nature protocols

SUPPLEMENTARY INFORMATION

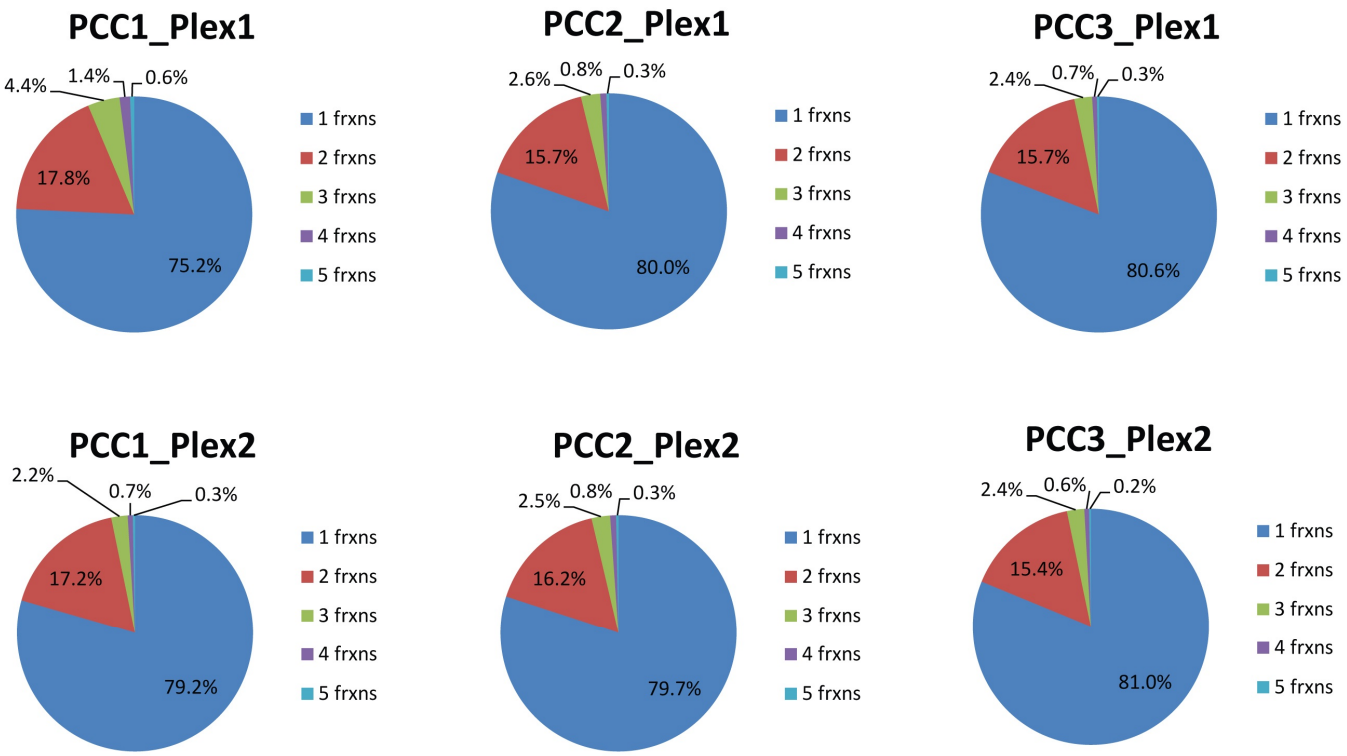
https://doi.org/10.1038/s41596-018-0006-9

In the format provided by the authors and unedited.

Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry

Philipp Mertins[1,2,6,7], Lauren C. Tang[1,7], Karsten Krug[1,7], David J. Clark[3,7], Marina A. Gritsenko[4,7], Lijun Chen[3], Karl R. Clauser[1], Therese R. Clauss[4], Punit Shah[3], Michael A. Gillette[1], Vladislav A. Petyuk[4], Stefani N. Thomas[3], D. R. Mani[1], Filip Mundt[1], Ronald J. Moore[4], Yingwei Hu[3], Rui Zhao[4], Michael Schnaubelt[3], Hasmik Keshishian[1], Matthew E. Monroe[4], Zhen Zhang[3], Namrata D. Udeshi[1], Deepak Mani[1], Sherri R. Davies[5], R. Reid Townsend[5], Daniel W. Chan[3], Richard D. Smith[4], Hui Zhang[3], Tao Liu[4] and Steven A. Carr[1*]

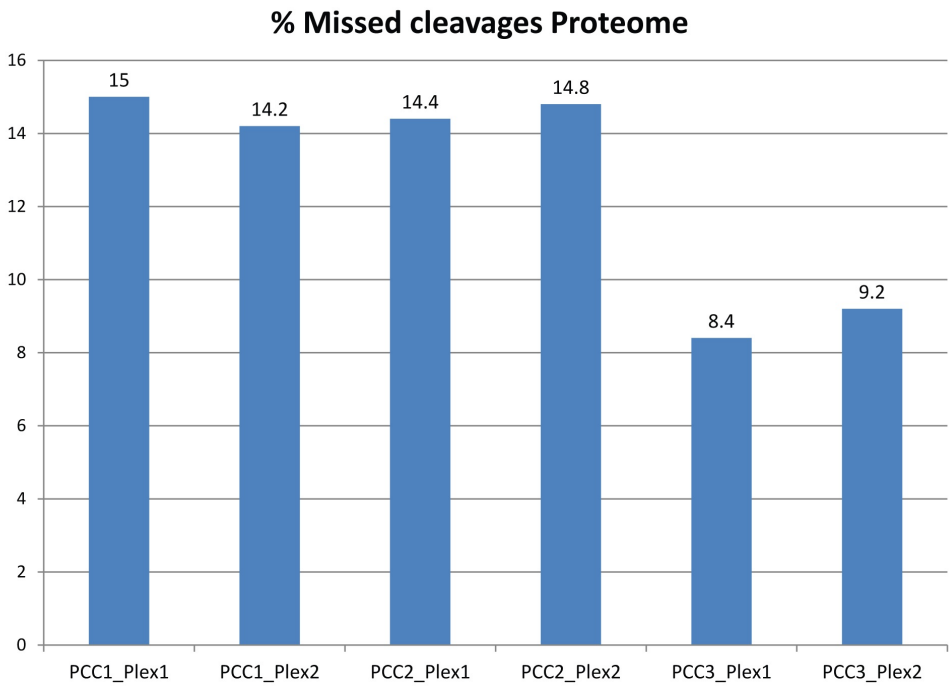
[1]Proteomics Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Berlin Institute of Health, Berlin, Germany. [3]Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD, USA. [4]Biological Sciences Division, Pacific Northwest National Laboratories, Richland, WA, USA. [5]Department of Medicine, Washington University School of Medicine, St. Louis, MO, USA. [6]Present address: Proteomics Platform, Max Delbrück Center for Molecular Medicine in the Helmholtz Society, Berlin, Germany. [7]These authors contributed equally: Philipp Mertins, Lauren C. Tang, Karsten Krug, David J. Clark, and Marina A. Gritsenko.. *e-mail: scarr@broad.mit.edu

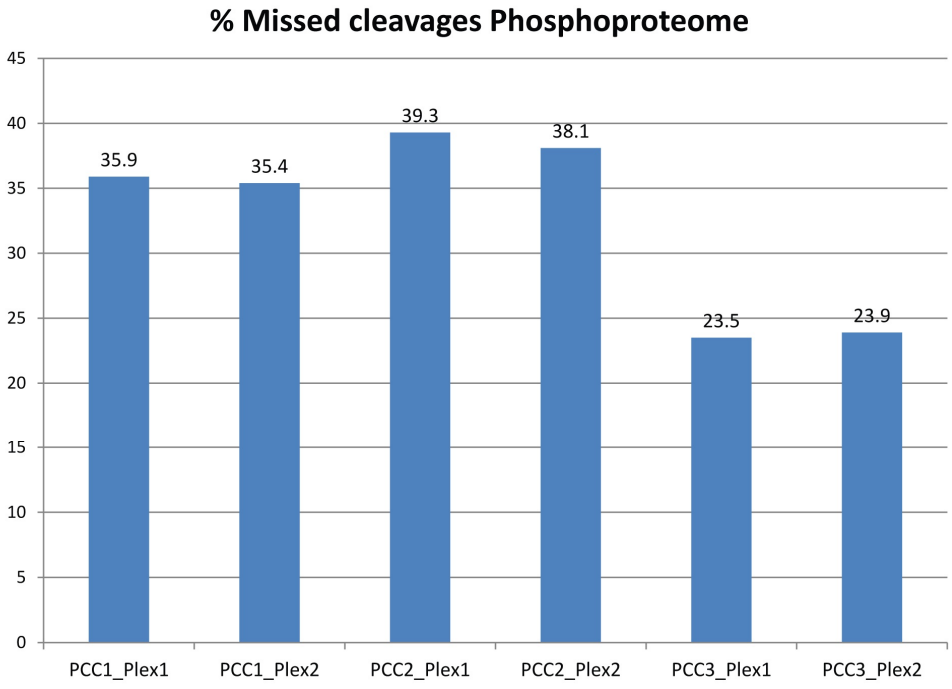


Supplementary Figure 1

Uniqueness per basic RP fraction for identified peptides.

The pie charts depict the distribution of the number of fractions a peptide has been detected per TMT plex which was highly reproducible across laboratories. More the 75% of peptides have been detected in a single fraction. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.

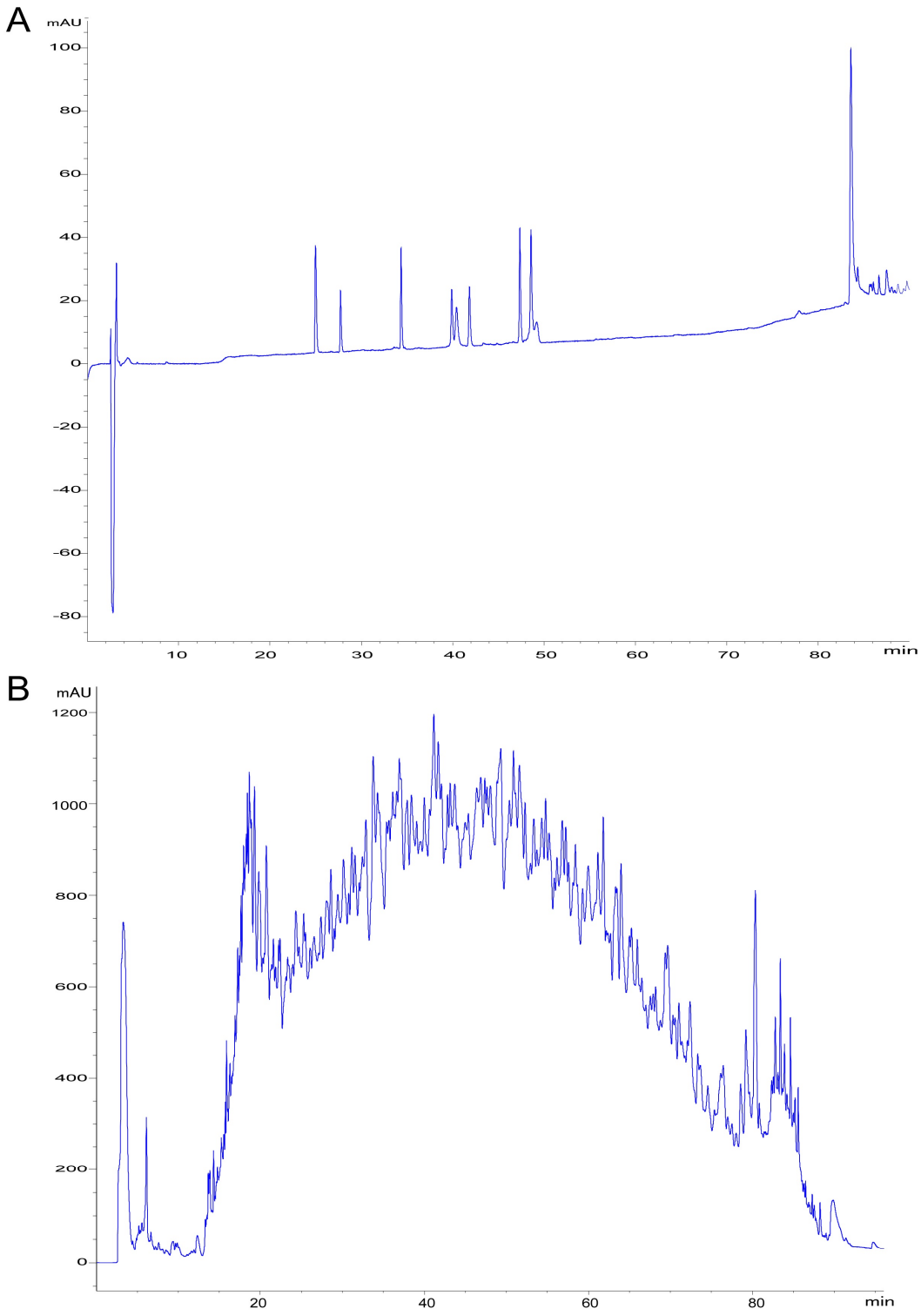




Supplementary Figure 2

Efficiency of tryptic protein digestion.

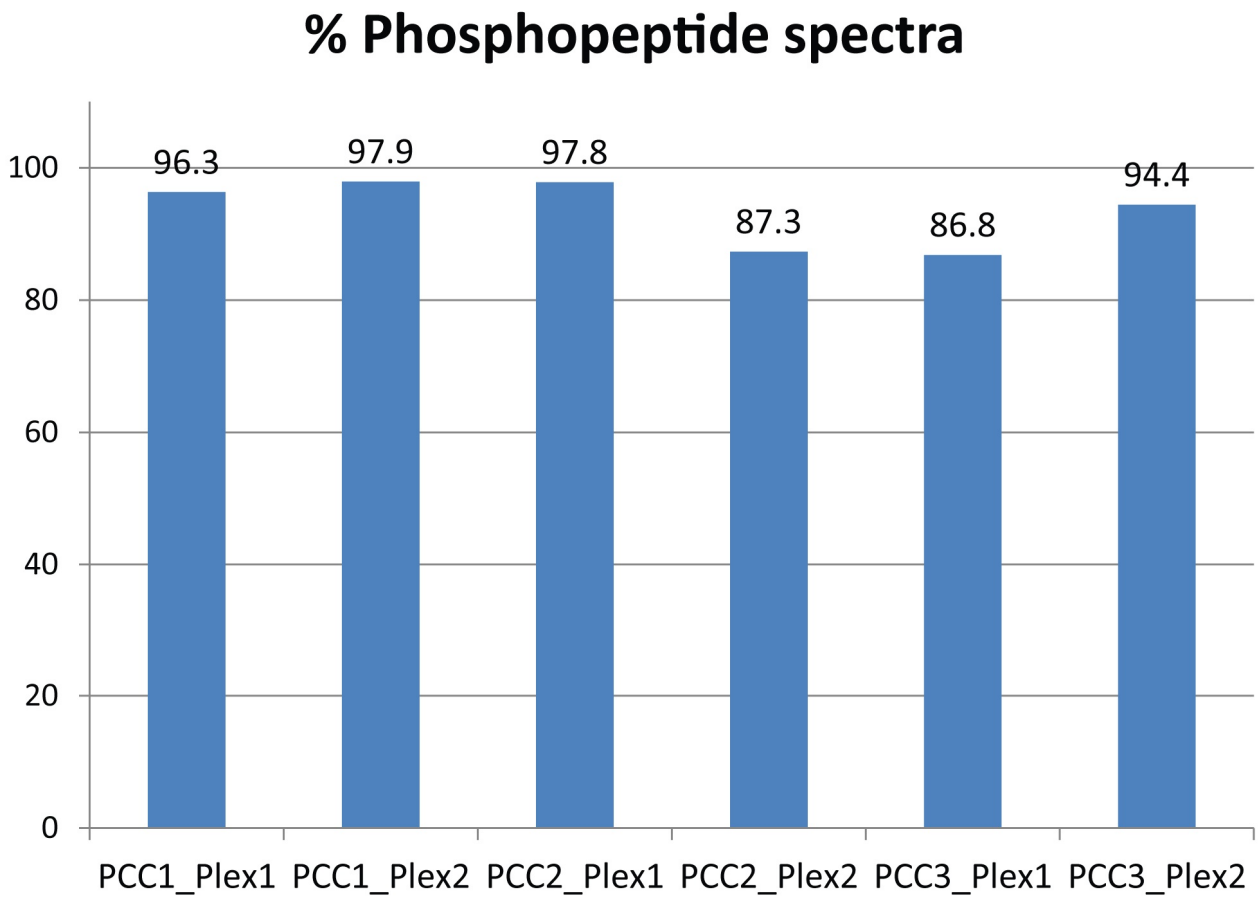
Missed cleavage rate for proteome (A) and phosphoproteome (B) data. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 3

QC of basic reversed phase chromatography.

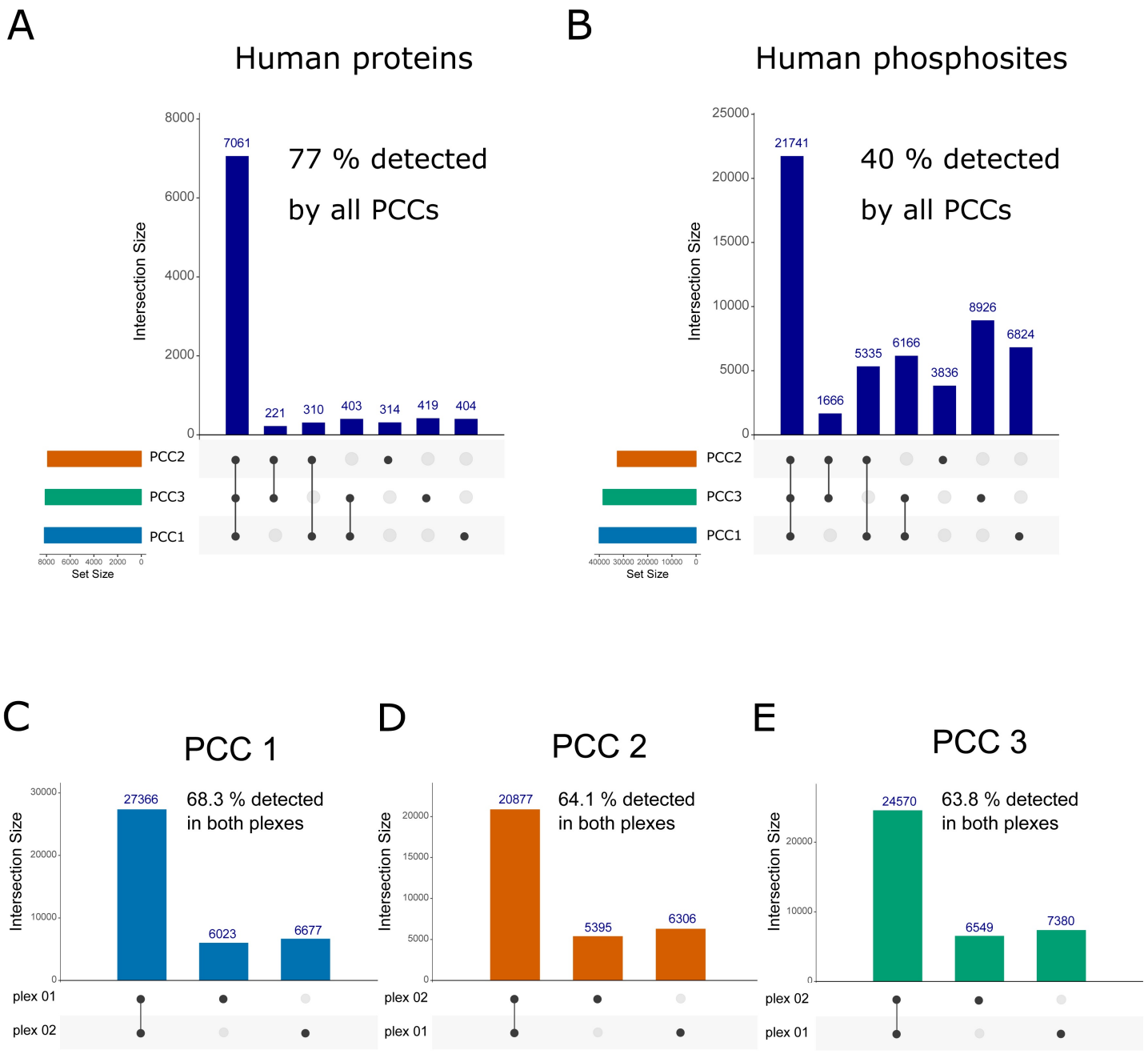
A) Chromatogram of non-human peptide standard mix on the offline HPLC column. B) Example of 4mg TMT-labeled sample fractionated on an Agilent Xtend Column.



Supplementary Figure 4

Enrichment specificity for phosphopeptides.

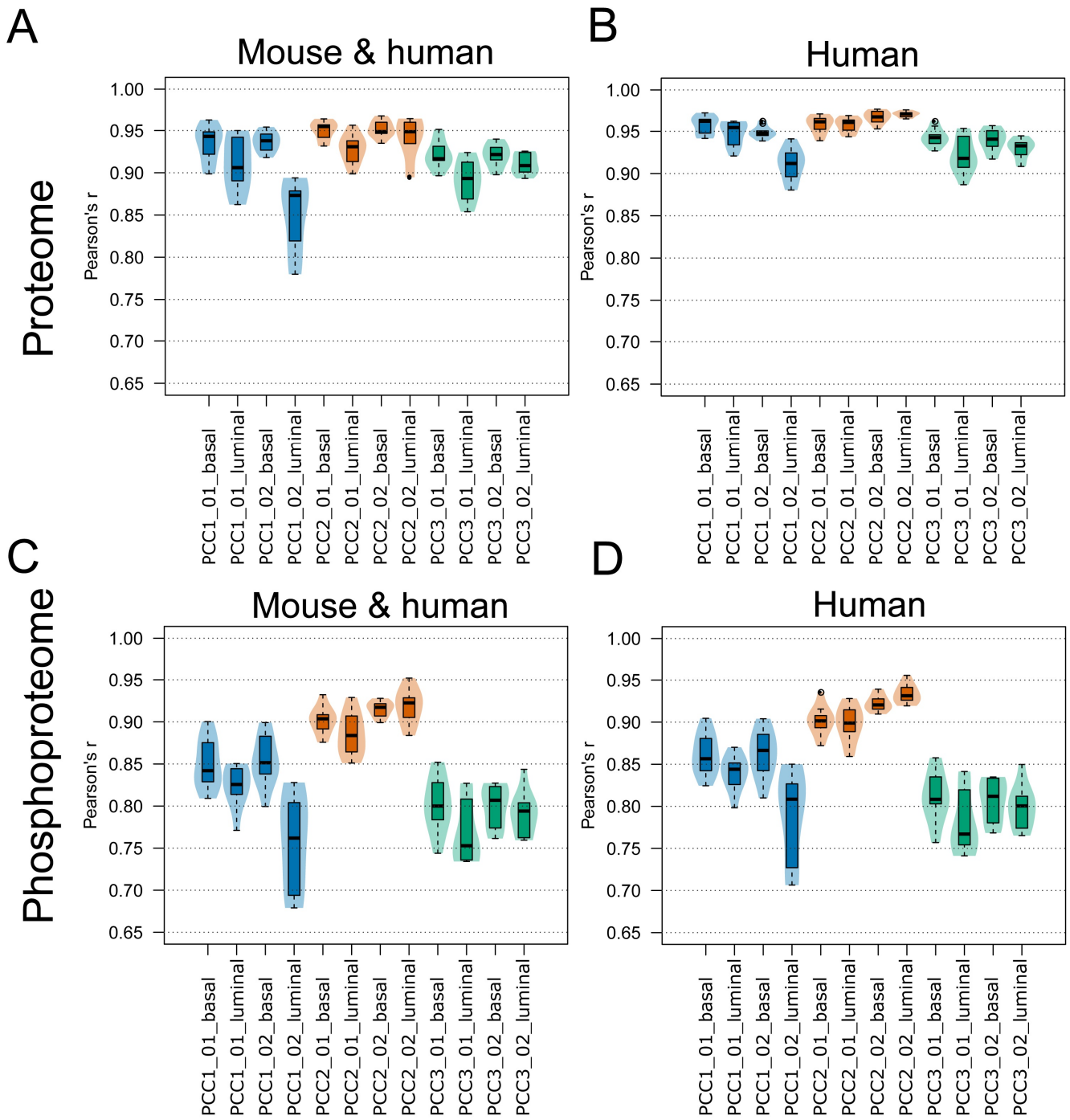
The bar chart depicts the percentage of phosphorylated peptides in the metal affinity enriched fractions. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 5

Number of quantified proteins and phosphorylation sites across three laboratories.

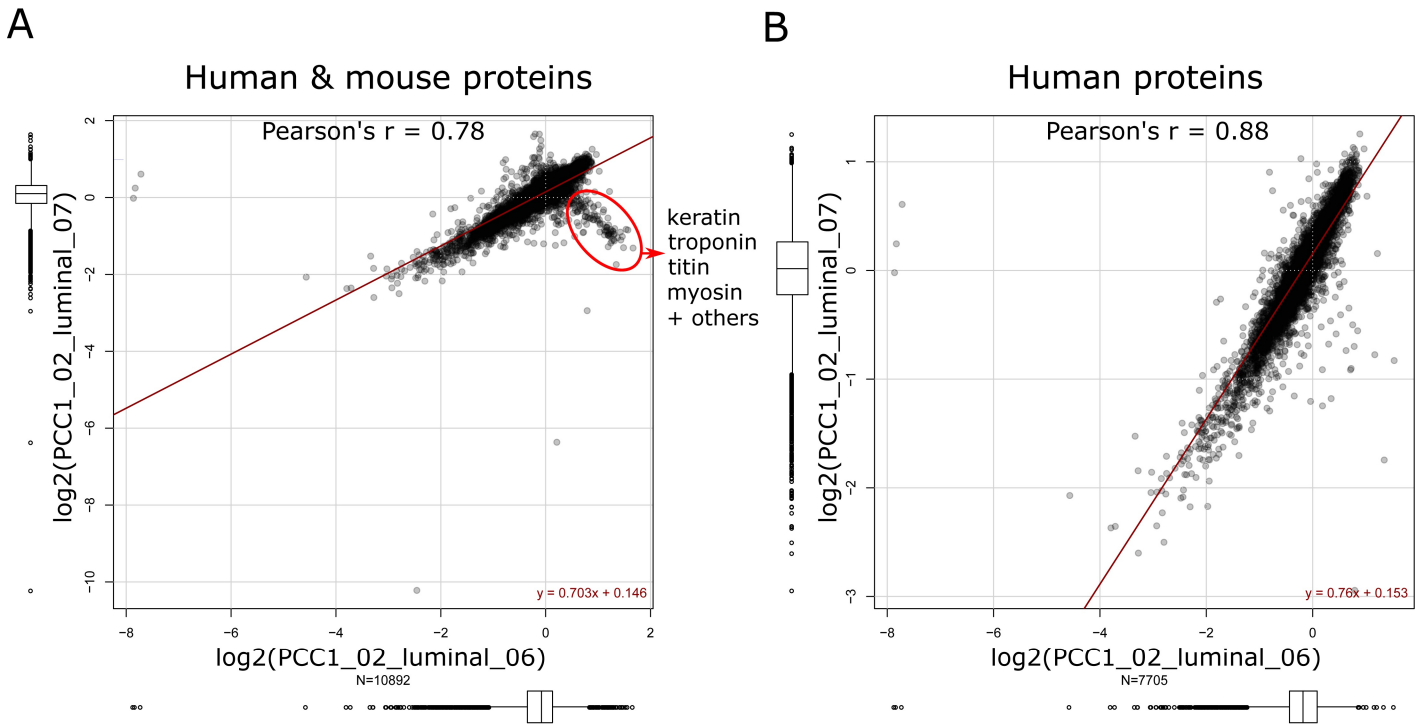
The total number of quantified A) proteins and B) phosphorylation sites that could be assigned to human are illustrated as 'UpSet' plots (Lex, 2014). Horizontal bars indicate total number of proteins or phosphorylation sites detected by each laboratory; vertical bars depict the number of jointly detected features as indicated by the layout matrix below. C) Comparison of detected and quantified human phosphosites in two TMT10-plexes acquired in PCC 1. D) and E) analogous plots for PCC 2 and PCC3, respectively. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 6

Intra-plex correlations calculated on human+mouse and human-specific proteins and phosphorylation sites.

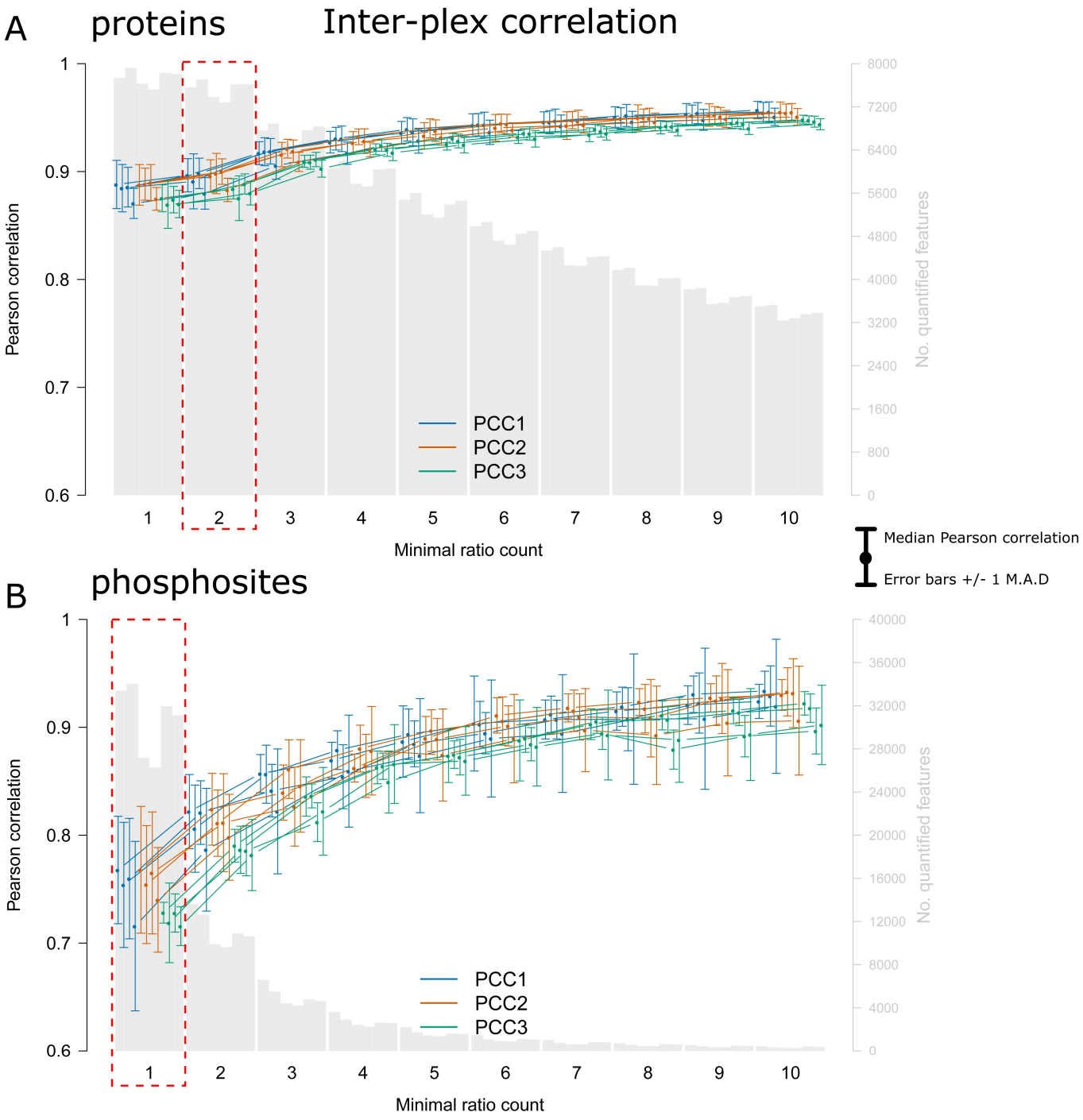
The box-whisker plots depict Pearson correlations of TMT ratios between intra-plex replicate measurements, separated into human+mouse and human-specific. A) Intra-plex correlations of human+mouse and B) human-specific proteins. C) Intra-plex correlations of human+mouse and B) human-specific phosphorylation sites. Focusing on proteins and phosphorylation sites with human origin resulted in lower intra-plex variability. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 7

Mouse-specific proteins contribute to high intra-plex variability.

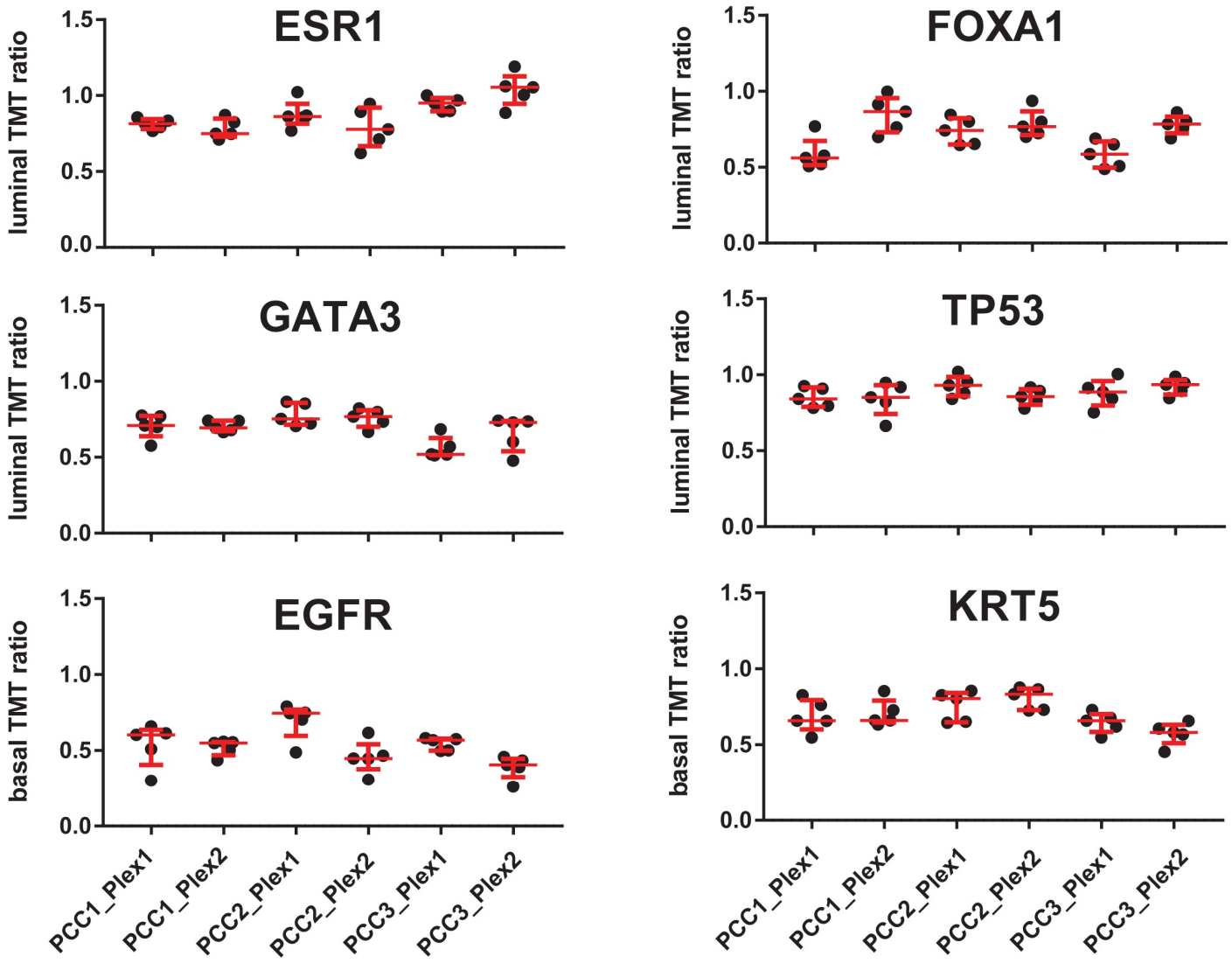
The example shown illustrates the impact of mouse proteins on reproducibility. A) Scatter plot of protein TMT ratios from human+mouse proteins measured in two luminal intra-plex replicate measurements. A small population of mouse proteins could not reproducibly quantified in the two replicates resulting in a moderate Pearson correlation of 0.78. B) After removal of mouse-specific proteins the correlation significantly improved to 0.88. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 8

Impact of ratio count pre-filtering on coverage and reproducibility.

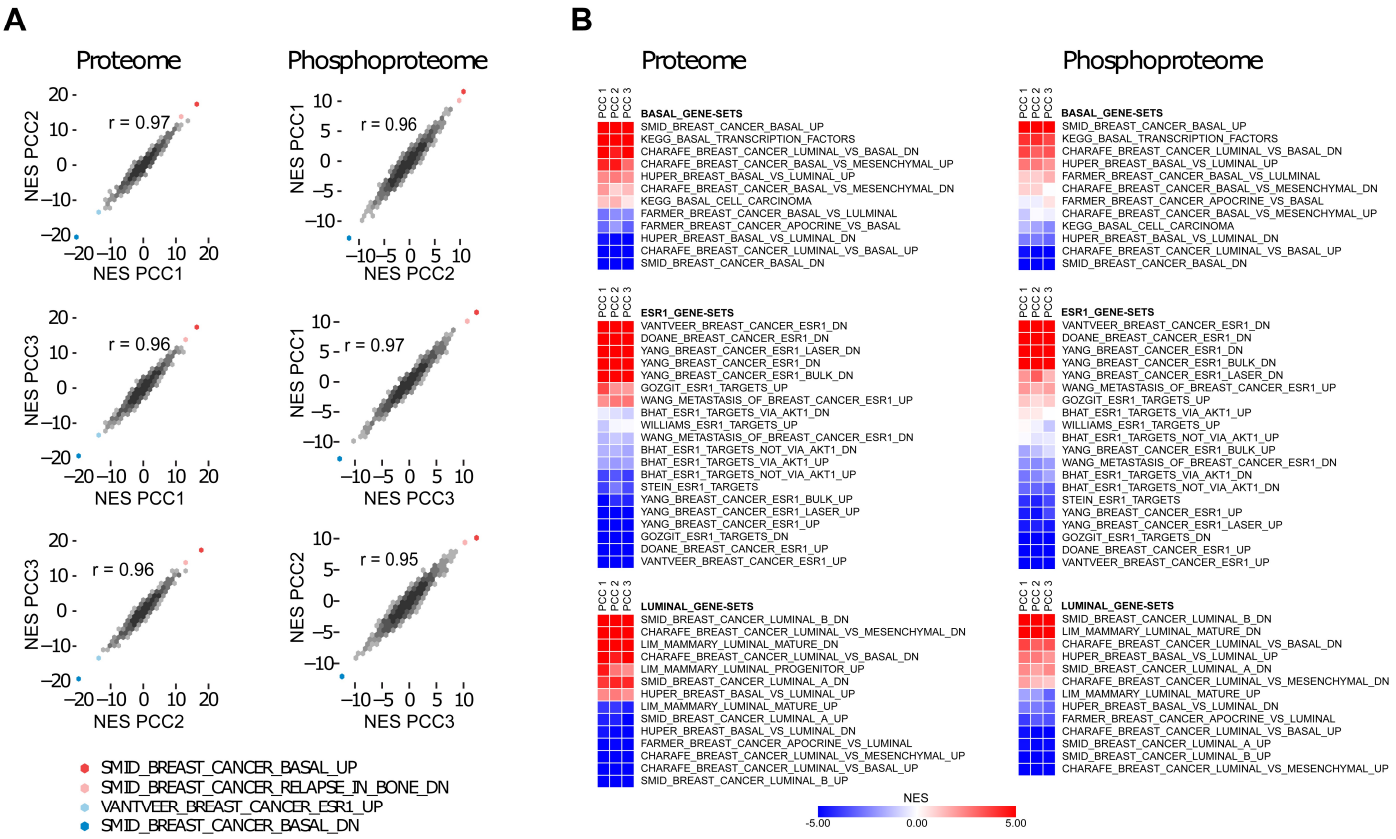
Inter-plex correlations were calculated for different numbers of ratio counts (x-axis) required to calculate A) protein and B) phosphorylation site ratios, respectively. Median Pearson correlations as a function of minimal ratio counts are depicted as lines (left axis). Error bars correspond to +/- 1 M.A.D. The number of quantified features at different minimal ratio counts are illustrated as bar plots (right axis). Red dashed boxes indicate the number of minimal ratio counts providing the best tradeoff between reproducibility and coverage and were subsequently used in this study. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 9

Breast cancer relevant proteins were consistently measured within and across laboratories.

Proteins for ESR1, FOXA1, GATA3, and TP53 show higher expression in luminal PDX samples, whereas proteins for EGFR and KRT5 are more abundant in basal PDX samples. Normalized TMT ratios of basal or luminal samples over a multi-mean denominator are shown. Red lines indicate the median of 5 measurements per TMT10-plex experiment with interquartile range bars. PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.



Supplementary Figure 10

Biology is very well recapitulated between all three centers.

A) Normalized enrichment scores (NES) show high correlation with Pearson r's of 0.96 to 0.97. The four most enriched gene-sets are color-coded and include SMID_BREAST_CANCER_BASAL_UP and SMID_BREAST_CANCER_RELAPSE_IN_BONE_DN, which are enriched in proteins and phosphoproteins that are high in the basal tumors compared to the luminal tumors. Conversely, SMID_BREAST_CANCER_BASAL_DN and VANTEER_BREAST_CANCER_ESR1_UP are enriched in proteins and phosphoproteins that are high in the luminal tumors (and therefore have low NES-scores). B) All gene-sets containing the terms: "BASAL", "ESR1" or "LUMINAL" are plotted as three heat-maps; one for each search term. Red indicates a high NES score and blue a low NES-score. Protein Characterization Center (PCC). PDX models used in this study were approved by the institutional animal care and use committee at Washington University in St. Louis.