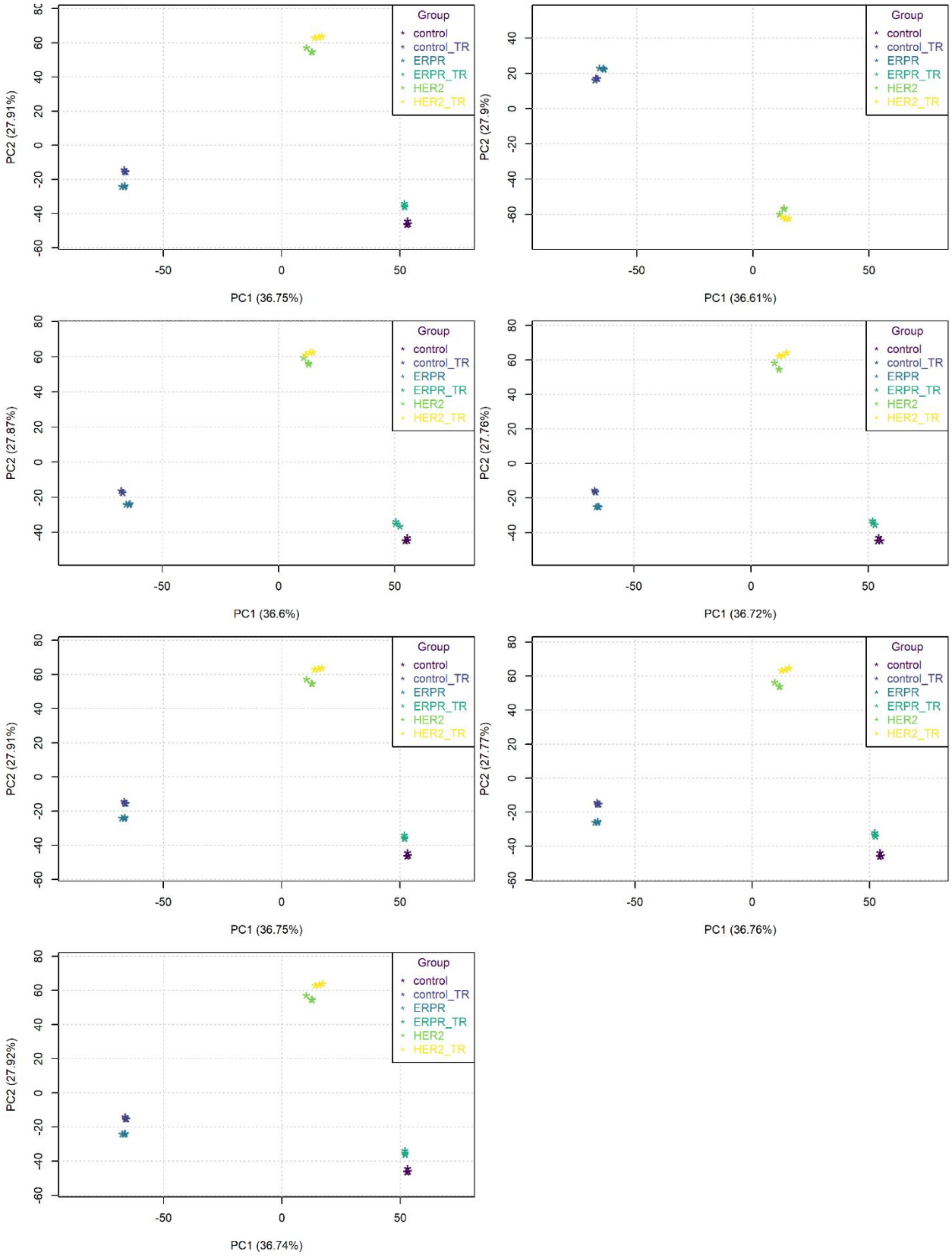

Figure S10 PCA plots of remaining normalization methods

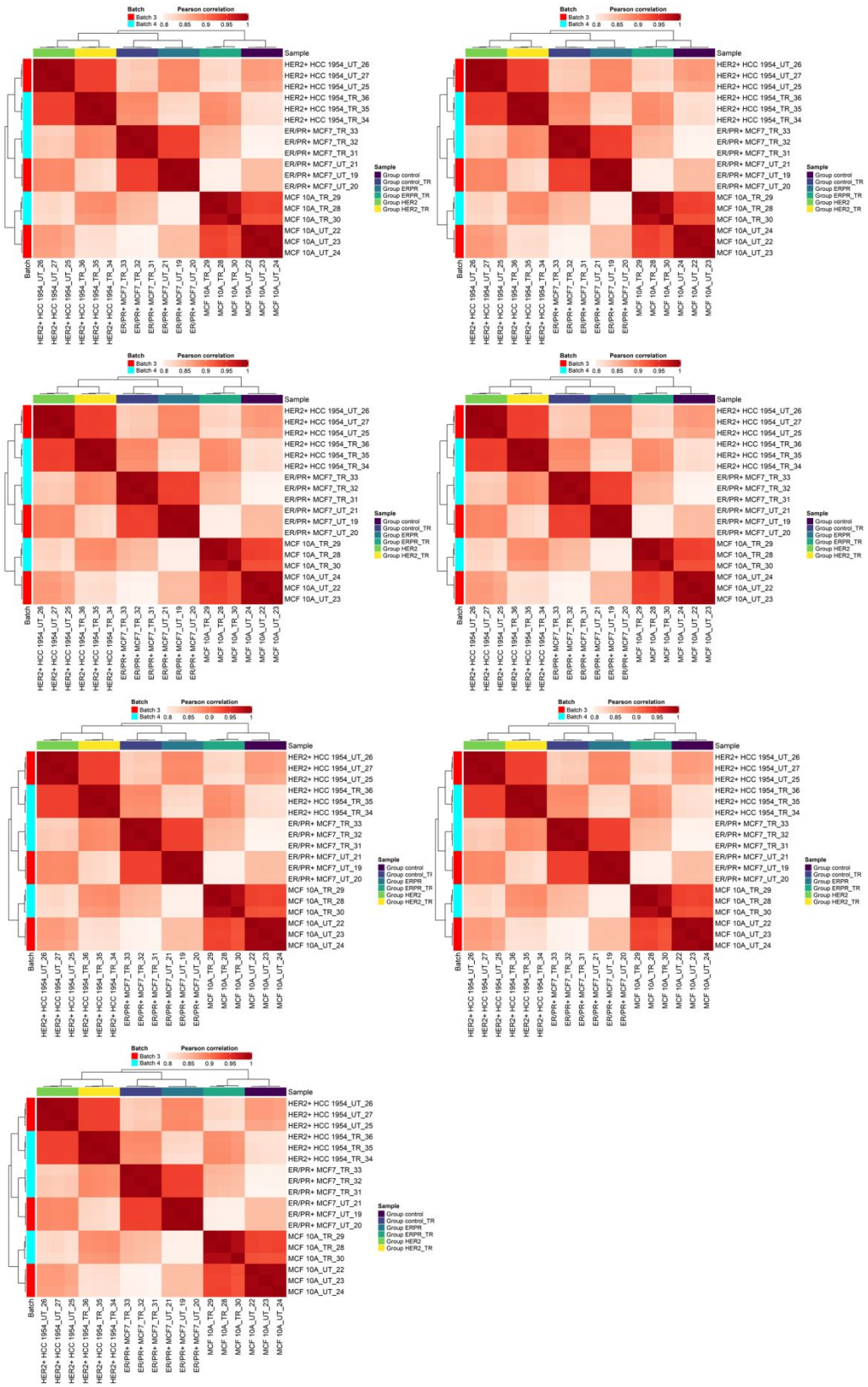

Figure S11 Correlation heatmaps of remaining normalization methods

S3. Spiked-in data

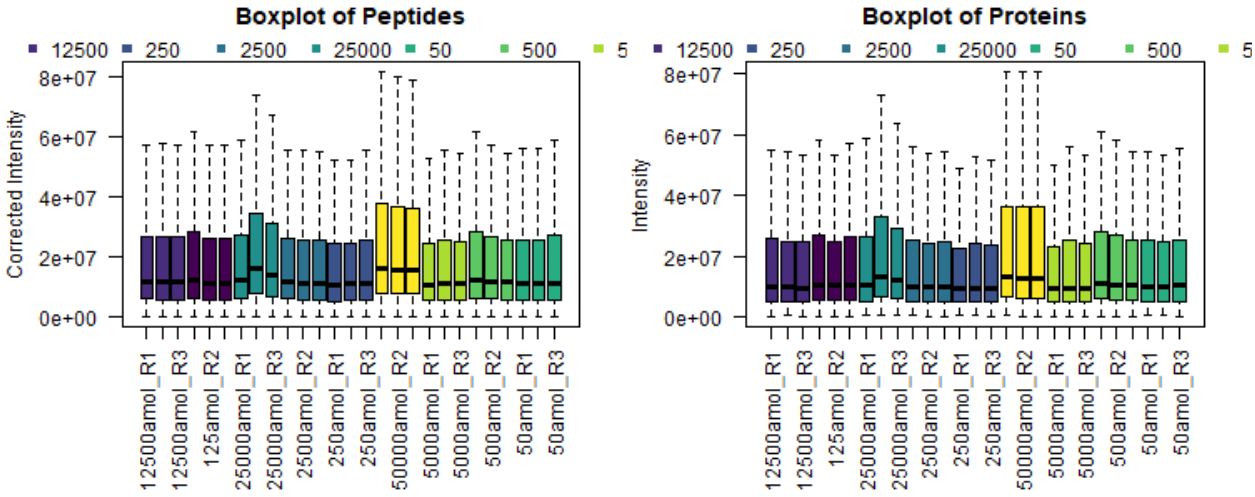

Figure S12 Intensity distribution (left: peptides; right: proteins)

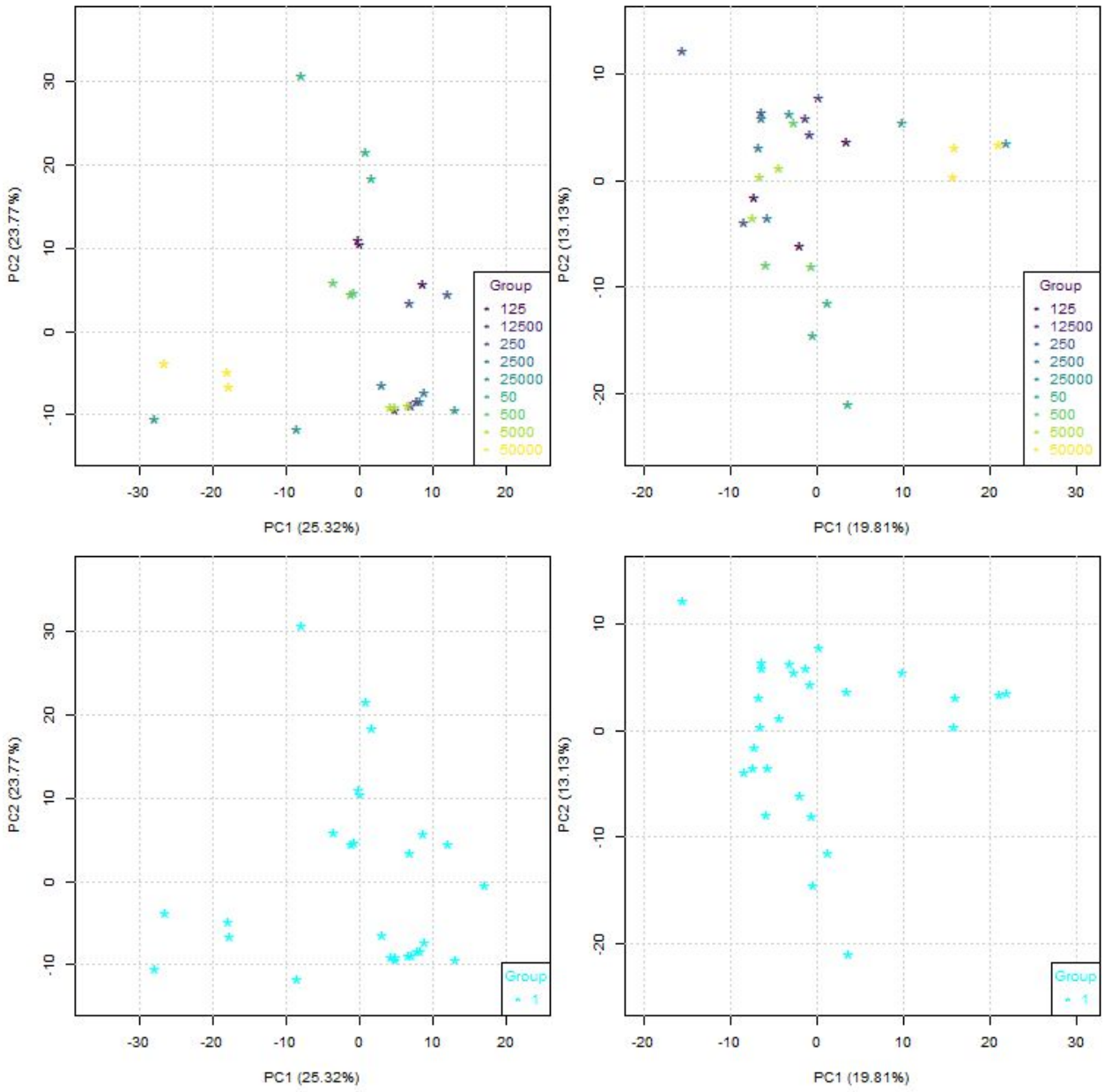

Figure S13 PCA plots before normalization (left: peptides; right: protein; top: colored by treatment; bottom: colored by batch)

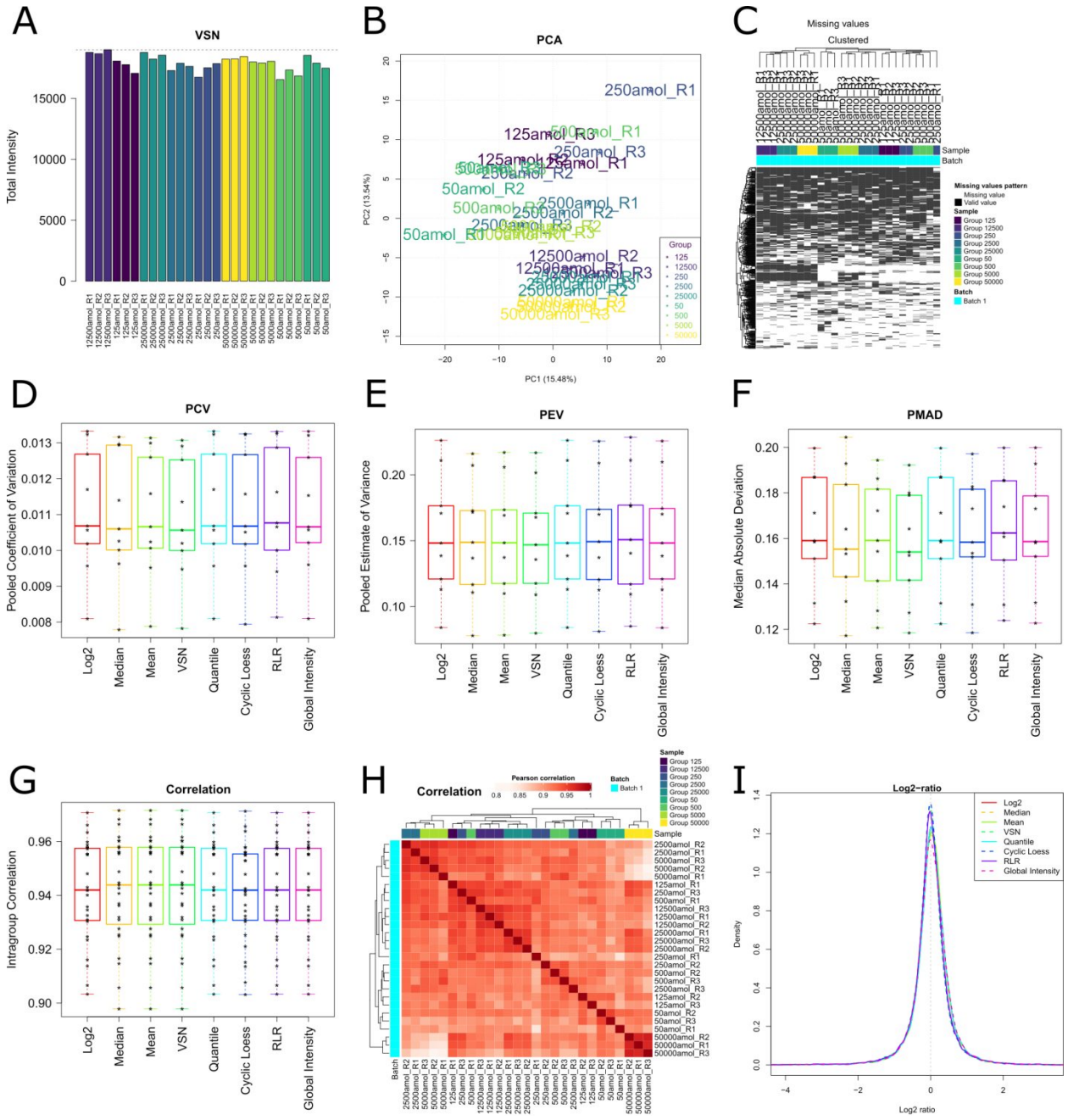

Figure S14 Evaluation of normalization and missing values. (A) Sum of normalized intensities using VSN by sample. (B) Principal component analysis plot based on data normalized by VSN. (C) Sample-clustered heatmap of missing data. (D) Pooled intragroup coefficient of variance comparing different normalization methods. (E) Pooled intragroup estimate of variance comparing different normalization methods. (F) Pooled intragroup median absolute deviation comparing different normalization methods. (G) Pair-wise sample correlations within a group for different normalization methods. (H) Correlation heatmap (all pair-wise samples). (I) Distribution of log2-ratios (all two-group combinations) for different normalization methods.

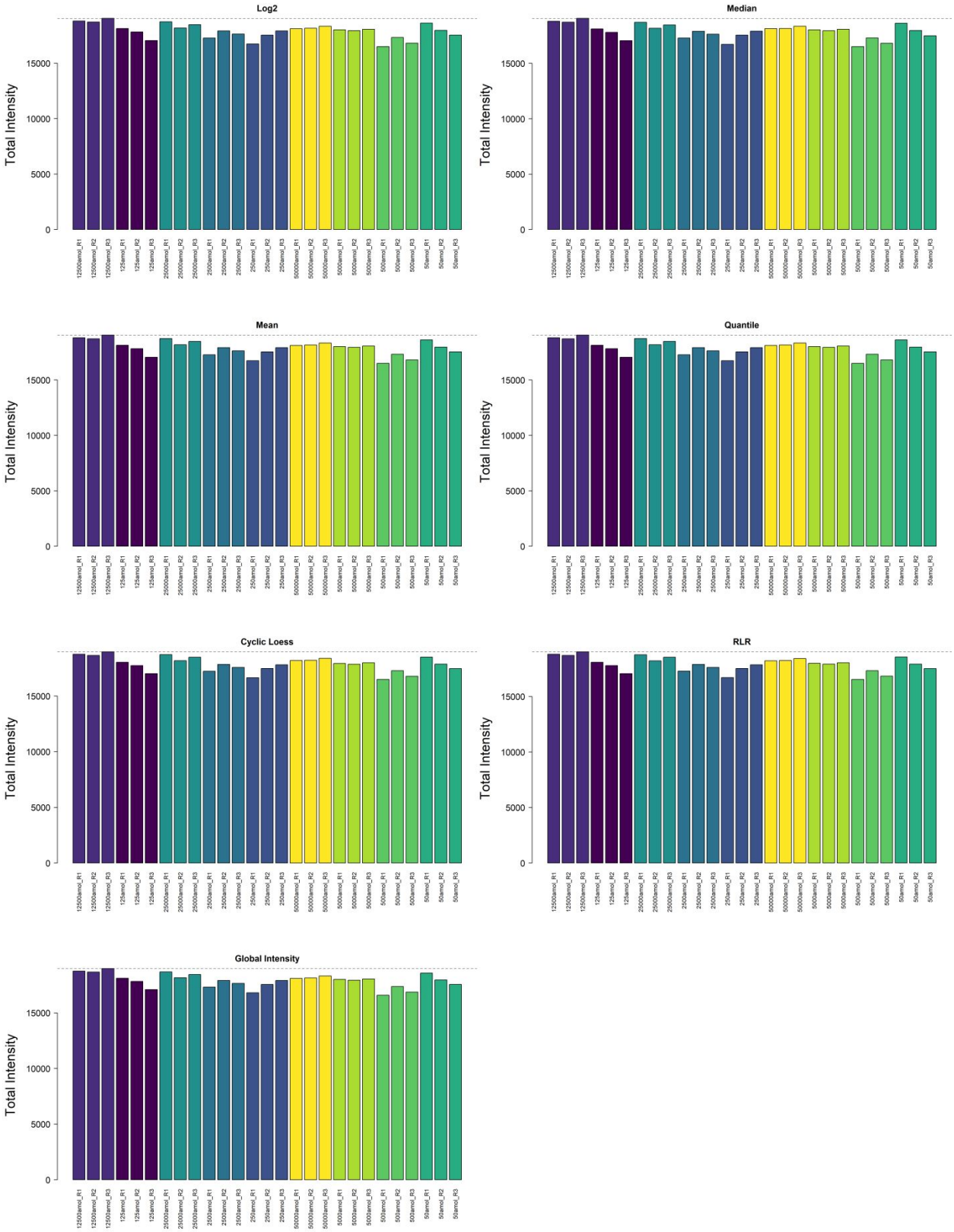

Figure S15 Total intensity of remaining normalization methods

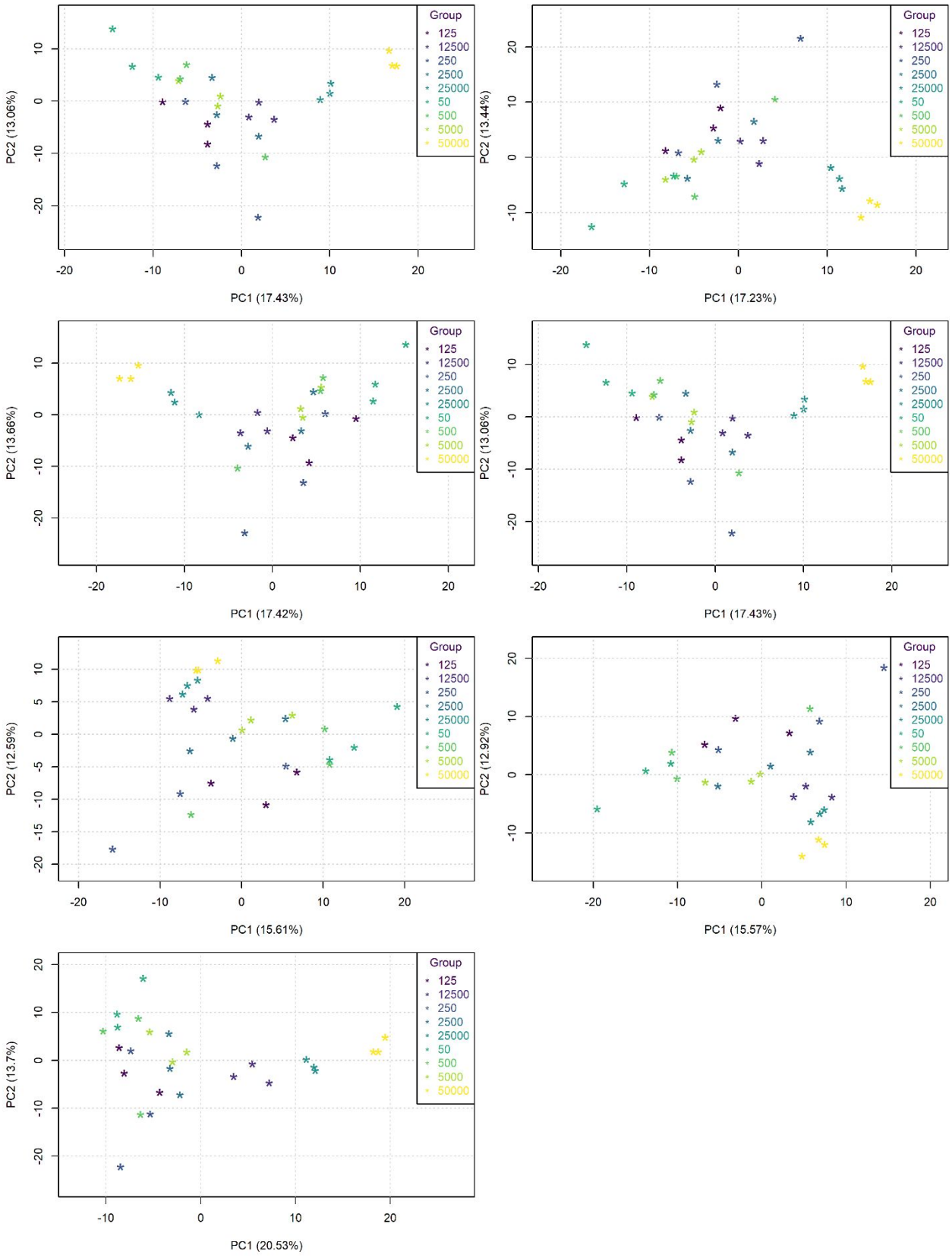

Figure S16 PCA plots of remaining normalization methods

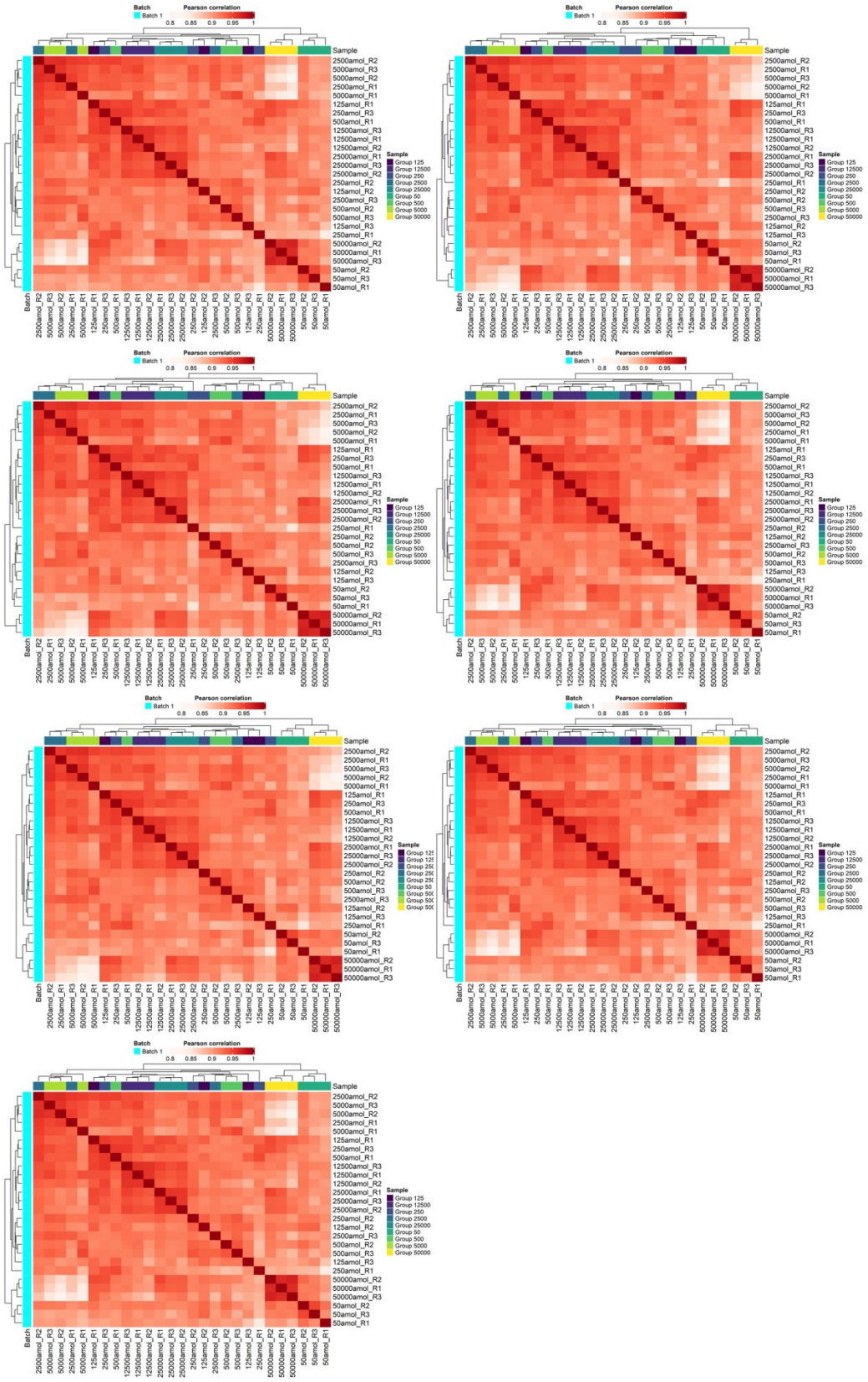

Figure S17 Correlation heatmaps of remaining normalization methods

Supplemental Methods

S1. Mouse data set

Animals

All protocols and experiments were performed in accordance with and approved by the University of Arkansas for Medical Sciences Institutional Animal Care and Use Committee. Adult wild type male and female C57BL/6 mice were purchased from Charles River Laboratories (Raleigh, NC) and housed in groups of 4–5 in a light-controlled environment (12 hr light/dark cycle). They received water and food pellets ad libitum, maintaining body weights averaging 26g. Mice were randomly selected for all experimental groups and divided among the studies. Treatment groups involved in the studies are as follows: saline-saline and methamphetamine (meth) (n=10/per group). All mice received intraperitoneal (ip) injections of either 1.5 mg/kg meth or saline every two hours for 4 consecutive injections. Mice were anesthetized with isoflurane and sacrificed 18 hour post the last injection

Drugs and Reagents

Methamphetamine was obtained from the National Institute of Drug Abuse (Bethesda, MD). Chemicals and other reagents were purchased from Fisher Scientific (Pittsburgh, PA) or ThermoFisher Scientific (Waltham, MA) unless otherwise stated.

Brain Dissections for Neuron Isolation

For these studies we isolated the nucleus accumbens (NAc) and dorsal striatum (S, containing the caudate and putamen regions) regions of the mouse brain. The dorsal striatum (S) and NAc

are enriched in dopaminergic neurons and are central in the reward pathways associated with psychostimulant SUDs (Chang et al. 2007).

After decapitation by guillotine, brains were removed from the base of the skull and dissected on ice. The S and NAc of each were extracted via freehand dissection and immediately placed in 1x phosphate buffered saline (PBS). The washed brain regions were dounce homogenized in radioimmunoprecipitation assay buffer (RIPA) lysis buffer (10 mM Tris-Cl (pH 8.0), 1 mM Ethylenediaminetetraacetic acid (EDTA), 0.5 mM Ethylene glycol tetraacetic acid (EGTA), 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% sodium dodecyl sulfate (SDS), and 140 mM NaCl) 5 times. The samples were incubated on ice for 30 min, spun at 4°C for 10 min at max speed, and the supernatant was transferred to a new tube for the whole lysate total protein analysis.

Mass Spectrometry of Total Protein

Purified proteins were reduced, alkylated, and digested using filter-aided sample preparation [17]. Tryptic peptides were labeled using a tandem mass tag 6-plex isobaric label reagent set (Thermo) following the manufacturer's instructions. Labeled peptides were separated into 36 fractions on a 100 x 1.0 mm Acquity BEH C18 column (Waters) using an UltiMate 3000 UHPLC system (Thermo) with a 40 min gradient from 99:1 to 60:40 buffer A:B ratio (buffer A: 0.1% formic acid, 0.5% acetonitrile, buffer B: 0.1% formic acid, 99.9% acetonitrile) under basic pH conditions, and then consolidated into 12 super-fractions. Each super-fraction was then further separated by reverse phase XSelect CSH C18 2.5 um resin (Waters) on an in-line 120 x 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 60 min gradient from 98:2 to 67:33 buffer A:B ratio. (buffer A: 0.1% formic acid, 0.5%

acetonitrile, buffer B: 0.1% formic acid, 99.9% acetonitrile). Eluted peptides were ionized by electrospray (2.25 kV) followed by mass spectrometric analysis on an Orbitrap Fusion Lumos mass spectrometer (Thermo) using multi-notch MS3 parameters as described in McAlister et al. [18]. MS data were acquired using the FTMS analyzer in top-speed profile mode at a resolution of 120,000 over a range of 375 to 1500 m/z. Following CID activation with normalized collision energy of 35.0, MS/MS data were acquired using the ion trap analyzer in centroid mode and normal mass range. Using synchronous precursor selection, up to 10 MS/MS precursors were selected for HCD activation with normalized collision energy of 65.0, followed by acquisition of MS3 reporter ion data using the FTMS analyzer in profile mode at a resolution of 50,000 over a range of 100-500 m/z. Proteins were identified and reporter ions quantified using MaxQuant (Max Planck Institute) database search against *Mus musculus* (June 2018) with a parent ion tolerance of 3 ppm, a fragment ion tolerance of 0.5 Da, and a reporter ion tolerance of 0.01 Da.

S2. Breast Cancer Cell Lines

This example data set consists of three different breast cancer cell lines (MCF10A a non-tumorigenic epithelial cell line, MCF7 an ER/PR + cell line, HCC1954 a HER2+ cell line) with and without 5 mM hydroxyurea (HU) treatment for 4 hours. Three replicates for each cell-line-treatment combination were analyzed. The 18 samples were multiplexed using two Tandem Mass Tag (TMT) TMT-10plex isobaric tag batches such that untreated and treated cell lines assembled one batch each.

Proteins were reduced, alkylated, and purified by chloroform/methanol extraction prior to digestion with sequencing grade modified porcine trypsin (Promega). Tryptic peptides were labeled using tandem mass tag isobaric labeling reagents (Thermo) following the manufacturer's instructions and combined into one 10-plex sample group. A pool sample was made by pooling

equal amounts of all 18 samples into one mixture and labeled with the TMT-131 reporter ion. The same sample was run on both TMT batches to provide a reference sample across the TMT batch sequencing runs. The labeled peptide multiplex was separated into 36 fractions on a 100 x 1.0 mm Acquity BEH C18 column (Waters) using an UltiMate 3000 UHPLC system (Thermo) with a 40 min gradient from 99:1 to 60:40 buffer A:B ratio under basic pH conditions, and then consolidated into 18 super-fractions. Each super-fraction was then further separated by reverse phase XSelect CSH C18 2.5 um resin (Waters) on an in-line 150 x 0.075 mm column using an UltiMate 3000 RSLCnano system (Thermo). Peptides were eluted using a 60 min gradient from 97:3 to 60:40 buffer A:B ratio. Eluted peptides were ionized by electrospray (2.15 kV) followed by mass spectrometric analysis on an Orbitrap Eclipse Tribrid mass spectrometer (Thermo) using multi-notch MS3 parameters from the McAlister et al. [18]. MS data were acquired using the FTMS analyzer in top-speed profile mode at a resolution of 120,000 over a range of 375 to 1500 m/z. Following CID activation with normalized collision energy of 35.0, MS/MS data were acquired using the ion trap analyzer in centroid mode and normal mass range. Using synchronous precursor selection, up to 10 MS/MS precursors were selected for HCD activation with normalized collision energy of 65.0, followed by acquisition of MS3 reporter ion data using the FTMS analyzer in profile mode at a resolution of 50,000 over a range of 100-500 m/z.

Proteins were identified and TMT MS3 reporter ions intensities obtained using a MaxQuant (Max Planck Institute) search against the UniProtKB database (November 2018) restricted to *Homo sapiens* with a parent ion tolerance of 3 ppm, a fragment ion tolerance of 0.5 Da, and a reporter ion tolerance of 0.003 Da. Scaffold Q+S (Proteome Software) was used to verify MS/MS based peptide and protein identifications (protein identifications were accepted if they could be established with less than 1.0% false discovery and contained at least 2 identified

peptides; protein probabilities were assigned by the Protein Prophet algorithm [19]). The MaxQuant output files “ProteinGroups.txt” and “peptides.txt” were used as input files for proteiNorm.

S3. Spike-in data

The data set was published by Ramus’ et al. [16] and describes a yeast lysate that was spiked with different known amounts of the UPS1 mixture and analyzed with LC-MS. The data consists of nine different UPS1 concentrations, each with three replicates, and provides the “ground truth” with known signals. Proteins were identified using MaxQuant (Max Planck Institute) search against the UniProtKB database (May 2020) restricted to *Saccharomyces cerevisiae* and the ups1-ups2-sequences.fasta from sigma Aldrich (https://www.sigmaaldrich.com/life-science/proteomics/mass-spectrometry/ups1-and-ups2-proteomic.html) with a parent ion tolerance of 3 ppm, a fragment ion tolerance of 0.5 Da, and a reporter ion tolerance of 0.003 Da.