Analysis of Limited Proteolysis-Coupled Mass Spectrometry Data

Authors

Luise Nagel, Jan Grossbach, Valentina Cappelletti, Christian Dörig, Paola Picotti, and Andreas Beyer

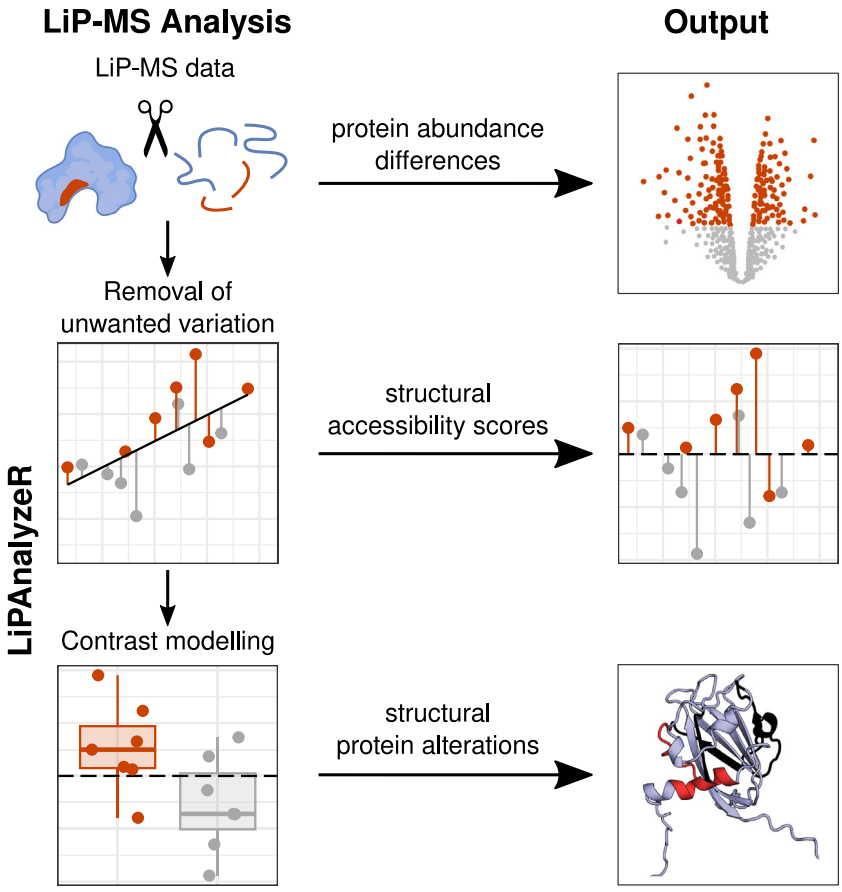
Correspondence

andreas.beyer@uni-koeln.de

In Brief

Limited proteolysis combined with mass spectrometry (LiP-MS) facilitates probing structural changes on a proteome-wide scale. Distinguishing different contributions to the LiP-MS signal, such as changes in protein abundance or chemical modifications, from structural protein alterations remains challenging. Here, we present a comparative exploration of approaches for the analysis of LiP-MS data. We introduce LiPAnalyzeR, a comprehensive computational pipeline that provides a uniquely powerful approach for deconvolving LiP-MS signals using a two-step approach.

Graphical Abstract



Highlights

- LiPAnalyzeR - a statistical framework for deconvoluting LiP-MS signal.
- Inferring structural protein alterations from LiP-MS data.
- Robust removal of unwanted variation and confounding variables.
- Outperforms alternative approaches.

2025, Mol Cell Proteomics *24*(*4*), 100934

https://doi.org/10.1016/j.mcpro.2025.100934

Analysis of Limited Proteolysis-Coupled Mass Spectrometry Data

Luise Nagel[1], Jan Grossbach[1], Valentina Cappelletti[2], Christian Dörig[2], Paola Picotti[2], and Andreas Beyer[1,3,4,*]

Limited proteolysis combined with mass spectrometry (LiP-MS) facilitates probing structural changes on a proteome-wide scale. This method leverages differences in the proteinase K accessibility of native protein structures to concurrently assess structural alterations for thousands of proteins *in situ*. Distinguishing different contributions to the LiP-MS signal, such as changes in protein abundance or chemical modifications, from structural protein alterations remains challenging. Here, we present the first comprehensive computational pipeline to infer structural alterations for LiP-MS data using a two-step approach. 1) We remove unwanted variations from the LiP signal that are not caused by protein structural effects and 2) infer the effects of variables of interest on the remaining signal. Using LiP-MS data from three species, we demonstrate that this approach outperforms previously employed approaches. Our framework provides a uniquely powerful approach for deconvolving LiP-MS signals and separating protein structural changes from changes in protein abundance, posttranslational modifications, and alternative splicing. Our approach may also be applied to analyze other types of peptide-centric structural proteomics data, such as FPOP or molecular painting data.

The function of a protein is closely linked to its structural state, which in turn affects corresponding phenotypes (1). Protein structural changes are important mediators of cell signaling, metabolic adaptation, molecular stress, and genetic variability (2, 3). Recent technological advances have made it possible to investigate protein structural changes on a proteomic scale (4–7). One method for simultaneously quantifying alterations in the structural state of thousands of proteins *in situ* is limited proteolysis coupled to mass spectrometry (LiP-MS). This method utilizes differences in the accessibility of protein regions to an unspecific protease, such as Proteinase K (PK) (4, 8, 9). LiP-MS has shed light on a variety of biological topics including metabolic adaptation (10), thermostability (11), cancer-related conformational changes (12, 13), protein–protein interactions (14–17) including proteome-specific protein–protein interactions (18), aging-related changes in yeast (19), *Caenorhabditis elegans* (20), and mice (21), as well as the use of multidimensional protein-structural changes as a new class of disease biomarkers (22).

To quantify changes in the structural protein accessibility between conditions, lysates from all conditions undergo short (i.e. limited) PK digestion, followed by a conventional complete trypsin digestion (LiP) (Fig. 1). Differences in the protein accessibility between the conditions will reflect in distinct PK digestion patterns (Fig. 1, yellow circled region) yielding altered peptide quantities. These accessibility changes may result from protein structural rearrangements or from variable shielding of surface areas, for example, through ligand binding. Mass spectrometry (MS) is used to quantify those variations in peptide abundance on a proteome-wide scale.

The analysis of LiP-MS data faces a number of challenges, some of which are common to molecular high-throughput data (e.g. correcting for technical variation such as batch effects, multiple hypothesis correction), while others are new and specific to LiP-MS data. First, changes in LiP peptide signals (i.e. the MS signal of a peptide subjected to the double digestion with PK and trypsin) may result from factors other than PK accessibility effects, that is, PK-independent effects, such as variation in protein abundance, alternative splicing, or post-translational modifications (PTMs). In order to account for such confounding effects, LiP-MS studies typically perform a second control digestion without limited PK digestion and only consisting of the trypsin digestion (TrP) (Fig. 1). This is based on the underlying notion that PK-independent effects manifest equally in both the LiP as well as the trypsin-only digestion. Therefore, MS signal changes that are common to both data are not due to differences in the PK accessibility. Analyses schemes that appropriately combine the data from LiP-treated and TrP control samples are required to infer the signal of PK accessibility changes. Second, the PK digestion generates many half-tryptic peptides, that is, peptides with a trypsin digestion site at one end

From the [1]Cologne Excellence Cluster for Aging and Aging-Associated Diseases (CECAD), University of Cologne, Cologne, Germany; [2]Institute of Molecular Systems Biology, Department of Biology, ETH Zurich, Zurich, Switzerland; [3]Faculty of Medicine and University Hospital of Cologne, and Center for Molecular Medicine, Cologne, University of Cologne, Cologne, Germany; [4]Institute for Genetics, Faculty of Mathematics and Natural Sciences, University of Cologne, Cologne, Germany

*For correspondence: Andreas Beyer, andreas.beyer@uni-koeln.de.

https://doi.org/10.1016/j.mcpro.2025.100934

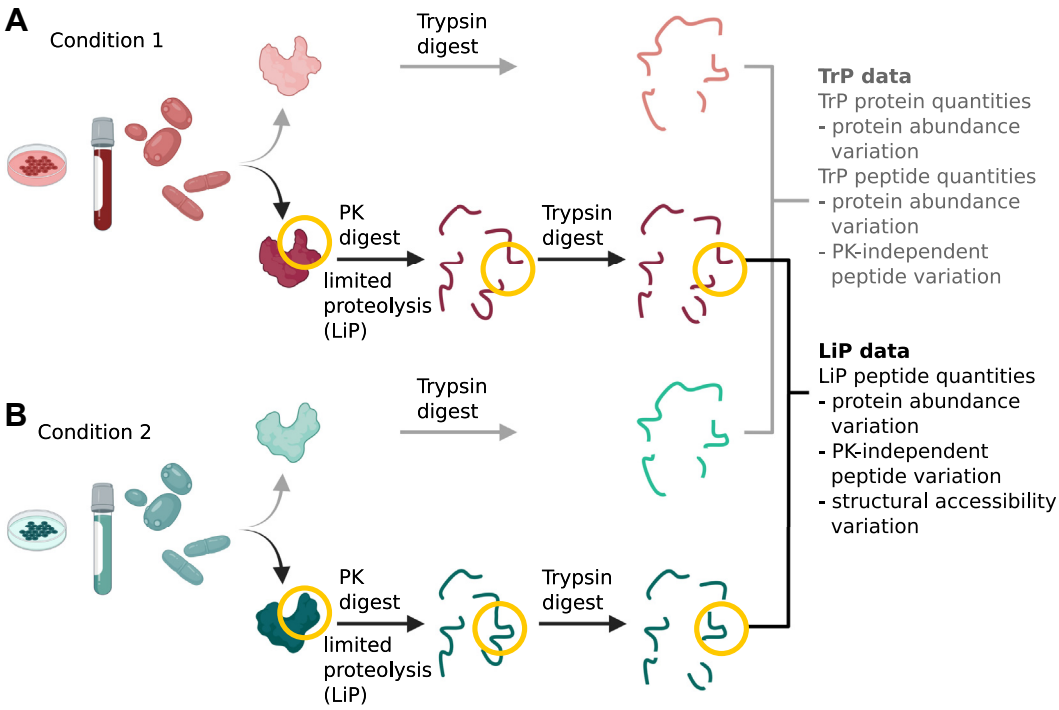


FIG. 1. **Experimental LiP-MS workflow.** The proteome from samples under different conditions are extracted in nondenaturing, native conditions and digested with (*A*) trypsin-only, (*B*) PK (short time, limited proteolysis) followed by trypsin. Regions with a difference in the three-dimensional conformation (*yellow*) show a different digestion pattern in the limited proteolysis step. Subsequently, LiP and TrP intensities are quantified via mass spectrometry. (Created with BioRender.com).

of the peptide, while the other end was digested by PK, due to the sequence-unspecific cleavage behavior of PK. These half-tryptic peptides may carry additional information about the specific location of protein structural changes. However, their analysis is hampered by the fact that they do not exist in the trypsin-only control samples.

To address all of the LiP data-specific challenges listed above, a computational approach is required that allows to distinguish the origin of variation in the LiP signal and subsequently identify the signal caused by differences in PK accessibility. As additional variation in the LiP signal can be induced by technical effects, such as batch effects, and biological variance, analysis approaches should also enable for the removal of this unwanted variation. Existing analysis frameworks for proteomics data do not fully address all of these challenges. Approaches, typically used in the analysis of MS data that are not specifically designed for LiP-MS data (23–25), are capable of correcting for covariates such as batch effects, but do not account for LiP-specific challenges in the data. State-of-the-art methods designed for or previously applied to LiP-MS data do not correct for all types of unwanted variation. For example, MSstatsLiP corrects LiP signals for variation caused by protein abundance changes, but it does not correct for peptide-level changes, for example, due to alternative splicing (9). We are not aware of any tool that appropriately corrects for LiP peptide signals for all PK-independent variations.

Hence, we have developed a comprehensive method which enables the removal of unwanted variation (RUV) from LiP-MS data with different experimental setups (Fig. 2). Our method subsequently estimated the changes in the structural accessibility between samples or conditions and is available as the R package LiPAnalyzeR. Our analysis reveals the importance of removing unwanted variation at the protein and peptide level, thereby creating a reference framework for the analysis of LiP-MS data.

EXPERIMENTAL PROCEDURES

Computational Method of LiPAnalyzeR

The RUV step removes unwanted variation from the LiP peptide matrix $Y_{LiP} = (y_{LiP,ij})$ with peptide i in sample j using a model that describes the contribution of unwanted covariables X to the measured peptide intensity:

$$Y_{LiP} = \beta_0 + \beta_1 X_{Prot} + \beta_2 X_{TrP} + \overline{\beta}_m \overline{X} + \epsilon \quad (1)$$

This model estimates the contribution of (whole) protein levels $X_{Prot} = (x_{Prot,ij})$, trypsin-only (TrP) peptide intensities $X_{TrP} = (x_{TrP,ij})$, and other unwanted covariates $\overline{X} = (x_{ijm})$ to the LiP signal variation of each peptide i independently. The matrices X_{Prot}, X_{TrP}, and unobserved errors $\epsilon = \epsilon_{ij}$ have the dimensionality $a \times b$; β_0, β_1, and β_2 are vectors of the length a with β_1 and β_2 being defined as ≥ 0; the $a \times b \times c$ matrix $\overline{X}$ provides further variation factors for RUV and $\overline{\beta}_m$ is a matrix of $a \times c$. Here, a = number of peptides, b = number of samples, and c = number of further covariables for RUV.

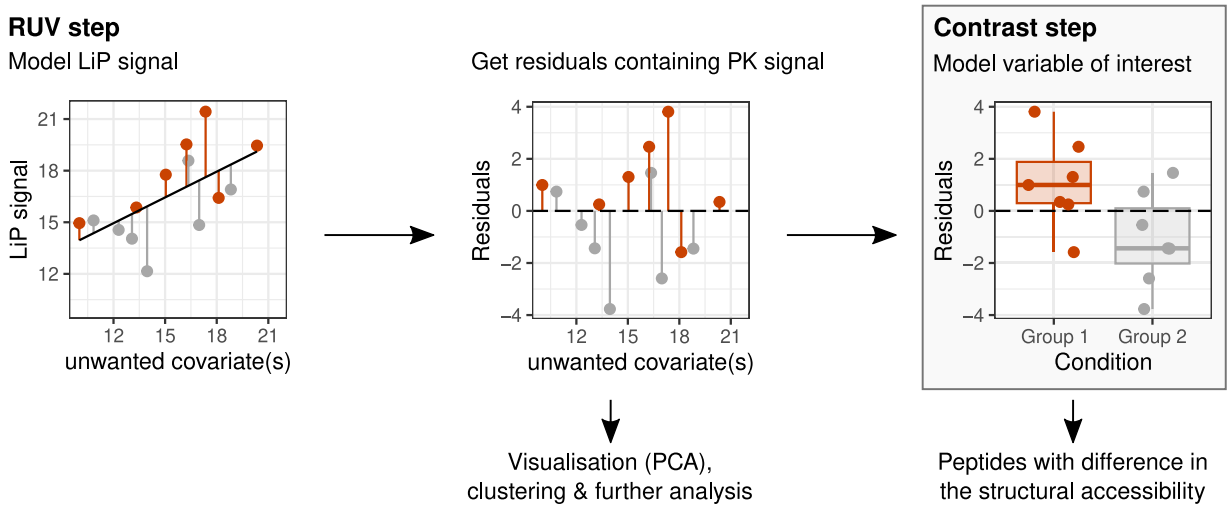


FIG. 2. **Overview LiPAnalyzeR pipeline.** Covariates, which are considered a source of unwanted variation (unwanted covariates), such as protein abundances, are regressed out from the LiP signal in an RUV step (*left*). Residuals containing the variation of interest are estimated from the RUV models. These residuals can be used for data visualization (PCA), clustering, and further downstream analysis (*center*). In the subsequent contrast step, the effects of the variable of interest are modeled on the residuals, identifying peptides with changes in the structural accessibility between conditions (*left*).

We then estimate all the variation which can be explained by X_{Prot}, X_{TrP}, and $\overline{X}$:

$$\tilde{Y}_{LiP} = \beta_0 + \beta_1 X_{Prot} + \beta_2 X_{TrP} + \overline{\beta}_m \overline{X} \quad (2)$$

defining the $a \times b$ matrix $\tilde{Y}_{LiP} = (\tilde{y}_{LiP,ij})$, which is subsequently used to estimate the $a \times b$ residual matrix $\hat{Y}_{LiP} = (\hat{y}_{LiP,ij})$:

$$\hat{Y}_{LiP} = Y_{LiP} - \tilde{Y}_{LiP} \quad (3)$$

$\hat{Y}_{LiP}$ contains the structural accessibility variation of interest, since all PK-dependent variation contained in trypsin-only data has been removed, as well as further variation caused by factors such as batch included in $\overline{X}$. Accessibility variation for the covariable(s) of interest n can now be inferred in the contrast model

$$\hat{Y}_{LiP} = \beta_{0'} + \overline{\beta}_n \overline{Z} \quad (4)$$

Were $\overline{Z} = (z_{jin})$ is a $a \times b \times d$ matrix and $\overline{\beta}_n$ is a matrix of $a \times d$, with d = number of variables of interest. $\overline{\beta}_n$ is then evaluated via t statistics, with the degrees of freedom used in the RUV model (Equation 1) being removed from the available degrees of freedom prior to estimating the p-value. Subsequently, we correct for multiple testing effects by computing corrected p-values using the Benjamini–Hochberg approach over all peptides from the same protein ('protein-wise FDR').

Combining the RUV (Equation 1) and contrast (Equation 2) model into one regression (section *'Superior performance of applying a constrained RUV model prior to the contrast model'*), results in

$$Y_{LiP} = \beta_0 + \beta_1 X_{Prot} + \beta_2 X_{TrP} + \overline{\beta}_p \overline{W} + \epsilon \quad (5)$$

where $\overline{W} = (w_{jip})$ is a $a \times b \times e$ matrix, combining $\overline{X}$ and $\overline{Z}$, and $\overline{\beta}_p$ is a matric of $a \times e$, combining $\overline{\beta}_m$ and $\overline{\beta}_n$, resulting in $e = c + d$.

Experimental Design and Statistical Rationale

The fission yeast data consists of three different isolates (JB50, JB759, JB760) and a mutant of one of these isolates (JB50 PYK1, see Experimental procedures) (26). For each of the four strains, five biological replicates were measured in matching LiP and TrP samples, a number of replicates representative of commonly used setups on LiP-MS experiments. This dataset provides a more diverse biological variation than other yeast datasets used in this study, as it contains four distinct yeast strains. Compared to the budding yeast data, this dataset has fewer replicates. It is used for visualizing general properties of LiP-MS yeast data and basic explorations of analyzing LiP-MS data. The budding yeast dataset consists of 11 biological replicates each of two well-characterized isolates, which were measured in matching LiP and Trp samples (27). Two technical replicates were measured for every individual sample. This dataset has a smaller amount of biological variation, since it only contains two different strains but provides a higher number of biological and technical replicates, therefore serving as a dataset to investigate consistency and reproducibility of results from inferring structurally accessibility variation using different approaches.

We additionally utilized an existing human cerebrospinal fluid (CSF) dataset containing 52 samples from healthy individuals each measured with LiP and Trp digest (22). This human data set differs from the yeast data, in that the samples differ more in study variables factors (such as age, sex, and environment) and also have a much more complex and varied genetic background. For instance, the budding yeast data contains only two fairly diverse genotypes, driving strong abundance effects for many protein levels. The human CSF data reflects greater proteome complexity than the yeast data and has many more PK-independent peptide effects, for example, due to alternative splicing. In addition to the fission yeast dataset, we use the human CSF for the general exploration of LiP-MS data properties and to benchmark our LiPAnalyzeR approach for inferring structural accessibility variation.

Limited Proteolysis and Liquid Chromatography-Mass Spectrometry of Fission Yeast

Sample Preparation and Limited Proteolysis—Five biological replicates of the *Schizosaccharomyces pombe* JB50, JB759, JB760, and JB50-PYK1 mutant strains were cultivated and harvested as described in Kamrad *et al.*, 2020 (26). Frozen cell pellets were resuspended in 400 µl cold LB, mixed with the same volume of acid-washed

glass beads (Sigma-Aldrich), transferred to a FastPrep-24 TM 5G Instrument (MP Biomedicals), and disrupted at 4 °C by eight rounds of bead-beating at 30 s with 200 s pauses between the runs. Samples were centrifuged (2 min, 1000*g*, 4 °C), supernatants collected, and protein concentrations determined with the bicinchoninic acid assay (Thermo Fisher Scientific). Proteome extracts were divided into two samples: a control sample (tryptic control, TrP) undergoing only tryptic digestion to measure protein abundance changes and a LiP sample subjected to a double-protease digestion with an unspecific protease followed by trypsin digestion to provide information on protein structural changes. PK (from Tritirachium album, Sigma Aldrich) was added to 100 μg of the LiP samples at an E:S ratio of 1:100 (w/w) and incubated for 5 min at 25 °C. The same volume of water was added to 100 μg of the control samples. Digestion reactions were stopped by heating LiP samples at 99 °C for 5 min, followed by the addition of sodium deoxycholate (Sigma Aldrich) to a final concentration of 5%. Control samples underwent the same procedure. Both LiP and TrP samples were then subjected to complete tryptic digestion under denaturing conditions. Peptides were reduced by incubation of samples with tris(2-carboxyethyl)phosphine (Thermo Fisher Scientific) to a final concentration of 5 mM for 30 min at 37 °C. Next, the alkylation of free cysteine residues was achieved by adding iodoacetamide (Sigma Aldrich) to a final concentration of 40 mM for 30 min at 25 °C in the dark. Samples were diluted with freshly prepared 0.1 M ammonium bicarbonate to a final concentration of 1% sodium deoxycholate. Samples were predigested with lysyl endopeptidase LysC (Wako Chemicals) at an enzyme/substrate ratio of 1:100. After 2 h at 37 °C, sequencing-grade porcine trypsin (Promega) was added to a final enzyme/substrate ratio of 1:100, and samples were incubated for 16 h at 37 °C under shaking at 800 rpm. Protease digestion was quenched by lowering the reaction pH (<3). The peptide mixtures were loaded onto Sep-Pak tC18 cartridges or 96 wells elution plates (Waters), desalted, and eluted with 80% acetonitrile, 0.1% formic acid. After elution from the cartridges, peptides were dried in a vacuum centrifuge, resolubilized in 0.1% formic acid, and analyzed by MS.

LC-MS/MS Data Acquisition—Peptide digestions were analyzed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific) equipped with a nanoelectrospray ion source and a nano-flow LC system (Easy-nLC 1000, Thermo Fisher Scientific). For shotgun LC-MS/MS data-dependent acquisition (DDA), 1 μl peptide digestion from each biological replicate was injected at a concentration of 1 mg/ml. One microliter of the same samples was also measured in data-independent acquisition (DIA) mode. Peptides were separated on a 40 cm × 0.75 μm i.d. column packed in-house with 1.9 μm C18 beads (Dr Maisch Reprosil-Pur 120). For LC fractionation, buffer A was 0.1% formic acid and buffer B was 0.1% formic acid in 100% acetonitrile using a linear LC gradient from 5% to 25% or 5% to 35% acetonitrile, respectively, over 120 min and a flow rate of 300 nl/min and the column was heated to 50 °C.

For DDA measurement on the Orbitrap Q Exactive Plus, MS1 scans were acquired over a mass range of 350 to 1500 m/z with a resolution of 70,000. The 20 most intense precursors that exceeded 1300 ion counts were selected for collision-induced dissociation and the corresponding MS2 spectra were acquired at a resolution of 35,000, collected for maximally 55 ms. All multiply charged ions were used to trigger MS-MS scans followed by a dynamic exclusion for 30 s. Singly charged precursor ions and ions of undefinable charged states were excluded from fragmentation.

For DIA measurements, 20 variable-width DIA isolation windows were recursively acquired (Supplemental Data File 1). The DIA isolation setup included a 1 m/z overlap between windows, as described in (37). DIA-MS2 spectra were acquired at a resolution of 17,500 with a fixed first mass of 150 m/z and an autoatic gain control target of 1 × 106. To mimic DDA fragmentation, normalized collision energy was 25, calculated based on the doubly charged center m/z of the DIA window. Maximum injection times were automatically chosen to maximize parallelization resulting in a total duty cycle of approximately 3 s. A survey MS1 scan from 350 to 1500 m/z at a resolution of 70,000, with autoatic gain control target of 3 × 106 or 120 ms injection time was acquired in between the acquisitions of the full DIA isolation window sets.

Library Generation and DIA—The data were searched in Spectronaut version 15.7.220308.50606 (Biognosys). Hybrid libraries for the tryptic control and the LiP samples consisting of the corresponding DDA and DIA runs were created based on a Pulsar search using the default settings, with the exception of digestion type, which was set to "semi-specific" for the LiP samples only, and the minimal peptide length, which was set to 6. The data were searched against customized fasta files (15,498 protein entries) (28). The targeted data extraction was performed in Spectronaut with default settings (number of missed cleavages permitted: 2, MS1 Mass Tolerance Strategy: dynamic, MS2 Mass Tolerance Strategy: dynamic, fixed modifications: Carbamidomethyl (C), variable modifications: acetyl (Protein N-term), Oxidation (M)) except for the machine learning, which was set to "across experiment," imputation which was deactivated and the data filtering, which was set to "Qvalue." The false discovery rate (FDR) was set to 1% on peptide and protein levels. Detailed search information can be found in Supplemental Data File 2 (LiP) and 3 (TrP-only). The LiP and tryptic control samples were searched separately. Precursor reports of the LiP (Supplemental Data File 4) and TrP-only (Supplemental Data File 5) are provided. Results were exported using the LiPAnalyzeR_SpectroScheme format.

Data Preparation—LiP and TrP quantities were extracted from the exported data from Spectronaut using the columns "*PEP.Quantity*" for peptide and "*PG.Quantity*" for protein quantities. Peptides missing a measurement in one of the samples were removed from any downstream analysis, resulting in 11,756 full-tryptic peptides from 2119 proteins and 7489 half-tryptic peptides from 1308 proteins. Peptide and protein quantities were log2-transformed. For visualization of the data, and all analysis were LiPAnalyzeR was not run (e.g. estimating LiP/TrP ratios), batch correction was performed on the data using the *removebatch(1)* function from the r-package *limma* (29). Batch here refers to the day at which the samples were measured in the MS. Analysis on the data without RUV models were performed on the batch-corrected data. If RUV was performed, it was run on the not batch corrected data and batch was instead added into the model as a coefficient.

Limited Proteolysis and Liquid Chromatography-Mass Spectrometry of Budding Yeast

Sample Preparation, Limited Proteolysis, and LC-MS/MS Data Acquisition—For this study, we used 11 biological replicates of the laboratory BY4716 *Saccharomyces cerevisiae* strain, an S288C derivative (MATα lys2Δ0), and 11 biological replicates of the wild isolate RM11-1a (MATa leu2Δ0 ura3Δ0 ho::KAN) (27). Single colonies of the two strains were picked from fresh plates and inoculated in synthetic complete medium (SC,Formedium) and grown for 6 h at 30 °C under shaking at 150 rpm. The precultures were inoculated into fresh SC medium cultures to a final A600 of 0.0004 and grown overnight at 30 °C under constant shaking at 150 rpm. When cultures reached A600 = 0.8 ± 0.1, the liquid medium was removed by 5 min centrifugation at 800*g*, RT. The pellets were washed with 20 ml of PBS 1X and centrifuged at 800*g*, RT. Cell pellets were finally resuspended in RT lysis buffer (100 mM Hepes, 1 mM MgCl2, 150 mM KCl, pH 7.5), flash-frozen, and stored at −80 °C. Cell lysis was performed as described above, using a FastPrep-24 TM 5G Instrument (MP Biomedicals), with the following settings: speed = 5.5 m/s, time = 30 s, eight cycles, rest

time = 200 s. Proteome extracts were then processed as described above and peptides analyzed on an Orbitrap Q Exactive Plus mass spectrometer (Thermo Fisher Scientific), see "LC-MS/MS data acquisition" section above.

Library Generation and DIA—The data were searched in Spectronaut version 15.7.220308.50606 (Biognosys). Libraries for the tryptic control and the LiP samples consisting of the DIA runs were created based on a Pulsar search using the default settings, with the exception of digestion type, which was set to "semi-specific" for the LiP samples only, and the minimal peptide length, which was set to 6. The data were searched against customized fasta files (13,434 protein entries) (30). The targeted data extraction was performed in Spectronaut with default settings (number of missed cleavages permitted: 2, MS1 Mass Tolerance Strategy: dynamic, MS2 Mass Tolerance Strategy: dynamic, fixed modifications: Carbamidomethyl (C), variable modifications: Acetyl (Protein N-term), Oxidation (M)) except for the machine learning, which was set to "across experiment," imputation which was deactivated and the data filtering, which was set to "Qvalue." The FDR was set to 1% on peptide and protein levels. Detailed search information can be found in Supplemental Data File 6 (LiP) and 7 (TrP-only). The LiP and tryptic control samples were searched separately. Precursor reports of the LiP (Supplemental Data File 8) and TrP-only (Supplemental Data File 9) are provided. Results were exported using the LiPAnalyzeR_SpectroScheme format.

Data Preparation—LiP and TrP quantities were extracted from the exported data from Spectronaut using the columns "*PEP.Quantity*" for peptide and "*PG.Quantity*" for protein quantities. Technical replicates were combined taking the mean of both measurements while omitting NAs. Peptides missing a measurement in one of the samples were removed from any downstream analysis, resulting in 18,126 full-tryptic peptides from 2411 proteins and 9953 half-tryptic peptides from 1343 proteins. Peptide and protein quantities were log2-transformed. For visualization of the data, and all analysis of LiPAnalyzeR was not run, batch correction was performed using the *removebatch(1)* function from the r-package *limma*. Batches refer to different sample handling (different times of cell culture) in the lab.

For comparing the results of different groups with each other, the data was split into technical and biological replicate groups. For the technical replicates, the two replicates per strain were each assigned to a different group, resulting in two technical replicates groups with 11 biological replicates of each strain in each of the groups (R1, R2). For creating the biological replicates groups (G1, G2), the summarizing technical replicates were divided by batch with G1 consisting of batch 1 to 5 and G2 batch 6 to 10.

Limited Proteolysis and Liquid Chromatography-Mass Spectrometry in Human CSF

Human CSF samples were prepared and analyzed as described by Mackmull *et al.* (22). The results from the Spectronaut search performed by Mackmull *et al.* were exported using the LiPAnalyzeR_SpectroScheme format.

Data Preparation—LiP and TrP quantities were extracted from the exported data from Spectronaut using the columns "*PEP.Quantity*" for peptide and "*PG.Quantity*" for protein quantities. Peptides missing a measurement in more than 20 samples in either the healthy or PD samples were removed from any downstream analysis, resulting in 9983 full-tryptic peptides from 1272 proteins/protein isoforms and 5020 half-tryptic peptides from 762 proteins/protein isoforms. Peptide and protein quantities were log2-transformed. Batch correction was performed on the complete dataset (healthy and PD individuals) using the *removebatch(1)* function from the r-package *limma*. As this dataset only has eight samples per batch, removing batch at this early stage of the analysis allows for a more stable batch correction as more samples can be included. Subsequently, only the 52 healthy CSF samples were exported and used for analysis in this work.

Limited Proteolysis and Liquid Chromatography-Mass Spectrometry of Spiked in α-Synuclein

Yeast samples with spiked in monomeric and fibril α-Synuclein as provided in the Spectronaut reports by Malinovska *et al.* (9) (Spectronaut report Lip samples: 2019-04-29_Normalization_LiP_MSStats_Report.csv, Spectronaut report TrP samples: 2019-04-29_Normalization_Trp_MSStats_Report.csv) (31) were used.

Data Preparation—LiP and TrP quantities were extracted from the exported data from Spectronaut using the columns "*PEP.Quantity*" for peptide and "*PG.Quantity*" for protein quantities. Technical replicates were combined taking the mean of both measurements while omitting NAs. Peptides missing a measurement in one of the samples were removed from any downstream analysis. Peptide and protein quantities were log2-transformed.

RESULTS

LiPAnalyzeR: A Tailored Bioinformatic Pipeline for Inferring Structural Variation from LiP-MS Data

The measured signal of a peptide from the LiP-MS assay ('LiP peptide') is affected by changes in PK accessibility and other (usually unwanted) covariates like protein abundance (Fig. 3*A*), changes of the proteoform (e.g. alternative splicing and PTMs; Fig. 3*B*), and technical factors such as batch effects. Changes in protein abundance, changes in peptide abundance, and changes in PK accessibility impact the relationships between the proteomics layers in distinct ways—a fact that is exploited by LiPAnalyzeR to disentangle their effects on the data. LiPAnalyzeR assumes that actual structural effects, reflected in changes of PK accessibility, only affect the signal of LiP peptides, while protein levels ('TrP proteins') and signals of peptides in the trypsin-only control measurements ('TrP peptides') report only on effects other than PK accessibility (Fig. 3*C*). Figure 3 visualizes the described effects that potentially alter LiP peptide intensities other than changes in PK digestion using example peptides from the fission yeast data. For example, in the case of Figure 3*A*, the example peptide has a protein abundance effect, that is, the abundance of that protein varies between the four strains (marked by color). This effect equally affects LiP and TrP peptides as well as the TrP protein abundance; therefore, all three measurements show a strong positive correlation with each other. In the case of Figure 3*B*, the peptide has a PK-independent peptide effect, that is, peptide intensity varies between strains, but not due to changes in PK accessibility. As a consequence, the two peptide measures TrP peptide and LiP peptide correlate with each other (left), but not with the total protein estimate TrP protein (middle and right). The example peptide shown in Figure 3*C* is differentially accessible for PK in the different strains. This results in TrP peptide signal variation that neither correlates with the TrP peptide estimate (left) nor with the whole protein estimate (middle). Yet, the signals of the TrP peptide and whole protein still correlate with each

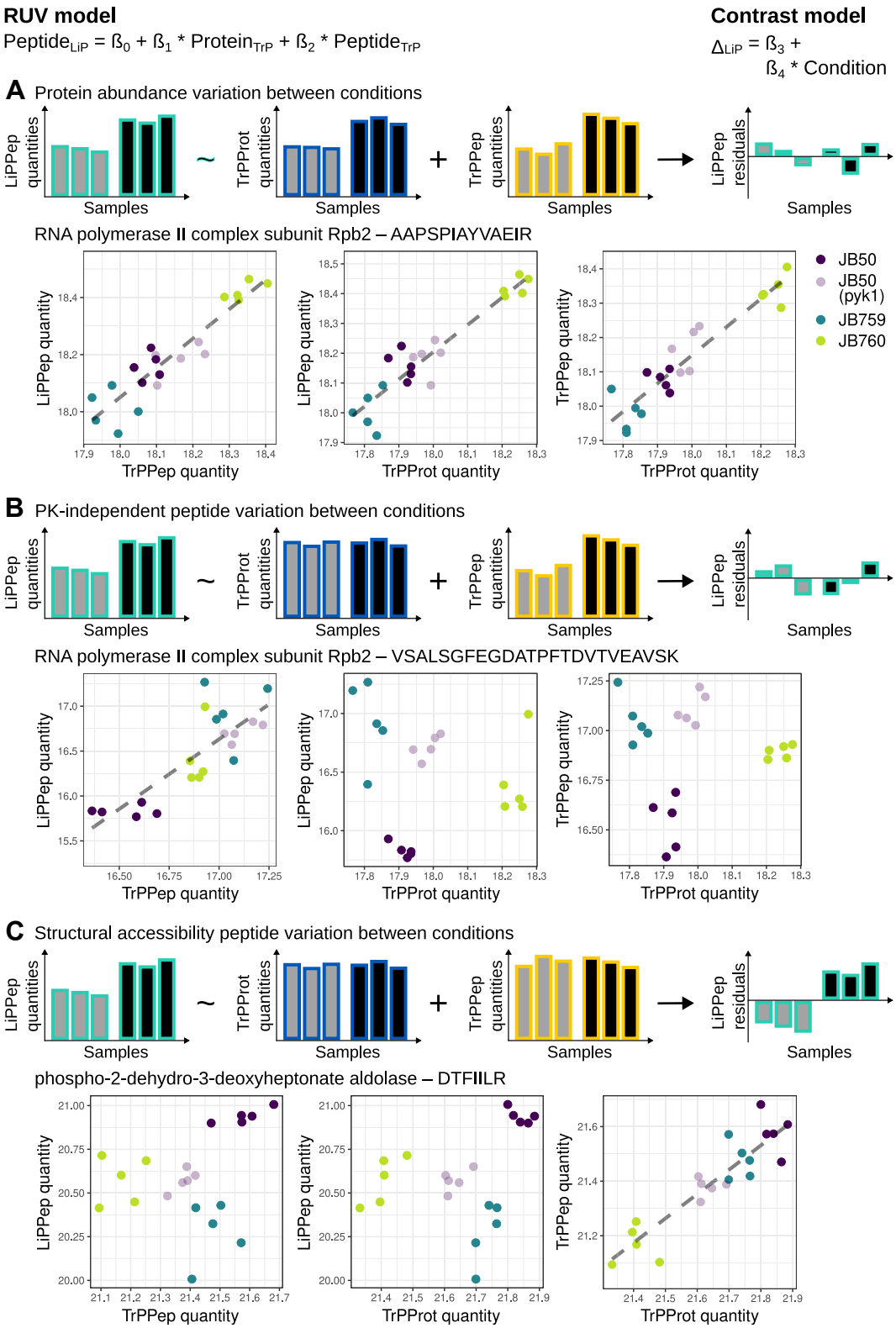


FIG. 3. **Computational approach for inferring accessibility changes between two or more conditions from LiP data.** Different types of variation in LiP data are exemplified with four fission yeast strains (for details, see the following section and methods). *A*, protein abundance variation. *Top*: Schematic overview of LiPAnalyzeR applied to a peptide containing a protein abundance effect. The RUV model removes protein abundance variation from the LiP peptides, and no structural variation is inferred in the subsequent contrast model. *Bottom*: LiP peptide, TrP peptide, and TrP protein quantities of a peptide showing a protein abundance difference between the JB50 and JB759 strain in the fission yeast

other (right), because they are unaffected by the variation in PK digestion. Therefore, signals of protein abundances (usually estimated from several peptides as single peptides are poor estimators of protein abundance (32)) and TrP peptides can be used to correct for biological signal variation of the LiP peptides that is not due to PK accessibility changes. LiPAnalyzeR removes unwanted variation (i.e. corrects for PK-unrelated covariates) by modeling the measured signal of a LiP peptide as a function of its covariates, regressing out the measured signals of those unwanted covariates, and retaining residuals containing the variation of interest. The residuals resulting from this model are utilized to estimate effects that variables of interest (e.g. different conditions) have on the structural protein accessibility in a second regression model.

Thus, LiPAnalyzeR infers structural accessibility variation in two steps (Figs. 2, 3): first correcting for all unwanted covariates (RUV model) and second modeling the effects of conditions of interest on the remaining signal variation (contrast model). The result of the first step is the corrected LiP signal (residual), from which the contribution of unwanted effects has been removed and that is hence independent of protein abundance variation or variation of peptide levels not related to PK-accessibility changes. These residuals can additionally be used for further downstream analysis, such as dimensional reduction or clustering. The RUV model may also be used to correct for additional biological covariates, for example, age, if their effect on structural accessibility variation is not of interest. The contrast model is used to estimate effect sizes, comparable to a fold change, for variables of interest and for obtaining estimates of statistical significance (*p*-values, see Experimental procedures for details).

Datasets Used for Benchmarking LiPAnalyzeR

Multiple datasets were utilized for exploring the analysis of structural accessibility changes from LiP-MS data and benchmarking our approach implemented in LiPAnalyzeR. These datasets encompass various species and experimental setups, including multiple yeast species and human data.

We used datasets from two distantly related yeast species—fission yeast (*S. pombe*) and budding yeast (*S. cerevisiae*). The fission yeast data consists of three different isolates (JB50, JB759, JB760) and a mutant of one of these isolates (JB50 PYK1, see Experimental procedures) (26). For each of the four strains, five biological replicates were measured in matching LiP and TrP samples. The budding yeast dataset consists of 11 biological replicates each of two well-characterized isolates, measured in matching LiP and TrP samples (27). Two technical replicates applying LiP and TrP digestion to the same sample twice were measured for every individual sample (see methods for detail).

We additionally utilized a human CSF dataset containing 52 samples from healthy individuals each measured with LiP and TrP digestion (22). This human data set differs from the yeast data, in that the samples contain more variation due to study variables, such as age, sex, or environment, and have a much more complex and varied genetic background.

LiP Peptide Intensities are Affected by Nonstructural Variation

Signals of almost all LiP peptides are positively correlated with their cognate TrP peptides and TrP proteins (Fig. 4*A*, Supplemental Fig. S1*A*), suggesting a strong contribution of PK-independent signals to LiP-peptide signals. This indicates the necessity to account for these PK-independent signals in the RUV step of LiPAnalyzeR. As the variation in the abundance of a protein is represented more robustly by summarizing the available levels of all TrP peptides than by a single TrP peptide (32), TrP proteins should be used to account for protein abundance effects in the LiP data during the RUV step. However, the signals of some LiP-peptides additionally need to be corrected for peptide-specific PK-independent effects, such as alternative splicing (Figs. 3*B*, 4, *B–E*).

An example of PK-independent peptide variation in our human CSF data (i.e. nonrandom intensity changes in some but not all of the TrP peptides of a protein) is alternative splicing that occurs in the haptoglobin-related protein (HPR, Uniprot ID: P00739). HPR has an alternative isoform (P00739-2) and additionally a very high sequence similarity with haptoglobin (HP, Uniprot ID: P00738) and its alternative isoform (P00738-2). All peptides present in the main isoform of HPR are used to visualize a peptide-specific signal caused by alternative splicing (Fig. 4*C*, top plot). The protein sequence has been annotated into five regions (R1-R5), each of which represents a contiguous segment of the protein that is covered by detectable peptides. Note that these protein regions are unrelated to the exon-intron structure of the genes as they purely result from the technical detectability of the peptides. Most HPR peptides (R2-R5) are present in all four isoforms of HPR and HP, but four peptides of region 1 (R1, residues 58–72) are absent in isoform two of HP (Fig. 4*C*, top

data (*gray dashed* line visualizes linear regression). Strains are plotted in different colors (JB50: *purple*, PYK1 mutant: *light purple*, JB759: *blue*, JB760: *green*). *B*, PK-independent peptide variation. *Top*: Schematic overview of LiPAnalyzeR applied to a peptide with a PK-independent peptide variation. The RUV model removes PK-independent peptide variation from the LiP peptides, and no structural variation is inferred in the subsequent contrast model. *Bottom*: LiP peptide, TrP peptide, and TrP protein of a peptide showing a PK-independent peptide variation between the JB50 and JB759 strain in the fission yeast data. All colors as in (*A*). *C*, structural accessibility variation. *Top*: Schematic overview of LiPAnalyzeR applied to a peptide with a structural accessibility variation reflected only in the LiP peptides. The RUV model does not account for the structural accessibility variation, hence the signal is still reflected in the resulting residuals and detected by the contrast model. *Bottom*: LiP peptide, TrP peptide, and TrP protein of a peptide showing variation in the structural accessibility between the JB50 and JB759 strain in the fission yeast data. All colors as in (*A*).

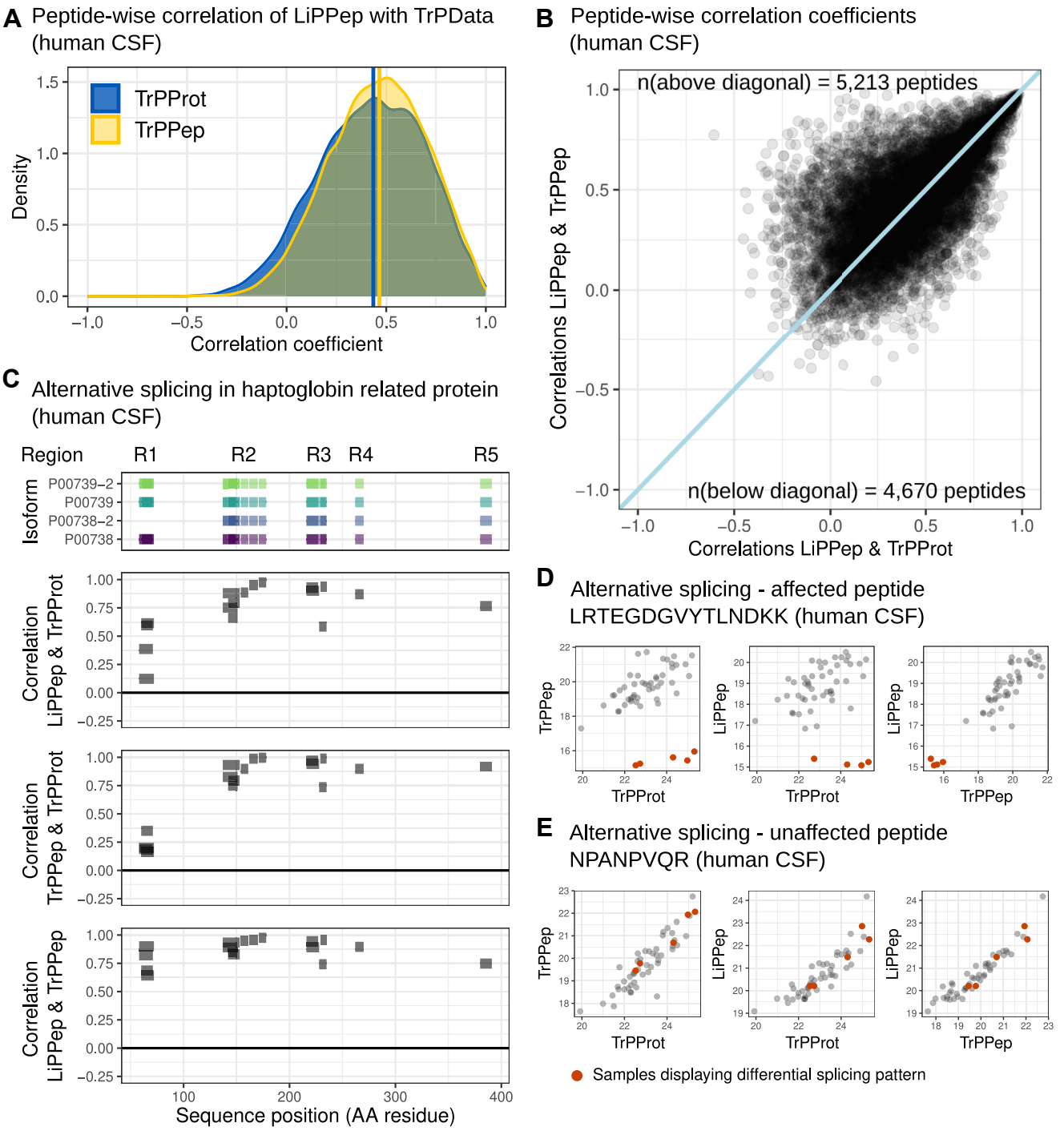


FIG. 4. **Investigation of PK-independent effects in LiP peptides, TrP peptides, and TrP protein quantities.** *A*, peptide-wise Pearson correlation coefficients of LiPPep quantities with the TrPPep (*yellow*: median = 0.466, *p*-value <0.0001) or TrPProt (*blue*: median = 0.435, *p*-value <0.0001) quantities from the human CSF data. *p*-values were estimated with the Wilcoxon signed rank test (alternative hypothesis: true location is not equal to 0). *B*, correlation coefficients from (*A*) plotted against each other. A line going through the origin with a slope of 1 is added (*light blue*). *C*, alternatively spliced haptoglobin-related protein (P00739) visualized across LiP and TrP data. All peptides belonging to the main isoforms of P00739 and occurring in LiP and TrP data are visualized along the protein residues of isoforms of haptoglobin and the haptoglobin-related protein (*top*). Peptide-wise Pearson correlation coefficients between LiPPep & TrPProt (second from *top*), TrPPep & TrPProt (third from *top*), and LiPPep &TrPPep (*bottom*) are visualized along the protein residues. The protein is divided into five regions (R1-R5), where R1 is affected by alternative splicing. *D*, quantities of an example peptide from the haptoglobin-related protein affected by alternative splicing (R1). Samples with alternative splicing in R1 are *red*. *E*, quantities of an example peptide from the haptoglobin-related protein showing no effect by alternative splicing (R3). Samples with alternative splicing in R1 are *red*.

plot). Pearson correlation coefficients were estimated for every peptide between LiP/TrP peptide and TrP protein signals across all samples (Fig. 4*C*). Strong positive correlations were observed for all peptides from R2-R5 (mean correlations: LiP peptide to TrP protein: 0.84, TrP peptide to TrP protein: 0.90), but not for peptides from R1 (mean correlations: LiP peptide to TrP protein: 0.43, TrP peptide to TrP protein: 0.23). TrP and LiP peptides located in R1 exhibit similar deviation from the estimated protein abundances, while being highly correlated with each other (mean correlation of LiP peptide to TrP peptide R1: 0.76). This suggests that they are commonly affected by alternative splicing. Investigating the behavior of the individual samples revealed that, for samples from five donors, peptides from region R1 showed LiP and TrP peptide signals that were lower than expected based on the estimated protein abundance as opposed to peptides from the unaffected regions R2-R5; we show examples of affected (Fig. 4*D*, red samples show an altered splicing pattern in R1) and unaffected peptides (Fig. 4*E*). This indicates that these five donors expressed lower levels of the alternative isoform of HPR (P00739-2) compared to the other donors. Protein abundance was dominated by peptides that are common to all isoforms and hence do not reflect the alternative splicing signal. Excluding peptide-specific PK-independent signal variation from the RUV step would have falsely identified a PK-accessibility difference between the two groups of donors for peptides from region R1.

These examples from fission yeast (Fig. 3*B*) and human (Fig. 4) LiP-MS datasets underline the importance of correcting for both protein- and peptide-specific—but PK-independent—variation in the LiP-MS signals. Further, the extent to which those corrections influence conclusions about specific PK effects depends on the dataset at hand. We hypothesize that accounting for peptide-specific variation in addition to protein abundance variation in the RUV step is more relevant for human samples than yeast samples. This is based on the fact that human proteomes are characterized by greater complexity than yeast proteomes, where especially alternative splicing substantially increases the diversity of proteoforms originating from the same gene. Indeed, TrP peptide signals and protein abundance were about equally correlated with LiP peptide signals in the human data, with the TrP peptides showing a slightly greater average correlation (Fig. 4*B*). As opposed to that, in case of the fission yeast data, LiP peptides signals were more closely correlated with the protein abundances, than with the matching TrP peptide signals (Supplemental Fig. S1*B*). These PK-independent effects influence the signal of both half-tryptic and full-tryptic peptides but are here exemplified only on full-tryptic peptides.

Strategies for Removing Unwanted Variation from LiP-MS Data

Based on biochemical principles, the PK-independent part of the LiP signal variation should be perfectly correlated with the corresponding TrP signal variation. Based on this, one should be able to correct for the PK-independent contribution by computing ratios of LiP and TrP signals (LiP/TrP ratios). In reality however, both the LiP and TrP signals are subject to measurement noise, which reduces the actual correlation between them. Simply computing LiP/TrP ratios would be suboptimal in that noise from the TrP measurements would be added to the LiP signal. The regression-based approach used by LiPAnalyzeR addresses this problem by computing the empirical association between the two measures (LiP and TrP peptide signal) for each peptide individually, thereby accounting for peptide-specific noise. To demonstrate this limitation of LiP/TrP ratios, we have correlated LiP/TrP ratios with the respective TrP signals (i.e. with the matching TrP peptide and the protein abundance estimated from the TrP peptides (TrPProt); Fig. 5). If the LiP/TrP ratios were corrected for the PK-independent signal variation included in both the LiP peptides and TrP measurements, the LiP/TrP ratios should show little correlation with the TrP signals. This however is not the case. Instead, many LiP/TrP ratios show strong negative correlations with the TrP peptide and TrP protein signals (Fig. 5, *A* and *B*). This bias is particularly strong for peptides where a significant structural accessibility effect is inferred based on LiP/TrP ratios, but not based on the RUV approach of LiPAnalyzeR (Supplemental Fig. S1, *C* and *D*; Supplementary text). Hence, computing ratios "corrects" for variation in the TrP data that is uncorrelated with the LiP measurements, resulting in false signals.

The RUV model of LiPAnalyzeR corrects the signal for TrP peptide and TrP protein. In principle, the TrP peptide signal is affected by all relevant PK-independent signal variations, that is, including variation of protein abundance and peptide effects. Therefore, it should be sufficient to correct for all PK-independent effects. However, LiPAnalyzeR includes protein abundance as an additional covariate, because single peptides are poor estimators of protein abundance (32). Integrating the information of several peptides from the same protein reduces noise in the data. Thus, PK-independent signal variation in the LiP measurements that is mainly due to protein abundance variation is generally better corrected by using aggregate measures of protein abundance based on multiple peptides. Correcting only for TrP protein without considering TrP peptide signal would not be sufficient to remove PK-independent peptide variation from the LiP signal. To demonstrate these points, LiP peptides were corrected by running the RUV regression of LiPAnalyzeR including either TrP peptide or TrP protein signals, but not both. Pearson's correlation coefficients were then computed between the residuals from these reduced RUV models and the respective TrP data type not used for the correction. If correcting only for TrP peptides or proteins alone is sufficient to correct for all PK-independent effects, the resulting residuals should be independent of the TrP data type not used in the RUV step. This, however, is not the case. Instead, LiP signals corrected using

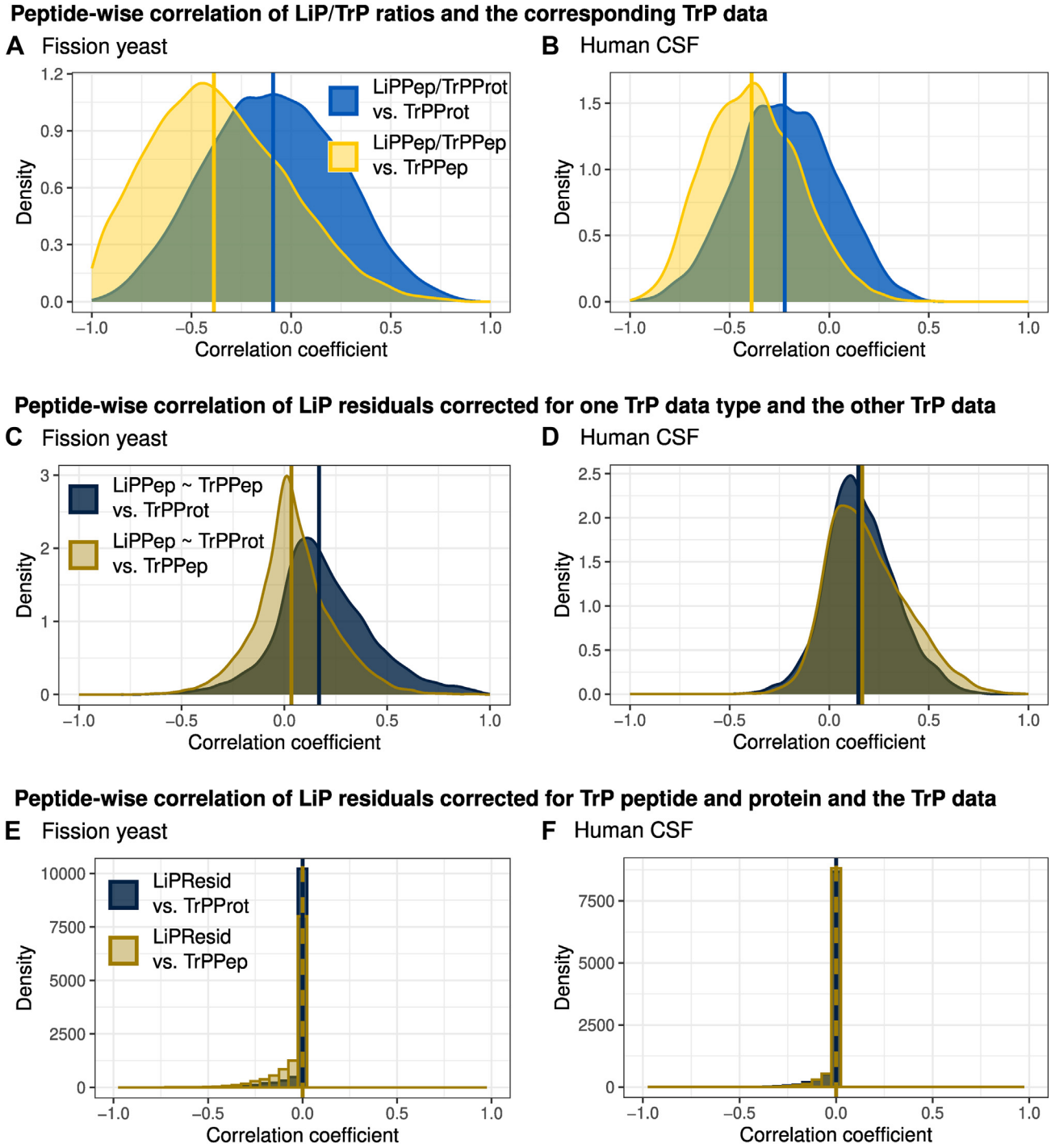


FIG. 5. **Removing unwanted PK-dependent variation from the LiP signal.** *p*-values were estimated with the Wilcoxon signed rank test (alternative hypothesis: true location is not equal to 0). *A* and *B*, peptide-wise Pearson's correlation coefficients between the ratio of LiP peptides to TrP peptides and TrP peptide quantities (*yellow*) as well as between the ratio of LiP peptides to TrP proteins and the TrP protein quantities (*blue*) in (*A*) fission yeast (*blue*: median = −0.090, *p*-value <0.0001; *yellow*: median = −0.388, *p*-value <0.0001) and (*B*) human CSF data (*blue*: median = −0.224, *p*-value <0.0001; *yellow*: median = −0.389, *p*-value <0.0001). *C* and *D*, peptide-wise Pearson's correlation coefficients between the residuals of the RUV step run regressing out TrP peptide or TrP protein signals from LiP peptide quantities and TrP peptide or protein quantities. Residuals estimated from models using TrP peptides as a variable are correlated against TrP protein quantities (*dark blue*) and residuals of those models with TrP proteins are correlated against the TrP peptide quantities (*dark yellow*) in both (*C*) fission yeast (*dark blue*: median = 0.168, *p*-value <0.0001; *dark yellow*: median = 0.034, *p*-value <0.0001)) and (*D*) human CSF data (*dark blue*: median = 0.146, *p*-value <0.0001; *dark yellow*: median = 0.166, *p*-value <0.0001). *E* and *F*, residuals estimated from RUV models using both TrP peptides and proteins as

the reduced RUV models still retain substantial positive correlation with TrP protein or TrP peptide signal (Fig. 5, *C* and *D*). In the case of fission yeast (Fig. 5*C*, median Pearson's correlation coefficient = 0.034), correcting only for protein abundance, but not for TrP peptides, induces a smaller bias than the human data (Fig. 5*D*, median Pearson's correlation coefficient = 0.166). This is in line with the earlier observation that the human data is more strongly affected by PK-independent peptide effects such as alternative splicing.

Applying the complete RUV model—including TrP peptide and protein quantities as covariates—to the LiP peptide completely removes bias in the overall correlations of the residuals for the vast majority of peptides (Fig. 5, *E* and *F*). However, a small negative bias remains due to the non-negative constraints LiPAnalzeR applies to the coefficients estimated for the TrP peptide and protein correction (see next section). Taken together, these results demonstrate that 1) a simple ratio approach leads to artifactual signals and 2) correcting for both TrP peptide signals and protein abundance is necessary.

Superior Performance of Applying a Constrained RUV Model Prior to the Contrast Model

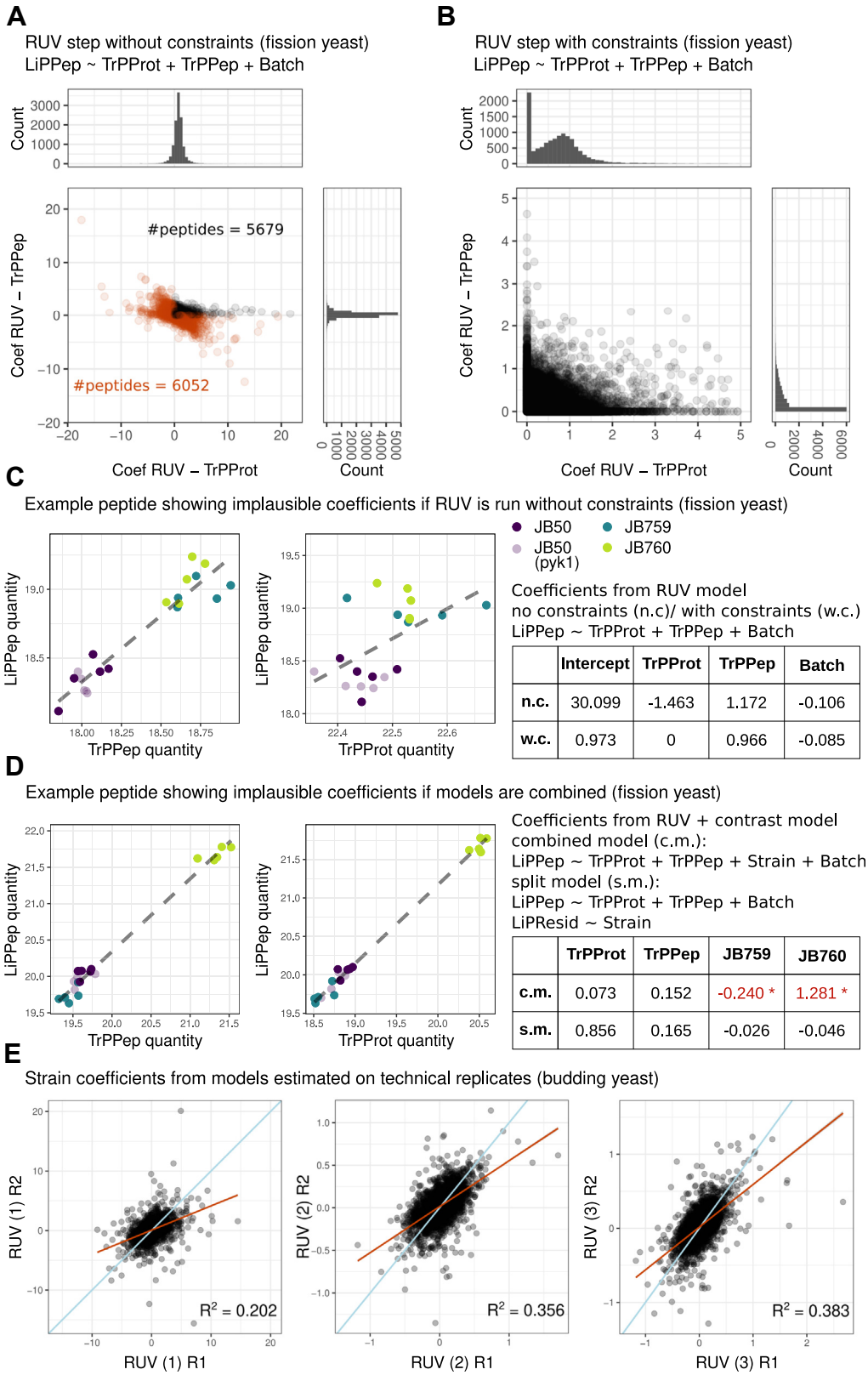
LiPAnalyzeR constrains the coefficients of the TrP peptide and protein signal to be non-negative to reduce overfitting. Negative coefficients would be biochemically implausible, since it can be expected that PK-independent effects affect the TrP and LiP data in the same direction, which is supported by the strong positive correlation of those data (Fig. 3*A*, Supplemental Fig. S1*A*). For instance, it would not be plausible that an increase in the abundance of a protein would result in a decrease in the signal of one of its LiP peptides. When LiP peptide signals are modeled without constraints, many of the TrP peptide and protein coefficients become negative (Fig. 6*A*, Supplemental Fig. S2*A*). Importantly, in the vast majority of cases, only one of the two coefficients is negative, which strongly suggests overfitting, where a too extreme positive coefficient estimated for one of the TrP data is "balanced" by a negative coefficient for the other.

In an example peptide from the fission yeast dataset for this type of overfitting, almost all variation in the LiP peptide signal could be explained by the TrP peptide signal, whereas the protein abundance shows a poor but still clearly positive correlation with the LiP peptide (Fig. 6*C*). However, an unconstrained RUV model results in a larger positive coefficient for the TrP peptide (1.172) and a negative coefficient for the protein (−1.463), which does not represent the observed behavior of the peptide and is biochemically implausible. Setting constraints to the RUV model results in a coefficient of zero for the protein abundance and estimates the correction coefficient for the TrP peptide to be very close to 1, which is biochemically plausible. Interestingly, the constrained RUV model sets one of the two coefficients at or close to zero for the vast majority of peptides (Fig. 6*B*, Supplemental Fig. S2, *B–D*). Thus, for most peptides, either the TrP peptide or the protein abundance is selected, but rarely both. Noticeably, more LiP peptides are TrP peptide driven in the human CSF data than in the fission yeast data (Fig. 6*B*, Supplemental Fig. S2*B*).

LiPAnalyzeR first removes unwanted variation (RUV step) and subsequently estimates effects for variables of interest (contrast step). In principle, it is possible to estimate the effects of unwanted covariates (such as PK-independent variation) together with the effects of interest in a single model, combining the RUV model and the contrast model. While technically feasible, this approach results in an increase in the number of false positive structural accessibility changes. For instance, the average protein abundance often depends on the genotype (33), hence the variable of interest (e.g. genotype) is confounded with a variable that is not of interest (e.g. protein abundance). As a consequence, LiP peptides from this protein are correlated with both genotype and protein abundance, even if the folding of the protein remains unaffected by the genotype. Combining both variables in a single model can result in modeling the protein abundance signal in the genotype (here: strain) variable, as it correlates with the LiP peptide signal. A combined model falsely assumes an effect of the genotype on the LiP peptide instead of correctly estimating the signal of protein abundance variation (Fig. 6*D*). Applying the two-step approach first completely removes the protein abundance effect and subsequently no genotype effect is inferred in the contrast model, resulting in a more conservative approach (Fig. 6*D*).

We demonstrate that the two-step constraint regression approach chosen for LiPAnalyzeR leads to more reproducible results by systematically exploring three different modeling options: 1) combining the RUV and contrast step into one model (Equation 5, no constraints to the model), 2) running RUV without constraints and subsequently the contrast model on the resulting residuals (Equations 1–4, no constraints in model 1), and 3) running RUV with constraints and subsequently the contrast model on the resulting residuals (Equations 1–4, constraints in model 1 as described in method section). We then used the consistency of strain coefficients estimated in the contrast models from different sets of replicates as a quality measure for the models: if the coefficients are more similar between replicates, the modeling approach is more robust and less prone to overfit. These approaches were applied to both technical and biological replicates of the

variables correlated against TrP peptide (*dark yellow*) and TrP protein (*dark blue*) in (*E*) fission yeast (*dark blue*: median = 0, *p*-value <0.0001; *dark yellow*: median = 0, *p*-value <0.0001) and (*F*) human CSF data (*dark blue*: median = 0, *p*-value <0.0001; *dark yellow*: median = 0, *p*-value <0.0001).



	Intercept	TrPProt	TrPPep	Batch
n.c.	30.099	-1.463	1.172	-0.106
w.c.	0.973	0	0.966	-0.085

	TrPProt	TrPPep	JB759	JB760
c.m.	0.073	0.152	-0.240 *	1.281 *
s.m.	0.856	0.165	-0.026	-0.046

FIG. 6. **Comparison of different regression approaches applied to remove TrP signal from LiP peptide quantities.** *A*, peptide-specific coefficients for TrPPep and TrPProt from RUV models without constraints applied to fission yeast data. Peptides with at least one negative coefficient are displayed in *orange*. Number of peptides per quadrant: *top left* = 4320, *top right* = 5679, *bottom left* = 224, *bottom right* = 1525. *B*,

budding yeast dataset to compare the stability of the inferred structural variation signals (see 'Experimetal procedures' for details). The strain effects estimated using the combined model were an order of magnitude larger than those resulting from the two-step approach and less consistent between replicates (Fig. 6*E*, Supplemental Fig. S3, *A–C*). The correction coefficients for TrP peptide and protein abundance were less consistent when using a combined model (Supplemental Fig. S3, *A–C*). Further, constraining the TrP peptide and protein coefficients in the RUV model results in higher similarities between replicates than an unconstrained RUV model. This further supports the notion from above that an unconstrained model results in overfitting. The comparison of results from these three approaches applied to the fission yeast data led to similar outcomes (Supplemental Fig. S4, *A–F*).

Taken together, these results demonstrate that using a constrained RUV model to remove unwanted variation and applying a contrast model in a second step leads to more conservative effect estimates and reduces overfitting.

Including Half-Tryptic Peptides can Improve Coverage and Sensitivity

The PK digestion of the LiP protocol produces many half-tryptic peptides. Since these half-tryptic peptides result from the PK digestion, they almost exclusively exist in the LiP samples, but not in the TrP samples. Half-tryptic peptides from the LiP samples have the potential to greatly increase the coverage of the proteins and hence the resolution of the LiP-MS assay, which motivates their inclusion in the analysis. For example, in the human CSF study, we analyze 5020 half-tryptic peptides along with 9982 full-tryptic peptides. However, since these PK-induced half-tryptic peptides do not exist in the trypsin-only control samples, half-tryptic LiP peptides, unlike full-tryptic LiP peptides, do not have a sequence-identical corresponding TrP peptide. Unfortunately, there is no trivial association of half-tryptic peptides from the LiP samples and matching full-tryptic peptides from the Trp control samples due to the complex patterns in which half-tryptic and full-tryptic peptides can overlap (Fig. 7*A*).

One strategy to deal with this ambiguity is to exclude half-tryptic peptides from the analysis, which leaves a part of the data unused. To maximally use the available data, LiPAnalyzeR also implements a conservative, data-driven approach for the identification of the best matching control peptide in the TrP samples specifically for half-tryptic peptides. First, LiPAnalyzeR correlates the signal of a half-tryptic LiP peptide with all full-tryptic peptides from the TrP data of the corresponding protein(s). Subsequently, the TrP peptide with the highest positive correlation is used in the RUV model to correct the signal of the corresponding half-tryptic LiP peptide (Fig. 7*B*, Supplemental Fig. S5*A*). All other aspects, such as correcting for protein abundance and applying the contrast model to the model residuals, are performed in the same way as for full-tryptic peptides. Using the most strongly correlated peptide for this correction introduces a sampling bias and may result in overcorrecting the signal of the half-tryptic peptide. If many full-tryptic peptides have been measured for a protein in the TrP condition, one of them may correlate well with a half-tryptic LiP peptide by chance alone. Indeed, half-tryptic LiP peptides correlate on average more strongly with the selected, highest correlating TrP peptide than with the corresponding TrP protein (Supplemental Fig. S5, *B* and *C*). By shrinking the residuals to zero, this approach reduces both the false positives, but also the true positive calls and hence is conservative. To visualize this aspect, we have run LiPAnalyzeR with and without using the most strongly correlating peptide during the RUV step (Supplemental Fig. S5, *D–F*). Adding the most correlated TrP peptide to the RUV model results in fewer half-tryptic peptides having significant effects, although the overall directionality of the signal (i.e. the coefficients) remains the same (Supplemental Fig. S5, *D–F*). It is also possible that PK-independent effects present in the half-tryptic peptide, such as a PTM, are not reflected in any full-tryptic peptide and therefore cannot be accounted for. Further, half-tryptic peptides tend to be more variable than full-tryptic peptides (Supplemental Fig. S5*G*), which could reduce the sensitivity for detecting statistically significant effects. Despite these limitations, half-tryptic peptides may still add useful information, especially when combined with the results from full-tryptic peptides from the same protein or protein region.

For example, if only a single full-tryptic peptide in a protein shows a change in PK accessibility, a half-tryptic peptide in

peptide-specific coefficients for TrPPep and TrPProt from RUV models with constraints (coefficients of TrPPep and TrPProt ≥ 0) in fission yeast data. *C*, LiP peptide quantities plotted against TrP peptide (*left*) and TrP protein (*middle*) quantities for the peptide RALIDSPCSEFPR from 60s ribosomal protein L14. Different fission yeast strains are displayed in different colors (JB50: *purple*, PYK1 mutant: *light purple*, JB759: *blue*, JB760: *green*). Coefficients from the RUV step without constraints (n.c.) and with constraints (w.c., coefficients of TrPPep and TrPProt ≥ 0) are displayed (*right*), all of them have a corresponding *p*-value >0.05. *D*, LiP peptide quantities plotted against TrP peptide (*left*) and TrP protein (*middle*) quantities for the peptide GLPLEAVTTIAK from dihydroxyacetone kinase Dak1 (Colors as in c)). Coefficients from a combined modeling (c.m) approach including the RUV and contrast in one step and from a split modeling (s.m.) approach where RUV and contrast modeling is applied separately are displayed (*right*). Coefficients with a *p*-value <0.05 are *red*. *E*, peptide-wise coefficients for strain effects estimated on the technical replicates of budding yeast data using different modeling approaches: (1) combining the RUV and contrast step into one model (Equation 5, no constraints to the model), (2) running RUV without constraints and subsequently the contrast model on the resulting residuals (Equations 1–4, no constraints in model 1) and (3) running RUV with constraints and subsequently the contrast model on the resulting residuals (Equations 1–4, constraints in model 1 as described in method section). A regression line (*red*) and a line going through the origin with a slope of one is added (*light blue*) is shown in each plot.

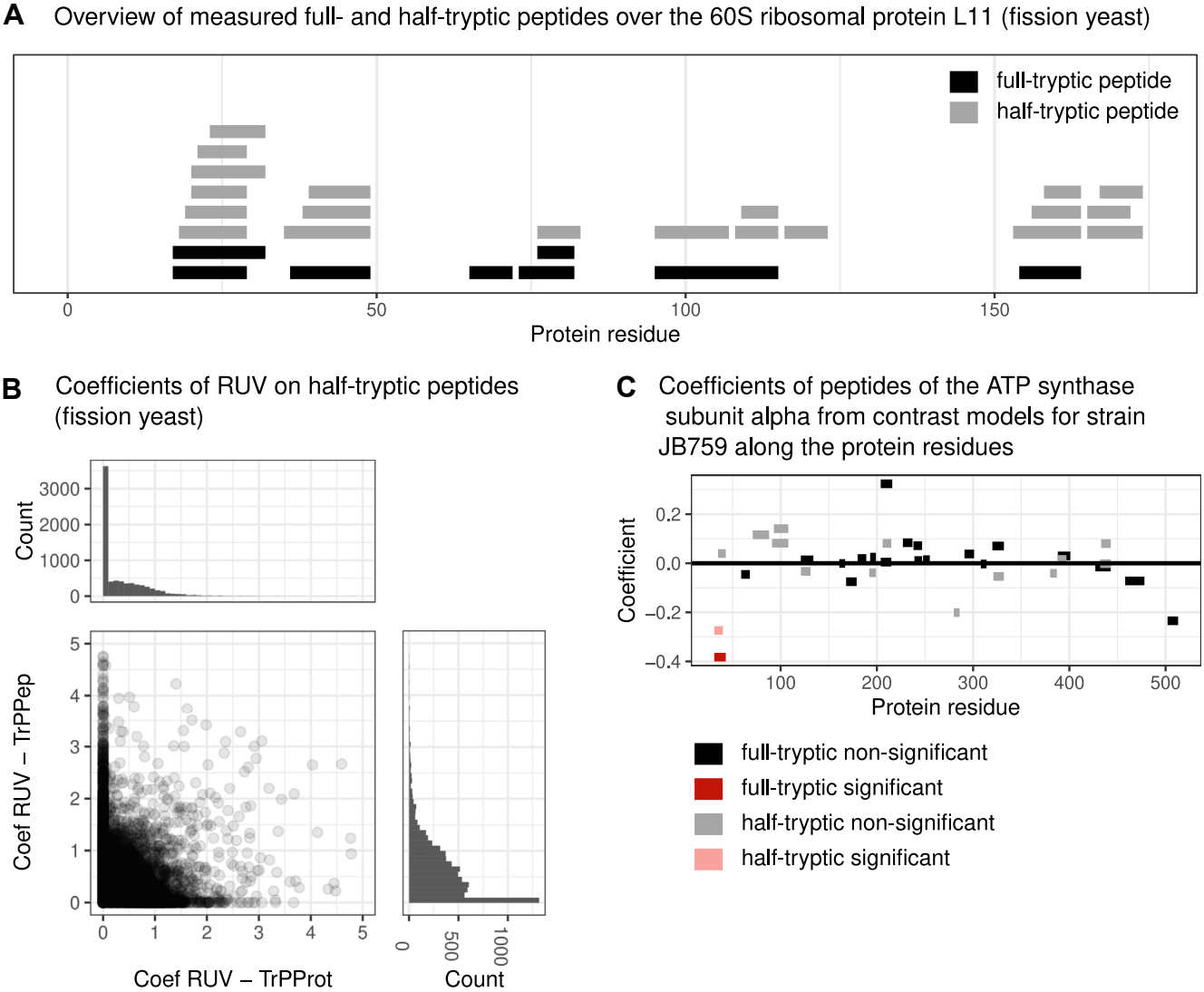


FIG. 7. **Inferring changes in the structural accessibility of half-tryptic peptides.** *A*, full-tryptic (*black*) and half-tryptic (*gray*) peptides detected in the 60S ribosomal protein L11 in fission yeast plotted along the protein residues. *B*, peptide-specific TrPPep and TrPProt coefficients estimated by the RUV models with constraints (coefficients of TrPPep and TrPProt ≥ 0) in fission yeast. *C*, strain coefficients estimated by the contrast models for full-tryptic and half-tryptic peptides of the ATP synthase subunit alpha in the JB759 strain using JB50 as the reference level. RUV for the full-tryptic peptides (*dark*) was performed with the default LiPAnalyzeR pipeline, while half-tryptic peptides (*light*) were analyzed in the HT-only modus. Significant peptides (i.e., those with a protein-wise FDR-corrected *p*-value <0.05 estimated from the contrast models) are displayed in *red*. *D*, strain coefficients estimated by the contrast models for full-tryptic and half-tryptic peptides of the phospho-2-dehydro-3-

the same region that also shows a structural signal can corroborate the finding (Fig. 7*C*). In the case of the phospho-2-dehydro-3-deoxyheptonate aldolase in the fission yeast data, a strong structural effect spanning large regions of the protein can be detected in the full-tryptic peptides between JB50 and JB759, but not between JB50 and JB760 (Fig. 7*D*). The same pattern can be observed just as strongly in the half-tryptic peptides, with there being numerous significantly changing peptides in JB759, but none in JB760. Thus, the half-tryptic peptides add more confidence to the observation that this aldolase is structurally different between JB50 and JB759, but not between JB50 and JB760.

In conclusion, LiPAnalyzeR offers three strategies for dealing with half-tryptic peptides: 1) completely excluding them from the analysis, 2) correcting peptide signals using both protein levels and the most strongly correlated TrP peptide, and 3) only correcting for protein levels, but not PK-independent peptide variation. While option 1) is clearly the most conservative approach, it ignores large parts of the data. Option 3) likely results in the best overall coverage of proteins while running the risk of more false-positive discoveries. Option 2) provides a conservative compromise by utilizing half-tryptic peptides to increase protein coverage and at the same time minimizing false-positives. Which approach is optimal will depend on the specific research question at hand.

Working Without Trypsin-Only Control Samples

Trypsin-only control measurements are essential for correcting for peptide-specific PK-independent effects. However, without such measurements, it is still possible to correct for protein-level effects. Without trypsin-only control measurements, protein abundances may be estimated from the respective LiP measurements, assuming that the majority of peptides of a protein are not affected by structural changes (Supplemental Fig. S5*F*). The correction for protein abundance variation in the RUV step is then performed using protein abundances estimated from the LiP data (LiPProt) while neglecting the correction of PK-independent peptide variation.

We first ran the RUV step on the fission yeast and human CSF data 1) with TrP peptide and protein quantities and 2) with LiP proteins only. Comparing the residuals from these two models showed a significantly higher correlation between the residuals in the fission yeast data than the human CSF data (Fig. 8*A*). This was expected, since we observed PK-independent peptide variation to be more prominent in human CSF as compared to yeast samples before. We subsequently estimated strain effects with the contrast model, using the residuals estimated from RUV alternatives described above and obtained a high similarity between them (Pearson's correlation coefficients of the strain coefficients estimated by LiPAnalyzeR: JB759 = 0.899, JB760 = 0.851, JB50 PYK1 mutant = 0.926). Therefore, using LiP proteins only, when no trypsin-only control is available, seems to be a reasonable proxy for the default RUV models where trypsin-only digested peptides (TrPPep) and TrPProt are included in the model.

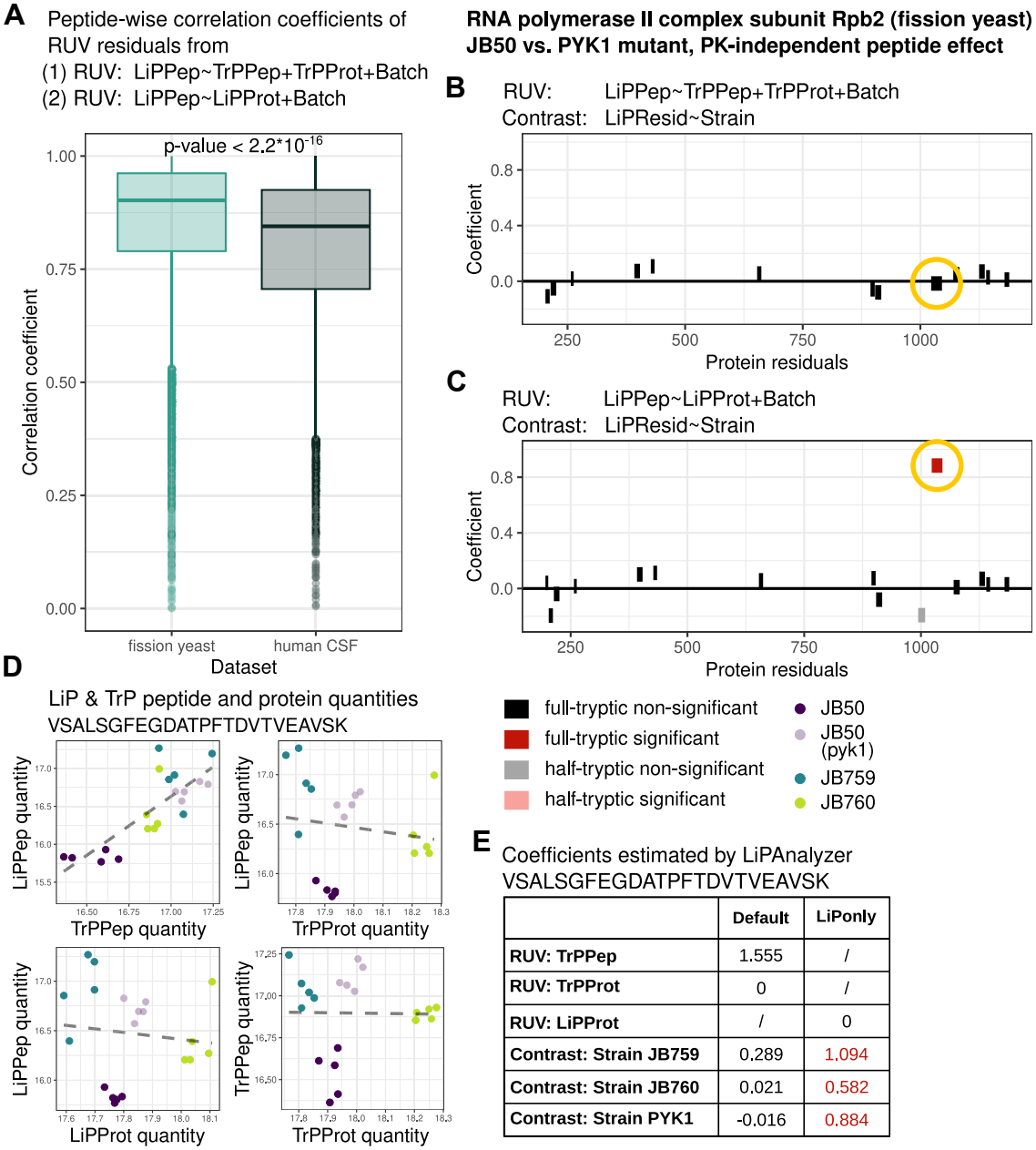
Running the RUV model utilizing the TrP measurements of the RNA polymerase II complex subunit Rpb2 (RPB2) in the fission yeast data leads to no strain-dependent difference in structural accessibility being inferred by the subsequent contrast model (Fig. 6, *B* and *E*). On the contrary, if the RUV step is run using only the LiP protein estimates, the contrast model infers a structurally accessibility variation between the strains for the peptide VSALSGFEGDATPFTDVTVEAVSK (Fig. 6, *C* and *E*). This is due to a PK-independent peptide effect which is reflected in both the LiP and TrP peptide measurements but not in protein abundance signals (Fig. 6*D*). Therefore, not including the TrP peptide into the model prevents the RUV model from accounting for this effect, resulting in the contrast model inferring a false-positive structural accessibility signal.

Utilizing LiP protein abundances instead of TrP measurements for correcting the LiP peptides is a valid option if no TrP data is at hand. However, this approach is limited by 1) the inability of correcting for PK-independent effects on peptides and 2) the presence of large scale effects, affecting a large proportion of LiP peptides, such as protein aggregation, thus affecting the estimation of LiP protein levels. How important the removal of PK-independent peptide variation is, depends on both the research question and the individual dataset/origin species.

Quantitative Assessment of Modeling Approaches

After having demonstrated the need to correct LiP data for protein abundance and PK-independent peptide variation, we wanted to quantify the consequences of including and excluding these covariates using the budding yeast data as an example, because we had the largest number of replicates in this dataset. First, we analyzed the consequences of separating the RUV step from the contrast model (Fig. 9*A*, left). Splitting the model, that is, first correcting for protein and peptide variation before computing the contrast between the BY and RM strains, reduced the number of significant peptides almost than fivefold (3397 significant peptides in the single-step model versus 685 peptides in the split model). Since effects on LiP peptide quantities are inevitable confounded with effects on protein abundance and TrP peptide quantities, we assume that most of these effects are actually false positives, implying that the split model drastically reduces the possible number of false positives. Next, we

deoxyheptonate aldolase in the JB759 (*left*) and JB760 strain using JB50 as the reference level. RUV for the full-tryptic peptides was performed with the default LiPAnalyzeR pipeline, while half-tryptic peptides were analyzed in the HT-only mode. Colors as in (c).



	Default	LiPonly
RUV: TrPPep	1.555	/
RUV: TrPProt	0	/
RUV: LiPProt	/	0
Contrast: Strain JB759	0.289	1.094
Contrast: Strain JB760	0.021	0.582
Contrast: Strain PYK1	-0.016	0.884

FIG. 8. **Inferring structural changes from LiP data without using the trypsin-control measurements.** *A*, peptide-wise Pearson's correlation coefficients between the residuals from the RUV step run with (1) TrP peptide and protein quantities and (2) LiP protein quantities in fission yeast and human CSF data (Student's *t* test: *p*-value <$2.2*10^{-16}$). (Median, center; first and third quartile, *lower* and *upper* hinges; largest/smallest value no further than 1.5 × interquartile range of the hinge, whiskers; data points beyond are defined as outliers and plotted individually). *B*, peptide-specific coefficients of the PYK1 mutant estimated in the RUV step using the default LiPAnalyzeR analysis using JB50 as the reference strain plotted along the protein residuals. Full-tryptic peptides (*dark*) and half-tryptic peptides (*light*) are depicted; peptides with a protein-wise FDR corrected *p*-value <0.05 are displayed in *red*. *C*, as (*B*), but the RUV models were run in the LiP only mode, only correction for LiPProt signal. *D*, LiP and TrP peptide and protein quantities of VSALSGFEGDATPFTDVTVEAVSK plotted against each other. Colors indicate different strains (JB50: *purple*, PYK1 mutant: *light purple*, JB759: *blue*, JB760: *green*). *E*, table with coefficients estimated by LiPAnalyzeR for the fission yeast peptide VSALSGFEGDATPFTDVTVEAVSK with the RUV step using (1) TrP peptide and protein quantities and (2) LiP protein quantities. Coefficients with a significant *p*-value are depicted in *red*.

evaluated the consequences of including and excluding individual covariates (Fig. 9*A*, right). Correcting for either protein abundance or TrP peptide both substantially reduced the number of significant peptides, while correcting for the protein had a bigger effect in this data. In fact, correcting for both protein and TrP peptide only minimally changed the numbers

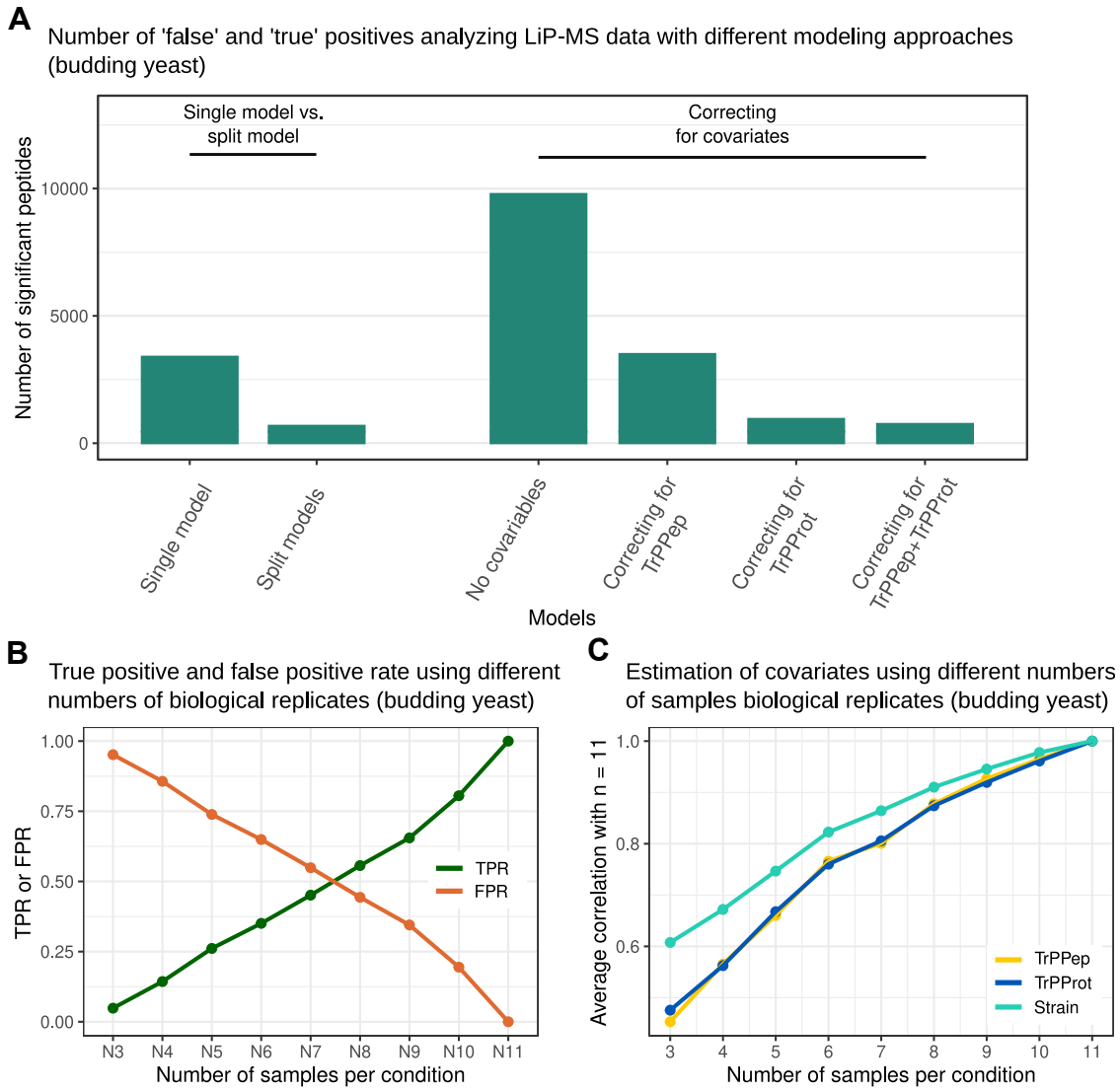


FIG. 9. **Performance comparison of modeling approaches and effect of the number of replicates.** *A*, number of significant peptides (*p*-value <0.05) when analyzing LiP-MS budding yeast data with different modeling approaches. The single versus split model comparison (two *left* models) were run without constraining coefficients to be non-negative. All other models set constraints to the covariates TrP peptide and/or TrP proteins (≥ 0). *B*, effect of varying the number of biological replicates per strain (i.e. sampling N = 3 to N = 11 replicates out of all 11 samples) were analyzed with LiPAnalyzeR using default settings. True positive rate (TPR) and false positive rate (FPR) are plotted, with TP and false positive (FN) peptides defined based on the effects inferred when using all samples (*p*-value <0.05, 11/11 per strain, 22/22 in total). *C*, models as in (*B*) but showing correlation of the coefficients for TrP peptides (*yellow*), TrP protein (*blue*), and strain (*turquoise*) to the coefficients obtained for the whole set of all replicates.

compared to only correcting for the protein levels. This is consistent with the notion that the proteome of budding yeast is comparably simple, devoid of any alternative splicing and hence very few peptide-specific PK-independent effects. However, we advise to correct for both protein and peptide quantities even in such a scenario to minimize the risk of individual false positives that otherwise might be misinterpreted.

Finally, we wanted to evaluate the consequences of having different numbers of replicates. Towards this goal, we sampled different numbers of biological replicates (from 3 to 10) from the budding yeast data. Subsequently, we determined true-positive rates and false-positive rates while assuming that peptides detected with the whole set of all replicate measurements (N = 11) were the "ground truth" peptides with real effects (Fig. 9*B*). As expected, the true-positive rate increased with increasing numbers of replicates, while the false-positive rate decreased. Further, we also correlated the model coefficients obtained with the reduced sets of replicates to those obtained with the full set, to quantify the error in those coefficients resulting from having fewer replicates (Fig. 9*C*). Where we observed some convergence of the model coefficients for higher numbers of replicates, we did

not observe any "saturation effect" of the true-positive and false-positive rates, that is, going from 9 to 10 and from 10 to 11 replicates still substantially improved the sensitivity and specificity of the peptide detection. Thus, we were unable to determine an upper limit of the number of replicates required to saturate the number of peptides with significant strain effects in this particular dataset.

We wish to emphasize that the number of replicates needed depends on several aspects, including the technical quality of the assay, the biological variation expected in the population, and the effect sizes that one aims to detect. Thus, the example calculations for budding yeast shown here do not necessarily apply to other setups.

DISCUSSION

LiPAnalyzeR implements a robust approach to remove unwanted variation from LiP-MS signals and subsequently identify regions in protein structures with accessibility changes. This framework consists of an RUV step that employs constrained regression, followed by a contrast model that estimates effect sizes and corresponding *p*-values representing PK-dependent signal changes. Our work has demonstrated that it is necessary to separate these two steps, as otherwise, effects of confounding variables may be incorrectly split between a LiP effect and a confounding effect, such as protein abundance change. This "effect bleeding" only occurs when two covariates are confounded, such as when batch is confounded with treatment or protein abundance is confounded with LiP effects. Whereas the former can usually be avoided by proper experimental design, the latter is practically impossible to prevent. If a treatment affects only the abundance of a protein, it will also affect the intensity of peptides measured in the LiP assay, even in the complete absence of structural changes. An approach that jointly models the LiP signal as a function of protein abundance and treatment runs the risk of incorrectly splitting the effect between the two, even though there is no effect of the treatment on PK accessibility. This 'effect bleeding' can also occur when analyzing other omics data sets, such as phosphoproteomics data or when inferring alternative splicing effects from MS data. LiPAnalyzeR could additionally be applied to analysis of other types of peptide-centric structural proteomics data, such as FPOP (34) or molecular painting data (35, 36), which are currently emerging as alternatives to LiP for the detection of structurally altered proteins in biological samples.

This work shows that TrP peptides and proteins are mostly positively correlated with the LiP peptides. Protein abundances are generally less noisy than single peptide measurements as they are estimated and averaged over multiple peptides. TrP protein abundance is therefore often a better estimator of the true protein level than the observed matching TrP peptide level and is therefore often preferentially selected by the RUV model for the correction step. If, however, peptide-specific PK-independent effects (such as alternative splicing or PTMs) dominate, our modeling approach will select the matching trypsin-only peptide to correct the LiP-MS signal. Consistent with this notion, we found that TrP peptides were selected much more frequently in the comparably complex human CSF data compared to the simpler fission yeast data. Hence, the importance for including TrP peptides to correct for peptide-specific effects depends 1) on the system (species) being studied and the 2) scientific question being asked.

When removing both TrP peptide and protein signals from the LiP signal in the RUV step, it is crucial to prevent the corresponding coefficients from becoming negative, as negative coefficients would be biochemically implausible. Constraining coefficients within plausible ranges reduces overfitting and increases the reproducibility of the effect estimate. When the RUV step was run without constraints, over half of all peptides in budding yeast had negative coefficients. In contrast, less than one third of the peptides in the human data were affected. This is consistent with the previous observation that peptide-specific effects are less common in the yeast data. If only a protein level effect is present in the LiP data, modeling it as a function of peptide and protein level increases the risk of overfitting. Overfitting can lead to "overcorrection" when attempting to remove the influence of a confounding variable. The approach presented in this work reduces the risk of "overcorrection" by imposing constraints on the RUV model. Using two variables (TrP peptide and TrP protein level) to correct for PK-independent effects is a conservative approach enabling the robust detection of PK-specific effects. Therefore, LiPAnalyzeR can also be run using either protein abundance or peptide level alone in the RUV step if this is deemed necessary for the question at hand.

The trypsin-only control experiment is used to correct for variation in the LiP-MS data that is not due to changes in the PK accessibility. Therefore, these control experiments can be omitted when no PK-independent variation in protein and peptide levels is expected, such as when mapping drug-protein binding using cell lysates (37, 38) or when structurally altered proteins are spiked into cell lysates (Supplemental Fig. S6, see supplementary text). As a consequence, the separate RUV step could be omitted. However, in all other cases, it is advantageous to obtain trypsin-only control data. If TrP data are unavailable, protein abundances estimated from the LiP data may also be used to correct for PK-independent signal variation. The underlying notion is that changes in PK accessibility will rarely affect the entire protein. Therefore, the average signal of all peptides of a protein can serve as a proxy for the PK-independent signal of any specific peptide in the same protein. However, this approach has two disadvantages compared to having a TrP control. It is not possible to correct for peptide-specific PK-independent effects (such as splicing) because there is no independent control measurement for individual peptides. Additionally, the estimation of protein abundance from the LiP data may be flawed due to the

possibility that PK digestion alters the average protein signal. In cases where proteins have only a few detected peptides, the average peptide signal may be biased by LiP effects in only one or two of these peptides. The significance of the TrP control relies on the dataset and research question at hand. If the dataset is anticipated to have more PK-independent peptide-level effects, such as splicing, it is crucial to include trypsin-control samples in the experimental setup. Our comparison of the fission yeast and human datasets confirmed this idea: omitting the TrP controls had a much greater impact on the LiP effect estimates in the human dataset than in the fission yeast data.

The RUV model produces residuals that are adjusted for confounding factors, such as protein abundance or PK-independent effects. This provides information on the structural state of the peptide. Therefore, in addition to estimating effect sizes in a contrast model, these residuals should be used for all downstream analysis focusing on LiP effects, for example, clustering of peptides based on their structural effects across conditions or visualizing structural changes (Supplemental Fig. S7).

LiPAnalyzeR facilitates the analysis of LiP-MS data from diverse experimental setups, thereby addressing a range of biological inquiries. It permits the examination of both categorical (e.g., strain, disease state) and continuous (e.g., concentration of inhibitor, age) variables of interest. The incorporation of technical and/or biological covariates into the RUV step enables the removal of unwanted variation that is confounded with the contrast of interest (e.g., age and sex in mammalian data). Variates can also be added within the contrast step, facilitating the investigation of multiple variables of interest within a single model. This approach also enables the modeling and evaluation of interactions between these variables. As is typical for high-throughput data, LiP-MS data frequently consists of multiple batches, necessitating batch correction. We recommend incorporating batch into the RUV model, regressing out the batch effect in the initial step, but it may also be added to the contrast step instead. Overall, LiPAnalyzeR supports a dataset-dependent adjustment of the covariates added in the RUV and contrast step.

LiPAnalyzeR is a versatile tool that can be applied beyond detecting protein structural effects. LiPAnalyzeR can also be used to detect PK-independent effects on the signal of TrP peptides by modeling TrP peptide signals as a function of protein levels. These effects might be indicative for the presence of alternative splicing or PTMs. As in the analysis of structural effects in LiP peptides, further variables such as batch membership may be included in the RUV model. LiPAnalyzeR also allows to infer differences in protein abundance between conditions using the same modeling approach as for peptide effects. These features enable the querying of a single LiP-MS dataset for a wide range of scientific questions and associating protein structural effects with protein abundance variation and changes in proteoforms.

Our work presents a unified and robust statistical framework for analyzing LiP-MS data. It addresses several challenges specific to this type of data and clearly demonstrates the need to separate effect estimation from (confounding) variable correction. This framework also allows for disentangling protein structural effects while quantifying changes in protein abundance and proteoforms, maximizing biological insight from this rich data.

DATA AVAILABILITY

The fission yeast data are available via ProteomeXchange with identifier PXD055352. The budding yeast data are available via ProteomeXchange with identifier PXD055783.

Code for all analysis performed and the corresponding figures are provided on GitHub: https://github.com/beyergroup/Analysis-of-LiP-MS-data. The R package LiPAnalyzeR can be installed from: https://github.com/beyergroup/LiPAnalyzeR.

Supporting information—This article contains supporting information (39).

Acknowledgments—We thank Natalie de Souza (Institute of Molecular Systems Biology, Department of Biology, ETH Zurich) for valuable input on the manuscript. We thank Jan Lackman (Proteomics Facility, CECAD, University of Cologne) for discussion and support on mass spectrometry data analysis. This project received funding from the European Research Council (grant agreement no. 866004). P. P. received funding from the Peter Bockhoff Stiftung and the ETH Zurich Foundation, Parkinson Schweiz, and the Open Research Data (ORD) Explore program of the ETH Domain. L. N. received funding from the German Academic Exchange Service (DAAD). A. B. received funding from the Deutsche Forschungsgemeinschaft (DFG, CRC 1310, grant agreement no. 325931972 and CRC 1678, grant agreement no. 520471345).

Author contributions—L. N., J. G., V. C., C. D., P. P., and A. B. writing–review and editing; L. N. and A. B. writing–original draft; L. N. visualization; L. N. software; L. N. methodology; L. N. investigation; L. N. formal analysis; L. N., J. G., and A. B. conceptualization; V. C. and C. D. data curation; P. P. and A. B. supervision; P. P. and A. B. resources; P. P. and A. B. funding acquisition.

Conflict of interest—P. P. is a scientific advisor for the company Biognosys AG (Zurich, Switzerland) and an inventor of a patent licensed by Biognosys AG that covers the LiP-MS method used in this protocol. The remaining authors declare no competing interests.

Abbreviations—The abbreviations used are: CSF, cerebrospinal fluid; DDA, data-dependent acquisition; DIA, data-independent acquisition; HP, haptoglobin; HPR, haptoglobin-related protein; LiP-MS, limited proteolysis combined with mass spectrometry; MS, mass spectrometry; PK,

proteinase K; PTM, posttranslational modification; RUV, removal of unwanted variation; TrPPep, trypsin-only digested peptides; TrPProt, protein abundance estimated from the trypsin-only digested data.

Received September 23, 2024, and in revised form, February 14, 2025 Published, MCPRO Papers in Press, March 7, 2025, https://doi.org/10.1016/j.mcpro.2025.100934

REFERENCES

1. Henzler-Wildman, K., and Kern, D. (2007) Dynamic personalities of proteins. *Nature* **450**, 964–972
2. Nussinov, R., Tsai, C.-J., and Jang, H. (2019) Protein ensembles link genotype to phenotype. *PLOS Comput. Biol.* **15**, e1006648
3. Tzeng, S.-R., and Kalodimos, C. G. (2012) Protein activity regulation by conformational entropy. *Nature* **488**, 236–240
4. Feng, Y., De Franceschi, G., Kahraman, A., Soste, M., Melnik, A., Boersema, P. J., *et al*. (2014) Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044
5. Savitski, M. M., Reinhard, F. B. M., Franken, H., Werner, T., Savitski, M. F., Eberhard, D., *et al*. (2014) Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784
6. De Souza, N., and Picotti, P. (2020) Mass spectrometry analysis of the structural proteome. *Curr. Opin. Struct. Biol.* **60**, 57–65
7. Kaur, U., Meng, H., Lui, F., Ma, R., Ogburn, R. N., Johnson, J. H. R., *et al*. (2018) Proteome-wide structural biology: an emerging field for the structural analysis of proteins on the proteomic scale. *J. Proteome Res.* **17**, 3614–3627
8. Schopper, S., Kahraman, A., Leuenberger, P., Feng, Y., Piazza, I., Müller, O., *et al*. (2017) Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **12**, 2391–2410
9. Malinovska, L., Cappelletti, V., Kohler, D., Piazza, I., Tsai, T. H., Pepelnjak, M., *et al*. (2023) Proteome-wide structural changes measured with limited proteolysis-mass spectrometry: an advanced protocol for high-throughput applications. *Nat. Protoc.* **18**, 659–682
10. Cappelletti, V., Hauser, T., Piazza, I., Pepelnjak, M., Malinovska, L., Fuhrer, T., *et al*. (2021) Dynamic 3D proteomes reveal protein functional alterations at high resolution in situ. *Cell* **184**, 545–559.e22
11. Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P. J., von Mering, C., *et al*. (2017) Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* **355**, eaai7825
12. [preprint] Granato, D. C., Normando, A. G. C., Carnielli, C. M., Trino, L. D., Busso-Lopes, A. F., Câmara, G. A., *et al*. (2023) Conformational changes in saliva proteome guides discovery of cancer aggressiveness related markers. *biorxiv*. https://doi.org/10.1101/2023.08.04.552034
13. Liu, F., and Fitzgerald, M. C. (2016) Large-scale analysis of breast cancer-related conformational changes in proteins using limited proteolysis. *J. Proteome Res.* **15**, 4666–4674
14. Sztacho, M., Šalovská, B., Červenka, J., Balaban, C., Hoboth, P., and Hozák, P. (2021) Limited proteolysis-coupled mass spectrometry identifies phosphatidylinositol 4,5-bisphosphate effectors in human nuclear proteome. *Cells* **10**, 68
15. Backe, S. J., Votra, S. D., Stokes, M. P., Sebestyén, E., Castelli, M., Torielli, L., *et al*. (2023) PhosY-secretome profiling combined with kinase-substrate interaction screening defines active c-Src-driven extracellular signaling. *Cell Rep.* **42**, 112539
16. Khanppnavar, B., Schuster, D., Lavriha, P., Uliana, F., Özel, M., Mehta, V., *et al*. (2024) Regulatory sites of CaM-sensitive adenylyl cyclase AC8 revealed by cryo-EM and structural proteomics. *EMBO Rep.* **25**, 1513–1540
17. Dörig, C., Marulli, C., Peskett, T., Volkmar, N., Pantolini, L., Studer, G., *et al*. (2024) Global profiling of protein complex dynamics with an experimental library of protein interaction markers. *Nat. Biotechnol.*, 1–15. https://doi.org/10.1038/s41587-024-02432-8
18. [preprint] Holfeld, A., Schuster, D., Sesterhenn, F., Stalder, P., Haenseler, W., Barrio-Hernandez, I., *et al*. (2023) Systematic identification of structure-specific protein–protein interactions. *biorxiv*. https://doi.org/10.1101/2023.02.01.522707
19. Paukštytė, J., López Cabezas, R. M., Feng, Y., Tong, K., Schnyder, D., Elomaa, E., *et al*. (2023) Global analysis of aging-related protein structural changes uncovers enzyme-polymerization-based control of longevity. *Mol. Cell* **83**, 3360–3376.e11
20. [preprint] Sui, X., Prado, M. A., Paulo, J. A., Gygi, S. P., Finley, D., and Morimoto, R. I. (2022) Global proteome metastability response in isogenic animals to missense mutations and polyglutamine expansions in aging. *biorxiv*. https://doi.org/10.1101/2022.09.28.509812
21. Shuken, S. R., Rutledge, J., Iram, T., Losada, P. M., Wilson, E. N., Andreasson, K. I., *et al*. (2022) Limited proteolysis–mass spectrometry reveals aging-associated changes in cerebrospinal fluid protein abundances and structures. *Nat. Aging* **2**, 379–388
22. Mackmull, M.-T., Nagel, L., Sesterhenn, F., Muntel, J., Grossbach, J., Stalder, P., *et al*. (2022) Global, in situ analysis of the structural proteome in individuals with Parkinson's disease to identify a new class of biomarker. *Nat. Struct. Mol. Biol.* **29**, 978–989
23. Choi, M., Chang, C. Y., Clough, T., Broudy, D., Killeen, T., MacLean, B., *et al*. (2014) MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **30**, 2524–2526
24. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., *et al*. (2016) The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat. Methods* **13**, 731–740
25. Röst, H. L., Sachsenberg, T., Aiche, S., Bielow, C., Weisser, H., Aicheler, F., *et al*. (2016) OpenMS: a flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **13**, 741–748
26. Kamrad, S., Grossbach, J., Rodríguez-López, M., Mülleder, M., Townsend, S., Cappelletti, V., *et al*. (2020) Pyruvate kinase variant of fission yeast tunes carbon metabolism, cell regulation, growth and stress resistance. *Mol. Syst. Biol.* **16**, e9270
27. Brem, R. B., Yvert, G., Clinton, R., and Kruglyak, L. (2002) Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755
28. Weith, M., Großbach, J., Clement-Ziza, M., Gillet, L., Rodríguez-López, M., Marguerat, S., *et al*. (2023) Genetic effects on molecular network states explain complex traits. *Mol. Syst. Biol.* **19**, e11493
29. Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., *et al*. (2015) Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47
30. Grossbach, J., Gillet, L., Clément-Ziza, M., Schmalohr, C. L., Schubert, O. T., Schütter, M., *et al*. (2022) The impact of genomic variation on protein phosphorylation states and regulatory networks. *Mol. Syst. Biol.* **18**, e10712
31. Cappelletti, V., and Malinovska, L. (2021) Data analysis of LiP-MS data for high-throughput applications. *Zenodo*. https://doi.org/10.5281/zenodo.5749994
32. Carrillo, B., Yanofsky, C., Laboissiere, S., Nadon, R., and Kearney, R. E. (2010) Methods for combining peptide intensities to estimate relative protein abundance. *Bioinformatics* **26**, 98–103
33. Foss, E. J., Radulovic, D., Shaffer, S. A., Ruderfer, D. M., Bedalov, A., Goodlett, D. R., *et al*. (2007) Genetic basis of proteome variation in yeast. *Nat. Genet.* **39**, 1369–1375
34. Li, K. S., Shi, L., and Gross, M. L. (2018) Mass spectrometry-based fast photochemical oxidation of proteins (FPOP) for higher order structure characterization. *Acc. Chem. Res.* **51**, 736–744
35. Chen, M. Z., Moily, N. S., Bridgford, J. L., Wood, R. J., Radwan, M., Smith, T. A., *et al*. (2017) A thiol probe for measuring unfolded protein load and proteostasis in cells. *Nat. Commun.* **8**, 474
36. Luchini, A., Espina, V., and Liotta, L. A. (2014) Protein painting reveals solvent-excluded drug targets hidden within native protein–protein interfaces. *Nat. Commun.* **5**, 4413
37. Piazza, I., Kochanowski, K., Cappelletti, V., Fuhrer, T., Noor, E., Sauer, U., *et al*. (2018) A map of protein-metabolite interactions reveals principles of chemical communication. *Cell* **172**, 358–372.e23
38. Morretta, E., Belvedere, R., Petrella, A., Spallarossa, A., Rapetti, F., Bruno, O., *et al*. (2021) Novel insights on the molecular mechanism of action of the anti-angiogenic pyrazolyl-urea GeGe-3 by functional proteomics. *Bioorg. Chem.* **115**, 105168
39. Bisaglia, M., Trolio, A., Bellanda, M., Bergantino, E., Bubacco, L., and Mammi, S. (2006) Structure and topology of the non-amyloid-β component fragment of human α-synuclein bound to micelles: implications for the aggregation process. *Protein Sci.* **15**, 1408–1416