

# Global variability analysis of mRNA and protein concentrations across and within human tissues

Christine Wegler<sup>1,2,†</sup>, Magnus Ölander<sup>1,†</sup>, Jacek R. Wiśniewski<sup>3</sup>, Patrik Lundquist<sup>1</sup>, Katharina Zettl<sup>3</sup>, Anders Åsberg<sup>4,5</sup>, Jøran Hjeltnes<sup>6,7</sup>, Tommy B. Andersson<sup>2</sup> and Per Artursson<sup>1,8,\*</sup>

<sup>1</sup>Department of Pharmacy, Uppsala University, Uppsala SE-75123, Sweden, <sup>2</sup>DMPK, Research and Early Development Cardiovascular, Renal and Metabolism, BioPharmaceuticals R&D, AstraZeneca, Gothenburg SE-43183, Sweden, <sup>3</sup>Biochemical Proteomics Group, Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Martinsried D-82152, Germany, <sup>4</sup>Department of Pharmacy, University of Oslo, Oslo NO-0316, Norway, <sup>5</sup>Department of Transplantation Medicine, Oslo University Hospital-Rikshospitalet, Oslo NO-0316, Norway, <sup>6</sup>Morbid Obesity Centre, Department of Medicine, Vestfold Hospital Trust, Tønsberg NO-3103, Norway, <sup>7</sup>Department of Endocrinology, Morbid Obesity and Preventive Medicine, Institute of Clinical Medicine, University of Oslo, Oslo NO-0316, Norway and <sup>8</sup>Department of Pharmacy and Science for Life Laboratory, Uppsala University, Uppsala SE-75123, Sweden

Received June 20, 2019; Revised August 27, 2019; Editorial Decision September 16, 2019; Accepted October 06, 2019

## ABSTRACT

**Genes and proteins show variable expression patterns throughout the human body. However, it is not clear whether relative differences in mRNA concentrations are retained on the protein level. Furthermore, inter-individual protein concentration variability within single tissue types has not been comprehensively explored. Here, we used the Gini index for in-depth concentration variability analysis of publicly available transcriptomics and proteomics data, and of an in-house proteomics dataset of human liver and jejunum from 38 donors. We found that the transfer of concentration variability from mRNA to protein is limited, that established ‘reference genes’ for data normalization vary markedly at the protein level, that protein concentrations cover a wide variability spectrum within single tissue types, and that concentration variability analysis can be a convenient starting point for identifying disease-associated proteins and novel biomarkers. Our results emphasize the importance of considering individual concentration levels, as opposed to population averages, for personalized systems biology analysis.**

## INTRODUCTION

Modern transcriptomics and proteomics technologies have enabled the comprehensive mapping of human mRNA and

protein levels in different cells and tissues (1,2). It has been shown that some genes and proteins are expressed in tissue-specific patterns, while others are ubiquitously expressed throughout the body (3–7). However, correlations between mRNA and protein concentrations are typically poor (8,9). Further, it is unclear to what extent the relative concentration differences on the mRNA level are transferred to the protein level (10).

Notable inter-individual differences have been observed within single tissue types on the mRNA level (11,12), but this has not been comprehensively investigated on the protein level. Due to the time-consuming nature of mass spectrometry-based proteomics analysis, large-scale proteomics studies of multiple tissues often use samples from few or sometimes pooled donors (5–7). Further, studies of single tissue types commonly use concentration levels from multiple donors as biological replicates, with little emphasis on inter-individual variability (13,14). However, by using few or average protein concentration levels, the outcome of holistic systems biology analysis would not necessarily reflect inter-individual differences. In-depth knowledge of protein concentration variability in single tissue types would enable systems biology to deliver results that better represent the entire population, thereby advancing the field of personalized medicine (15,16).

We reasoned that in-depth concentration variability analysis could be used to address these two knowledge gaps: the transfer of relative concentration differences from mRNA to protein and the inter-individual protein concentration variability in single tissue types. For this purpose, we first

\*To whom correspondence should be addressed. Tel: +46 18 471 44 71; Email: per.artursson@farmaci.uu.se

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

used publicly available transcriptomics and proteomics data from paired human tissues, and then generated a unique proteomic dataset of liver and jejunum samples from 38 human donors. Our results show that mRNA variability is poorly reflected at the protein level, and that ‘reference genes’ proposed for normalization of ‘omics’ data, with low concentration variability on the mRNA level, had higher variability on the protein level across different tissues and within multiple samples of the same tissue type. Thus, proper normalization would necessitate specific reference genes for different types of samples and data, which would not be practical in proteomics. Furthermore, we show that proteins with essential cellular functions have low protein concentration variability within single tissue types. On the other hand, many proteins vary substantially between donors and we noted a high variability in proteins related to disease, suggesting that protein concentration variability analysis can be used as a starting point for the identification of disease-associated proteins. We also demonstrate that many proteins show large inter-individual concentration variability, with implications for personalized systems pharmacology.

## MATERIALS AND METHODS

### Samples for proteomics

Human liver and jejunum samples were obtained from 38 donors that had provided informed consent as part of the COCKTAIL study (17). Samples were collected from patients undergoing bariatric surgery preceded by a 3-week fasting period. The patients were between 23 and 63 years old, with BMI-values ranging between 30 and 63 kg/m<sup>2</sup>.

### Data origin

For concentration variability analysis across human tissues (Supplementary Data S1), we used mRNA and protein concentrations from 29 paired human tissue types from previously published transcriptomics and proteomics datasets (7). For within-tissue variability analysis on the mRNA level, we used mRNA concentrations from liver (175 donors) and small intestine (137 donors) obtained from the GTEx project (18).

Within-tissue variability analysis on the protein level was performed using proteomics data of the 38 human liver and jejunum samples obtained here. For the proteomics analysis, proteins were solubilized with 2% SDS and processed with the multi-enzyme digestion filter-aided sample preparation (MED-FASP) protocol, using LysC and trypsin (19). Peptides were separated on a C<sub>18</sub> column (50 cm and 75 μm inner diameter) using a 2-h acetonitrile gradient in 0.1% formic acid at a flow rate of 300 nl/min. The LC was coupled to a Q Exactive HF or Q Exactive HF-X mass spectrometer (Thermo Fisher Scientific) operating in data dependent mode with survey scans at a resolution of 60 000, AGC target of 3 × 10<sup>6</sup> and maximum injection time of 20 ms. The top 15 most abundant isotope patterns were selected from the survey scan with an isolation window of 1.4 *m/z* and fragmented with nCE at 27. The MS/MS analysis was performed with a resolution of 15 000, AGC target of 1 × 10<sup>5</sup> and maximum injection time of 60 ms. The

resulting MS data were analyzed with MaxQuant (version 1.6.0.16) (20), where proteins were identified by searching peptides against a decoy version of the UniProtKB (May 2013). Carbamidomethylation was set as a fixed modification, and protein discovery rates were specified as 0.01, allowing a maximum of two missed cleavages. Spectral raw intensities were normalized with variance stabilization (*vs*n) (21). Protein concentrations were calculated with the Total Protein Approach (22).

### Calculation of concentration variability

Concentration variability analysis was performed with the Gini index, using a modified version of the *ineq* R package, version 0.2-13 (<https://CRAN.R-project.org/package=ineq>). More specifically, Gini indices were calculated for each mRNA or protein based on the mRNA or protein concentration levels in the 29 tissues (for the across-tissue analysis) or based on the protein concentration levels in the 38 human liver or jejunum samples (for the within-tissue analysis). Thus, mRNAs or proteins with low Gini indices (close to zero) had similar concentration levels across the 29 tissues or across the 38 donors in the across-tissue and within-tissue analysis, respectively.

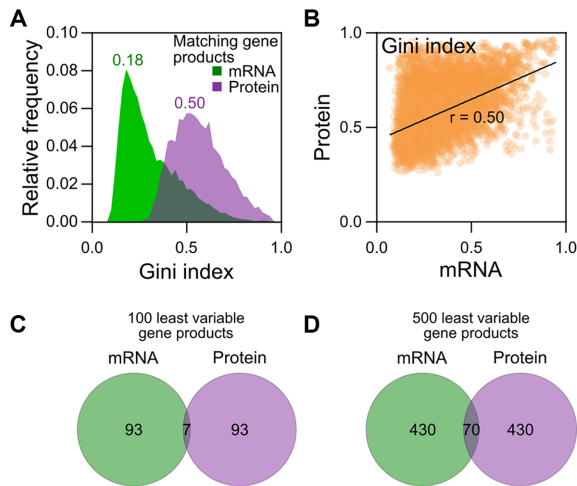
### Statistical analysis

Statistical analyses were performed using GraphPad Prism, version 7.03, and partial least squares (PLS) modeling using SIMCA, version 15.0.0.4783. Functional annotation clustering of GOBP and KEGG terms was performed with DAVID, version 6.8 (23), where clusters were considered significant above an enrichment score of 1.3 (corresponding to a *P*-value < 0.05). Statistical details are provided at relevant places in the ‘Results’ section and figure legends. For the analysis of the proteomics datasets, the first protein ID was selected when multiple protein IDs were separated by ‘;’. This enabled matching between transcriptomics and proteomics datasets.

## RESULTS

### Variability in mRNA and protein concentration levels across different tissue types

The Gini index is a measure of variability in a dataset (24), and was recently introduced for describing the variability of mRNA levels across different tissue types (25). The Gini index ranges from 0 to 1, where lower values indicate similar concentration levels across samples. Here, we extended the variability analysis to the closer-to-phenotype protein level to investigate whether the concentration variability was similar on the mRNA and protein levels. For this purpose, we calculated Gini indices (Supplementary Data S1) using mRNA and protein concentrations in 29 paired human tissue types from a previously published dataset (7). To ensure reliable Gini index calculations, we only included mRNAs and proteins that were detected in at least 15 tissues. The frequency distribution of Gini indices from the proteomics data was shifted toward higher values compared to the transcriptomics data, reflecting an overall higher variability at the protein level. The distribution of protein concentration

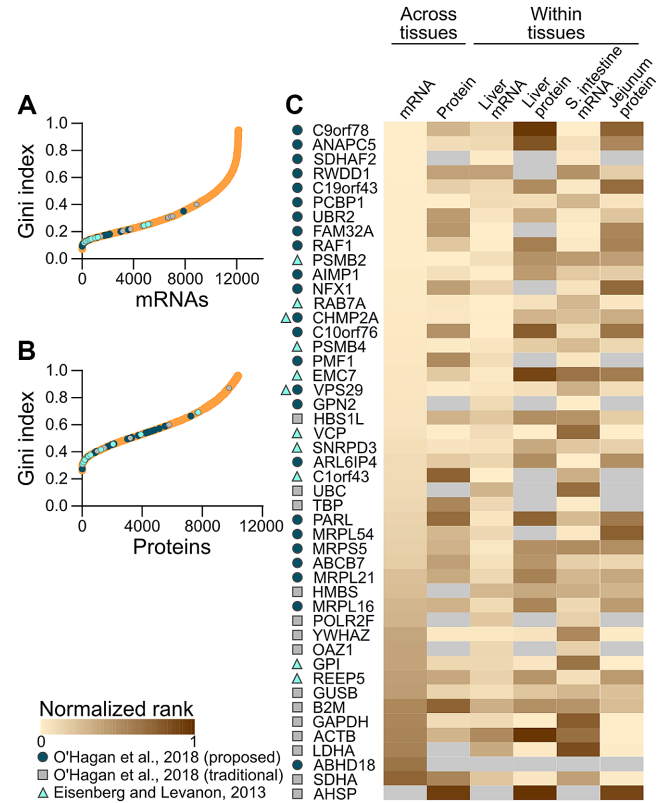


**Figure 1.** Variability in mRNA and protein concentration levels across different tissue types. (A) Concentration variability distributions of matching mRNAs and proteins ( $n = 8828$ ) across 29 paired human tissue types, based on previously published transcriptomics and proteomics data (7). Numbers in figure denote mode (bin width: 0.02). (B) Correlation of mRNA and protein concentration variability ( $n = 8828$ ), using transcriptomics and proteomics data from the 29 tissue types (7). (C) Overlap of the 100 least variable mRNAs and proteins across tissues (7). (D) Overlap of the 500 least variable mRNAs and proteins across tissues (7);  $r =$  Pearson's correlation coefficient.

variability across tissues peaked at a Gini index of 0.50, while the distribution of mRNA concentration variability peaked at 0.18 (bin width 0.02; Figure 1A and Supplementary Data S1). Gini indices were negatively correlated with median concentration levels across the 29 tissues at the protein level ( $r = -0.51$ ; Supplementary Figure S1). However, this was only a general trend, and many highly abundant proteins still had high Gini indices (and vice versa). On the other hand, no such correlation was observed at the mRNA level ( $r = -0.03$ ). Gini indices calculated from the transcriptomics data showed some discrepancies with the corresponding values from the proteomics data ( $r = 0.50$ , Figure 1B). In addition, there was very little overlap among the 100 and 500 least variable gene products between the transcriptomics and proteomics datasets ( $<14\%$ ; Figure 1C and D). The fact that low variability at the mRNA level across different tissues not necessarily results in low variability at the protein level is likely due to translational or post-translational regulatory mechanisms, and presumably reflects tissue-specific phenotypes (26).

**Reference gene variability across different tissue types**

Historically, ubiquitously expressed gene products have been used as references for e.g. western blots (27) and qPCR (28). More recently, the use of reference genes has also been proposed to improve normalization of more complex omics data (29,30). Large-scale transcriptomics studies have provided lists of genes with low mRNA concentration variability as potential references (25,30). We investigated the concentration variability of mRNAs from these proposed genes, as well as from some traditionally used reference genes (25), across different tissue types in the transcriptomics and proteomics datasets. Unsurprisingly, the ma-



**Figure 2.** Reference gene variability in transcriptomics and proteomics data. (A and B) Gini indices of mRNAs and proteins in transcriptomics and proteomics data across 29 tissue types (7). Previously proposed reference genes are highlighted. (C) Normalized ranks of reference genes on the mRNA and protein levels across tissues, as well as across donors in transcriptomics data from 175 liver and 137 small intestine samples (18) and in proteomics data of liver and jejunum from 38 donors. Ranks were normalized by the number of entries in each separate dataset, and sorted by the rank in the transcriptomics data across tissues.

majority of the newly proposed references (23 of 33) were among the 10% least variable mRNAs in the transcriptomics dataset (Figure 2A and C). On the other hand, the traditionally used references (e.g. GAPDH and ACTB) generally showed much higher variability, possibly due to the diverse reasons for their selection.

To assess the utility of references selected from transcriptomics data for normalization at the protein level, we studied the variability of the 33 proposed references (25,30) in the proteomics dataset across multiple tissue types. We found that protein concentrations of the proposed references were highly variable across the different tissue types, with Gini indices covering  $\sim 75\%$  of the entire variability spectrum (Figure 2B and C). This high variability on the protein level questions the suitability of selecting reference genes for proteomics based on transcriptomics information.

**Inter-individual reference gene variability in single tissue types**

Differences in protein concentrations across different tissue types, resulting from tissue specialization, might preclude the universal application of transcriptomics-based reference

genes. However, the reference genes could still have a low concentration variability (and thus be useful for normalization) within a more homogeneous sample group, such as multiple samples of the same tissue type from different donors. Therefore, we analyzed reference gene variability on the mRNA level in liver (175 donors) and small intestine (137 donors) using previously published transcriptomics data (18), and on the protein level in liver and jejunum samples from 38 human donors (Supplementary Data S1). Similarly to the observation across different tissue types, many of the proposed references (14 of 33 from the liver and 19 of 33 from the small intestine) were among the 10% least variable mRNAs (Figure 2C; Supplementary Figure S2a and S2b). On the protein level, a few of the references were among the 10% least variable (5 in both liver and jejunum). However, the overall variability was, surprisingly, as high as in the multi-tissue proteomics dataset (Figure 2C). This indicates that caution should be exercised if selecting reference genes for normalization of proteomics data is based on transcriptomics information.

#### Variability in protein concentrations in single tissue types

Inspired by this unexpectedly high protein concentration variability in single tissue types, we performed more in-depth analysis of within-tissue variability in the proteomes of the 38 liver and jejunum samples. For the calculation of Gini indices, we included proteins that were detected with at least three unique peptides. The final datasets comprised 5968 and 7662 proteins for liver and jejunum, respectively. The variability distributions were significantly lower than in the across-tissue proteomics data, with peaks at Gini indices of 0.12 and 0.16 in liver and jejunum, respectively, compared to 0.54 across tissues ( $P < 0.0001$ , one-way ANOVA followed by Tukey's multiple comparisons test; Figure 3A and Supplementary Data S1). Using previously published transcriptomics data (18), we observed relatively similar within-tissue variability distributions on the mRNA level compared to the protein level, with peaks at Gini indices of 0.20 and 0.18 in liver and small intestine, respectively (Supplementary Figure S3a and S3b). In general, the variability distribution of protein concentrations in the jejunum samples was shifted toward higher values compared to liver ( $P < 0.0001$ , one-way ANOVA followed by Tukey's multiple comparisons test). More specifically, Gini indices ranged between 0.01–0.84 and 0.04–0.94, with medians of 0.21 and 0.24 in liver and jejunum, respectively. Gini indices from the two tissues showed a relatively poor correlation ( $r = 0.48$ ; 5496 matching proteins). However, when excluding a relatively small number of the most discrepant values between the two datasets (469 proteins), we observed a much stronger correlation ( $r = 0.71$ ; Figure 3B and Supplementary Figure S4A). This shows that the overall variability across donors is similar in liver and jejunum.

We further investigated the least variable proteins in liver and jejunum, and found that 34% and 53% of the 100 and 500 least variable proteins, respectively, were overlapping between the two tissue types (Figure 3C). This demonstrates that proteins with low inter-individual variability in liver often also have a low variability in jejunum, suggesting that different tissue types share many proteins that

require concentration levels within narrow intervals. Biological functions represented by the 34 proteins with the lowest Gini indices in both tissue types included protein folding (CCT2, CCT8, HSP90AA1, HSP90B1, HSPA8 and PDIA3), energy metabolism (HADHA, MDH2, ENO1, PGK1, ATP5A1 and ATP5B) and vesicular trafficking (CLTC, COPB2, DYNC1H1, GDI2 and RAB11B).

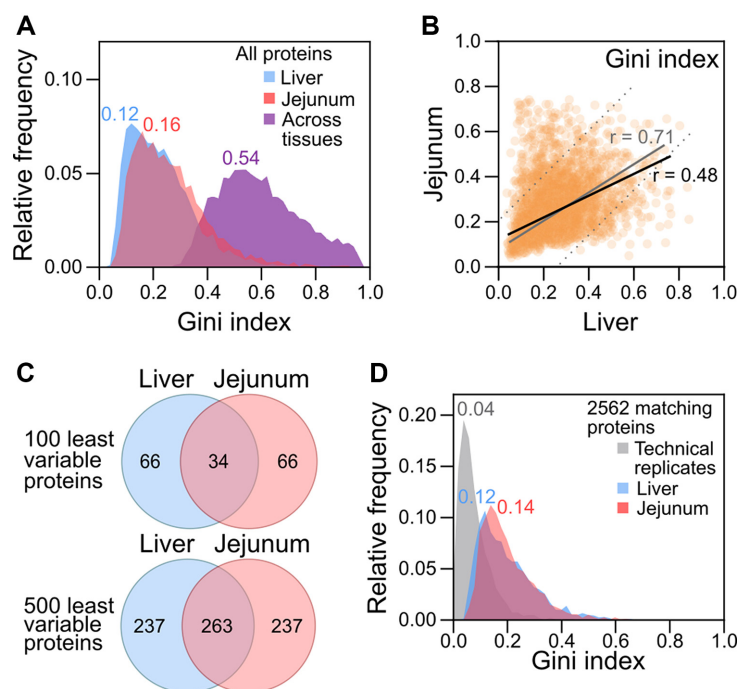
To verify the biological relevance of our findings, and to exclude that the variability was introduced by measurement errors, we compared the Gini indices from the liver and jejunum proteomics data with a previous dataset of Caco-2 cells analyzed in triplicate (31). In theory, the cell samples should have Gini values close to 0 for all proteins, as they were replicates cultured together under identical conditions. For this comparison, we used the 2562 proteins that were detected with at least three peptides in all three datasets. Indeed, the replicate samples had significantly lower Gini indices, with a median value of 0.04 compared to 0.12 and 0.14 in liver and jejunum, respectively ( $P < 0.0001$ , one-way ANOVA followed by Dunnett's multiple comparisons test; Figure 3D). Interestingly, however, some proteins showed relatively high variability even among the replicates. In contrast to the least variable proteins (below the 5th percentile), the most variable proteins (above the 95th percentile) in the replicate samples were over-represented by nucleic acid binding proteins and transcription factors (Supplementary Figure S4B). Nevertheless, the generally low variability among the replicates shows that the Gini indices we observed in the liver and jejunum samples mainly resulted from biological differences and not from technical measurement errors.

We further validated the biological relevance of our analyses by comparing Gini indices in the 38 liver samples with a separate in-house proteomics dataset containing 15 liver samples from another patient group. A stronger correlation was found between the two liver datasets than between the different tissue types (i.e. the 38 liver and jejunum samples; Supplementary Figure S4C). Thus, the results in this study should not be specific to particular datasets.

#### Within-tissue variability in relation to protein properties

We next investigated the characteristics of proteins from the entire within-tissue variability spectra in liver and jejunum. Proteins were classified by type using the PANTHER classification system (32). Median Gini indices of the protein classes from liver and jejunum were highly correlated ( $r = 0.87$ ), showing that similar types of proteins had low and high variability in both tissues (Figure 4A). For example, lyases, oxidoreductases, and chaperones were among the least variable protein classes, while cell adhesion molecules and extracellular matrix proteins showed high variability.

Further, we used the Prolocate database of protein localization in rat liver (33) to study the subcellular distribution of the 2095 proteins that matched between our liver dataset and Prolocate. In general, proteins with high Gini indices were more widely distributed across subcellular compartments (Figure 4B). The cytosol was the major compartment for 46% of the 2095 proteins, followed by mitochondria (17%) and the endoplasmic reticulum (ER; 16%; Figure 4C). We then specifically considered the 10% least and



**Figure 3.** Variability in protein concentrations in single tissue types. **(A)** Variability distributions of proteins in human liver ( $n = 5968$ ) and jejunum ( $n = 7662$ ) from 38 donors, and in proteomics data across 29 tissues ( $n = 10373$ ) (7), comparing within-tissue and across-tissue variability. Numbers in figure denote the mode (bin width: 0.02). **(B)** Correlation of protein concentration variability between the liver and jejunum datasets. The black line shows the regression with all proteins included, and the grey line shows the regression after exclusion of highly discrepant proteins (outside the dotted lines) between the two tissues. **(C)** Overlap of the 100 and 500 least variable proteins in the liver and jejunum datasets. **(D)** Variability distributions of 2562 matching proteins in Caco-2 cells, analyzed as triplicates, and human liver and jejunum, comparing technical and biological variability. Numbers in figure denote the mode (bin width: 0.02);  $r =$  Pearson's correlation coefficient.

the 10% most variable proteins. Here, differences were seen for other compartments than the cytosol between the two extremes (Figure 4C and Supplementary Figure S5A). The mitochondria and ER were the major compartments for many of the least variable proteins. On the other hand, the plasma membrane (PM) and lysosome were major compartments for the most variable proteins. This is logical in light of the fasting period that the patients undergo prior to surgery (see ‘Materials and Methods’ section), as fasting is known to trigger an increase in the number and activity of lysosomes (34).

We then used the DAVID database (23) to characterize the biological functions of the least and most variable proteins. The 100 least variable proteins represented similar basal cellular functions in both tissues, such as carbohydrate metabolism, protein processing and translation (Figure 4D and Supplementary Data S2). The 100 most variable proteins were more random, with few enriched functions. However, extending the analysis to the 500 most variable proteins resulted in interesting findings. Enriched functions included extracellular matrix organization in the liver, possibly due to varying levels of collagen deposition related to liver status, as well as drug metabolism in the jejunum, which could be related to induction of intestinal metabolic enzymes by dietary and environmental factors (Supplementary Figure S5B and Data S2).

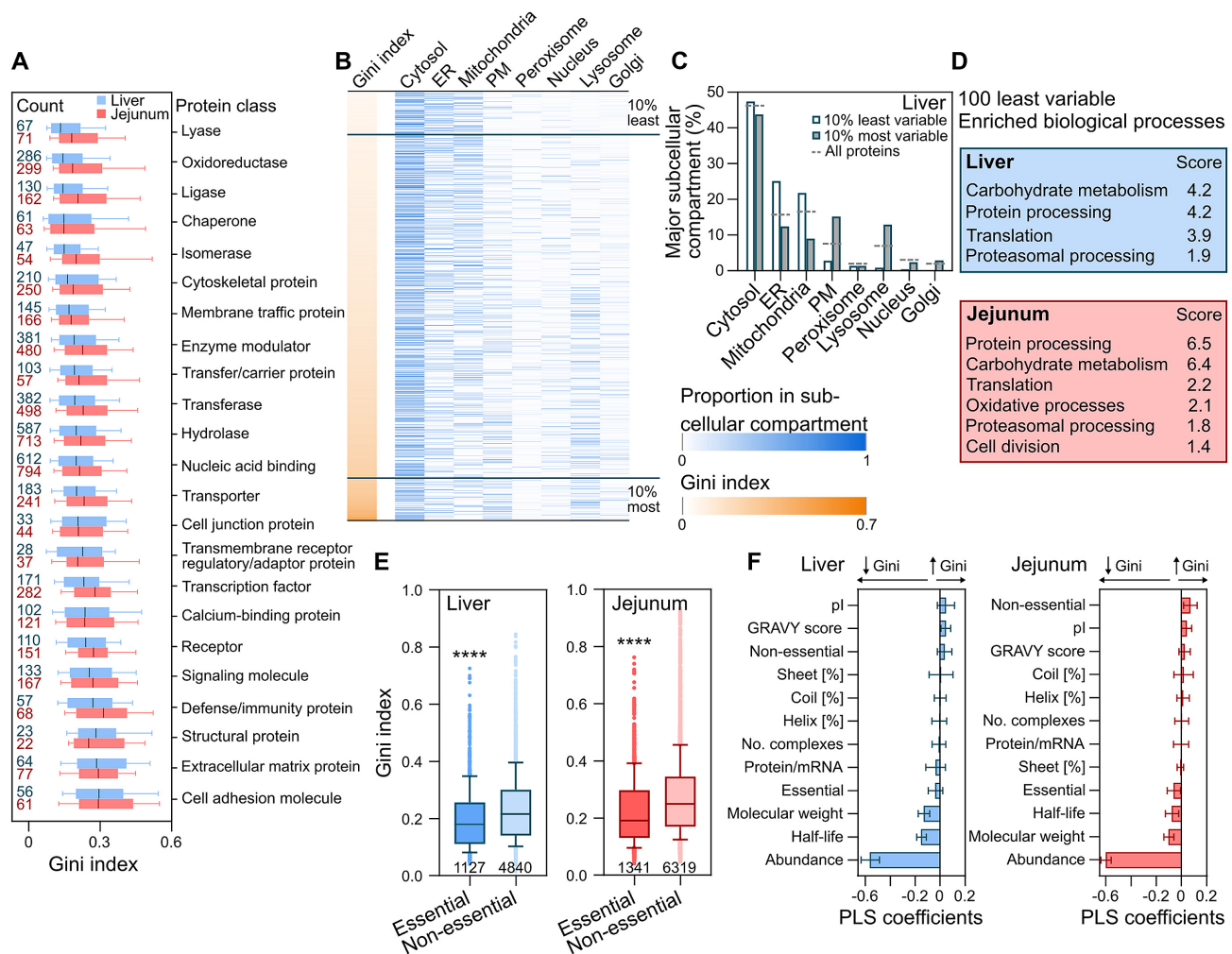
To further probe the involvement of the least variable proteins in basal functions, we compared proteins from the ‘core essentialome’ (35), i.e. proteins essential for cell sur-

vival, with non-essential proteins. Indeed, essential proteins had significantly lower Gini indices in both tissues ( $P < 0.0001$ , Student's  $t$ -test), showing that these proteins were found at similar concentrations in all donors (Figure 4E). This is logical, as cells need to maintain these proteins at certain levels to ensure survival. The corresponding mRNAs to these essential proteins also had significantly lower Gini indices across donors in the external transcriptomics dataset of liver and small intestinal samples (18) (Supplementary Figure S6A and S6B). Further, similar analysis of the external proteomics dataset containing 15 liver samples from another patient group supported these results (Supplementary Figure S6C).

To provide a summary of factors that influence within-tissue variability, we performed PLS modeling of the effect of various protein properties on Gini indices. We included GRAVY score (hydrophobicity), secondary structure (36), essentiality (35), isoelectric point (37), protein complex participation (38), protein/mRNA ratio (6), turnover rate (39), molecular weight and protein concentration (Supplementary Data S1). The results of the models generally indicated that highly abundant, large, essential, hydrophilic proteins with low turnover rates have low concentration variability in single tissue types (Figure 4F).

### Capturing expected biological variability

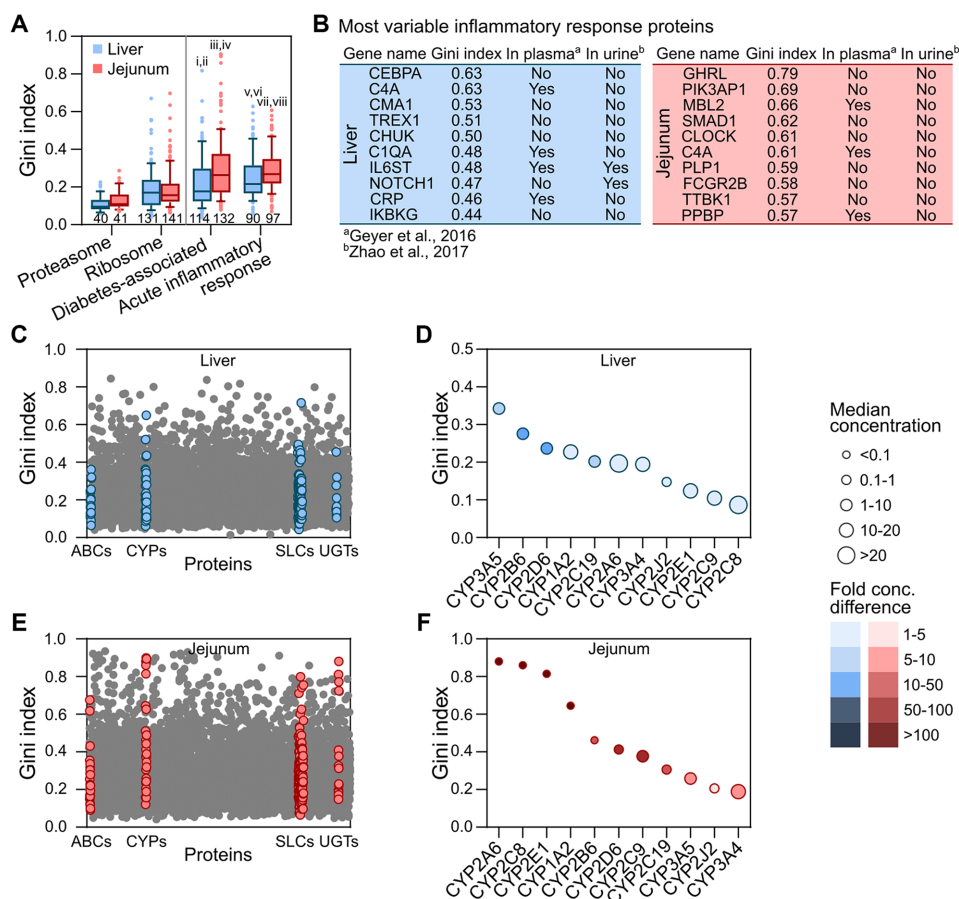
We next investigated whether the Gini index could capture a biological process where high variability in protein con-



**Figure 4.** Within-tissue variability in relation to protein properties. **(A)** Gini indices of proteins in different classes in human liver and jejunum from 38 donors. The number of proteins in each class is shown in the figure. Boxes range between the 25th and 75th percentiles, lines show medians and whiskers denote the 10th and 90th percentiles. **(B)** Subcellular distribution of proteins in human liver, sorted by Gini index. The 10% least and most variable proteins are highlighted. **(C)** Major subcellular compartment of the 10% least and most variable proteins in human liver. Dashed lines show the corresponding values for all proteins. **(D)** Enriched biological processes in the 100 least variable proteins in human liver and jejunum. Scores show the enrichment score of the functional annotation clustering. **(E)** Gini indices of essential and non-essential proteins in human liver and jejunum. The number of proteins is displayed under each boxplot. Boxes range between the 25th and 75th percentiles, lines show medians and whiskers denote the 10th and 90th percentiles. **(F)** The effect of various protein properties on Gini indices, assessed by PLS modeling. Error bars indicate 95% confidence intervals for the model coefficients. \*\*\*\* $P < 0.0001$ , one-way ANOVA followed by Dunnett's multiple comparisons test.

concentrations across donors was expected. For this, we extracted proteins that were associated with diabetes in the DisGeNET database (40) and proteins annotated with the term 'inflammatory response' in the Gene Ontology (41). Diabetes-associated proteins were likely to have variable concentration levels across the 38 donors, since a subgroup of the patients (one-third) suffered from diabetes. Further, the patients had varying degrees of obesity, and inflammation is known to increase with increasing body weight (42). The Gini indices for these proteins were compared with proteins from the proteasome and ribosome, i.e. core cellular structures that were expected to have similar concentration levels across the donors. As expected, diabetes-associated proteins had significantly higher Gini indices than the core structures in both liver and jejunum (one-

way ANOVA followed by Dunnett's multiple comparisons test; Figure 5A). This was further investigated by dividing the donors into two groups: patients with diabetes and patients without diabetes, and calculating the Gini indices for each group (Supplementary Figure S7A). As patients in the respective groups were expected to have similar concentrations of diabetes-associated proteins (e.g. high levels of a protein in patients with diabetes but low levels in patients without diabetes), low Gini indices were expected in each group (Supplementary Figure S7A). Indeed, diabetes-associated proteins had significantly lower Gini indices in both groups compared to the Gini indices obtained when including all 38 donors (one-way ANOVA followed by Holm-Sidak's multiple comparisons test; Supplementary Figure S7B and S7C). Acute inflammatory response proteins also



**Figure 5.** Variability in proteins associated with inflammation, diabetes, and drug metabolism. (A) Gini indices of proteins from core cellular structures, with low expected variability, compared with Gini indices of diabetes-associated and inflammatory response proteins, with high expected variability. The number of proteins is displayed under each boxplot. Boxes range between the 25th and 75th percentiles, lines show medians, and whiskers denote the 10th and 90th percentiles. (B) The ten most variable inflammatory response proteins in liver and jejunum. (C) Gini indices of all proteins quantified in liver samples, ordered by gene name. Important transporter and metabolic enzyme families are highlighted. (D) Important drug metabolizing CYP enzymes quantified in liver samples, sorted by Gini index. (E) Gini indices of all proteins quantified in jejunum samples, ordered by gene name. Important transporter and metabolic enzyme families are highlighted. (F) Important drug metabolizing CYP enzymes quantified in jejunum samples, sorted by Gini index. Roman numerals indicate statistical significance from one-way ANOVA followed by Dunnett's multiple comparisons test, in comparisons of diabetes-associated proteins with proteasome (i, iii;  $P < 0.0001$ ) and ribosome (ii, iv;  $P < 0.05$  and  $P < 0.0001$ , respectively), and comparisons of inflammatory response proteins with proteasome (v, vii;  $P < 0.0001$ ) and ribosome (vi, viii;  $P < 0.001$  and  $P < 0.0001$ , respectively) in liver and jejunum, respectively.

had significantly higher Gini indices than the core structures in both liver and jejunum (one-way ANOVA followed by Dunnett's multiple comparisons test; Figure 5A). These results further demonstrate the applicability of this approach for observing biologically relevant variability. Interestingly, some of the most variable inflammatory proteins have also been detected in plasma (43) and urine (44). Both well-established biomarkers, such as CRP (45), and less investigated proteins were found among the ten most variable inflammatory proteins, indicating that the latter could potentially be used as biomarkers for tissue inflammation (Figure 5B).

### Variability in proteins related to drug metabolism

One progressing branch of systems biology is systems pharmacology, in which biological network structures are used to predict drug action (46). For this type of analysis, it is imperative to know the quantities of proteins involved

in drug disposition. Therefore, we investigated the variability of the most important proteins involved in drug transport and metabolism. These included proteins from the two main transporter families, ATP-binding cassette (ABC) and solute carrier (SLC), and the two most prominent enzyme families in phase I and phase II metabolizing enzymes, i.e. the cytochrome P450s (CYPs) and UDP-glucuronosyltransferases (UGTs) (47–50). In general, these proteins covered large parts of the variability spectrum in both liver and jejunum (Figure 5C and E). More specifically, Gini indices for the CYPs responsible for the majority of drug metabolism (50) ranged from 0.09 to 0.34 and 0.19 to 0.88 in liver and jejunum, respectively (Figure 5D and F). This spread in Gini indices was also reflected by the fold differences of protein concentrations across donors (i.e. the ratio of maximum and minimum values), where the majority of the enzymes (67%) had a fold difference above 10. Fold differences are heavily affected by extreme values,

but provide more tangible information for systems pharmacology analysis. The high variability we observed for many important metabolic enzymes implies that comprehensive systems pharmacology analysis, tailored to the patient, requires thorough consideration of inter-individual protein concentration variability.

## DISCUSSION

Concentrations of mRNAs and proteins show considerable variability across tissues, and it is well known that mRNA and protein levels are poorly correlated. However, little is known about the transfer of relative concentration differences from mRNA to protein. Furthermore, the variability in protein concentrations within single tissue types from multiple donors has not been comprehensively characterized. Here, we first used the Gini index to characterize mRNA and protein concentration variability across tissues using a publicly available dataset. We then investigated inter-individual differences within single tissue types, using in-house proteomics data of 38 human liver and jejunum samples. We found substantial variability differences between transcriptomics and proteomics data across tissues, and identified common themes in the single tissue types among proteins at both ends of the variability spectrum.

We observed a discrepancy in concentration variability between the mRNA and protein levels across tissue types. This could be explained by various factors. Tissue specialization constitutes a logical partial explanation, as different proteins are required at different levels to establish specific phenotypes (4). Furthermore, protein levels are controlled by a complex interplay of post-transcriptional processes. Protein synthesis consumes more energy than mRNA transcription (51,52), meaning that tissues will only produce proteins at levels that are required for their tissue-specific functions, whereas mRNA levels do not need to be as tightly controlled (53). Moreover, regulation of synthesis and degradation rates can also largely affect the variability in protein concentration irrespective of the corresponding mRNA levels (54).

We found that the discrepancy in variability between mRNA and protein levels across tissue types was also prominent among transcriptomics-based reference genes proposed for use in normalization procedures. Further, the references showed different variabilities in proteomics datasets containing 38 samples of human liver and jejunum. In essence, these results mean that optimal normalization would require the selection of different reference genes for transcriptomics and proteomics data, and for different cells and tissues. In fact, specific reference genes might even be necessary for different samples of the same tissue type in different conditions (55). The complexity of this task makes the use of reference genes a highly impractical normalization approach for proteomics data. Instead, more sophisticated normalization methods, such as intensity-based variance stability normalization (vs<sub>n</sub>), should be used (21,56).

In the analysis of our 38 liver and jejunum samples, we observed that protein concentration variability was lower within single tissue types than across different tissues. Variability was generally higher in jejunum than in liver. Demographic differences between sample types could be ruled

out as a cause for the higher variability, since the samples originated from the same 38 donors. Therefore, other explanations need to be considered. The liver is a relatively homogeneous tissue, where hepatocytes constitute 80% of the volume (57), whereas the jejunum is composed of several distinct tissue layers (58). Further, hepatocytes have a comparatively small proteome in relation to other cell types in the liver (59). This could explain why more proteins were detected in the jejunum samples. The layered nature of the jejunum means that the pinch-biopsy technique used for obtaining jejunal biopsies can result in slightly different sample compositions due to variable sampling depth (60). Deeper biopsies would contain higher proportions of collagen-rich submucosal connective tissue, thus diluting the tissue-specific expression of the jejunal epithelium to varying degrees. Indeed, we observed high variability in the major intestinal collagen types (61) between the jejunum samples (Supplementary Figure S5C). Further, inter-individual differences in jejunum length (62) could affect the relative position of the jejunal sampling site (17). This could influence the variability due to regional differences in protein concentrations in the human small intestine (63). Such sampling effects have previously been observed on the mRNA level in human lung tissue (12).

We also note that several technical parameters of the proteomics analysis itself may affect the Gini index. For instance, it is well known that the number of peptides detected for a certain protein affects the reliability of its quantification (64). We did observe a tendency that larger numbers of peptides used for detection resulted in lower Gini indices, but the correlations were relatively weak ( $r = -0.37$  and  $r = -0.35$  for liver and jejunum, respectively; Supplementary Figure S8A and S8B). To partly account for the problem, we only included proteins that were detected with at least three unique peptides. Another problem is represented by plasma membrane proteins, which are often low-abundant, show limited solubilization, and contain hydrophobic domains that are problematic to ionize, which all makes them more difficult to analyze (65). However, in the proteomics workflow used here, the lysis buffer contained high concentration of SDS that facilitates the solubilization of proteins with hydrophobic domains. Nevertheless, the challenge of quantifying low abundant proteins still remains, and is a plausible reason to that plasma membrane proteins were over-represented among the most variable proteins in the liver and jejunum. The difficulties involved in their analysis thus indicate that their Gini indices should be interpreted with a measure of caution.

Even though the protein concentration variability was generally higher in jejunum than in liver, the least variable proteins in each tissue were involved in similar essential cellular processes. The least variable proteins also seemed to capture basal differences in tissue physiology. For instance, the constant proliferation and migration of enterocytes along the villus axis (66) likely accounts for the low variability of proteins related to cell division in the jejunum. This highlights the relevance of using the Gini index for variability analysis. The Gini index also provides a simple approach to identify variability in proteins of importance for systems pharmacology analysis, such as transporters and metabolic enzymes, as well as proteins involved in bi-

ological processes that deviate from the ‘normal’ situation (e.g. healthy controls). For example, we found high variability in proteins involved in diabetes and inflammation, indicating varying degrees of these disease-related processes in the different donors. By identifying highly variable proteins and subsequent comparison of the actual protein levels with disease status in the different patients, the Gini index might provide an unbiased starting point for biomarker discovery in a more general sense. Supporting this, previous studies have shown that hypervariably expressed genes are largely associated with human diseases (67).

In summary, we found that the transfer of concentration variability from mRNA to protein across tissue types is limited, in addition to the already established poor correlation between quantitative levels of mRNA and protein. Our analysis also indicates that specific, rather than universal, reference genes would be required for different omics levels and sample types. This indicates that reference gene normalization is not feasible for proteomics data. At the level of single tissue types (liver and jejunum), we found that proteins with low concentration variability across donors were involved in essential cellular processes. On the other hand, proteins with variable concentrations reflected varying degrees of disease, indicating that this type of variability analysis could be a simple aid in early biomarker discovery. Furthermore, we observed that proteins within single tissue types covered a wide variability spectrum, which implies that individual concentration levels should be taken into consideration for truly personalized outcomes of systems biology analysis.

## DATA AVAILABILITY

Concentration variability data (Gini indices), together with protein properties for the liver and jejunum datasets, are provided in Supplementary Data S1. Proteomics data of liver and jejunum are available upon email request to the corresponding author, provided that the proposed use of data is approved by the COCKTAIL-study steering committee and are according to the consent given by the participants and Norwegian laws and legislations (see Supplementary trial protocol).

## SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NARGAB Online.

## ACKNOWLEDGEMENTS

We thank Cecilia Karlsson and Shalini Andersson, Cardiovascular, Renal and Metabolism, Innovative Medicines and Early Development Biotech Unit, AstraZeneca, Jens Kristoffer Hertel, Morbid Obesity Centre, Department of Medicine, Vestfold Hospital Trust, and Dr André Mateus, Uppsala University, for valuable discussions. We also thank Lars Thomas Seeberg and colleagues at the Department of Gastrointestinal Surgery, Vestfold Hospital Trust for providing biopsies from jejunum and liver, and Ida Robertsen, Department of Pharmacy, University of Oslo, for sample logistics. Finally, we are grateful to the patients at the Morbid Obesity Center, Vestfold Hospital Trust, for participating in

the COCKTAIL study and accepting tissue biopsies to be taken.

**Author Contributions:** Conceptualization: C.W., M.Ö., T.B.A. and P.A.; Methodology: C.W. and M.Ö.; Resources: K.Z., J.R.W., J.H. and P.A.; Data curation: C.W.; Investigation: C.W. and M.Ö.; Formal analysis: C.W. and M.Ö.; Visualization: C.W. and M.Ö.; Writing—original draft: C.W., M.Ö. and P.A.; Writing—review and editing, C.W., M.Ö., J.R.W., P.L., A.Å., J.H., T.B.A. and P.A.; Funding acquisition: T.B.A. and P.A.

## FUNDING

Swedish Research Council [5715 to T.B.A., P.A., 01951 to P.A.].

*Conflict of interest statement.* None declared.

## REFERENCES

- Aebersold, R. and Mann, M. (2016) Mass-spectrometric exploration of proteome structure and function. *Nature*, **537**, 347–355.
- Wang, Z., Gerstein, M. and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.*, **10**, 57–63.
- Melé, M., Ferreira, P.G., Reverter, F., DeLuca, D.S., Monlong, J., Sammeth, M., Young, T.R., Goldmann, J.M., Pervouchine, D.D. and Sullivan, T.J. (2015) The human transcriptome across tissues and individuals. *Science*, **348**, 660–665.
- Uhlén, M., Fagerberg, L., Hallström, B.M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, Å., Kampf, C., Sjöstedt, E. and Asplund, A. (2015) Tissue-based map of the human proteome. *Science*, **347**, 1260419.
- Kim, M.-S., Pinto, S.M., Getnet, D., Nirujogi, R.S., Manda, S.S., Chaerkady, R., Madugundu, A.K., Kelkar, D.S., Isserlin, R. and Jain, S. (2014) A draft map of the human proteome. *Nature*, **509**, 575–581.
- Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A.M., Lieberenz, M., Savitski, M.M., Ziegler, E., Butzmann, L., Gessulat, S. and Marx, H. (2014) Mass-spectrometry-based draft of the human proteome. *Nature*, **509**, 582–587.
- Wang, D., Eraslan, B., Wieland, T., Hallstrom, B., Hopf, T., Zolg, D.P., Zecha, J., Asplund, A., Li, L.H., Meng, C. *et al.* (2019) A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.*, **15**, e8503.
- Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.*, **13**, 227–232.
- Edfors, F., Danielsson, F., Hallström, B.M., Käll, L., Lundberg, E., Pontén, F., Forsström, B. and Uhlén, M. (2016) Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.*, **12**, 883.
- Fortelny, N., Overall, C.M., Pavlidis, P. and Freue, G.V.C. (2017) Can we predict protein from mRNA levels? *Nature*, **547**, E19.
- Hughes, D.A., Kircher, M., He, Z., Guo, S., Fairbrother, G.L., Moreno, C.S., Khaitovich, P. and Stoneking, M. (2015) Evaluating intra- and inter-individual variation in the human placental transcriptome. *Genome Biol.*, **16**, 54.
- McCall, M.N., Illei, P.B. and Halushka, M.K. (2016) Complex sources of variation in tissue expression data: analysis of the GTEx lung transcriptome. *Am. J. Hum. Genet.*, **99**, 624–635.
- Wiśniewski, J.R., Duś-Szachniewicz, K., Ostasiewicz, P., Ziółkowski, P., Rakus, D. and Mann, M. (2015) Absolute proteome analysis of colorectal mucosa, adenoma, and cancer reveals drastic changes in fatty acid metabolism and plasma membrane transporters. *J. Proteome Res.*, **14**, 4005–4018.
- Hamdeh, S.A., Shevchenko, G., Mi, J., Musunuri, S., Bergquist, J. and Marklund, N. (2018) Proteomic differences between focal and diffuse traumatic brain injury in human brain tissue. *Sci. Rep.*, **8**, 6807.
- Chen, R. and Snyder, M. (2012) Systems biology: personalized medicine for the future? *Curr. Opin. Pharmacol.*, **12**, 623–628.

16. Tavassoly, I., Goldfarb, J. and Iyengar, R. (2018) Systems biology primer: the basic methods and approaches. *Essays Biochem.*, **62**, 487–500.
17. Hjelmseth, J., Åsberg, A., Andersson, S., Sandbu, R., Robertsen, I., Johnson, L.K., Angeles, P.C., Hertel, J.K., Skovlund, E., Heijer, M. *et al.* (2018) Impact of body weight, low energy diet and gastric bypass on drug bioavailability, cardiovascular risk factors and metabolic biomarkers: protocol for an open, non-randomised, three-armed single centre study (COCKTAIL). *BMJ Open*, **8**, e021878.
18. Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F. and Young, N. (2013) The genotype-tissue expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.
19. Wiśniewski, J.R. and Mann, M. (2012) Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phosphoproteomic analysis. *Anal. Chem.*, **84**, 2631–2637.
20. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367.
21. Huber, W., Von Heydebreck, A., Sultmann, H., Poustka, A. and Vingron, M. (2002) Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **18**, S96–S104.
22. Wiśniewski, J.R. and Rakus, D. (2014) Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome. *J. Proteomics*, **109**, 322–331.
23. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44.
24. Ceriani, L. and Verme, P. (2012) The origins of the Gini index: extracts from Variabilità e Mutabilità (1912) by Corrado Gini. *J. Econ. Inequal.*, **10**, 421–443.
25. O'Hagan, S., Muelas, M.W., Day, P.J., Lundberg, E. and Kell, D.B. (2018) GeneGini: Assessment via the gini coefficient of reference "Housekeeping" genes and diverse human transporter expression profiles. *Cell Syst.*, **6**, 230–244.
26. Cenik, C., Cenik, E.S., Byeon, G.W., Grubert, F., Candille, S.I., Spacek, D., Alsallakh, B., Tilgner, H., Araya, C.L. and Tang, H. (2015) Integrative analysis of RNA, translation and protein levels reveals distinct regulatory variation across humans. *Genome Res.*, **25**, 1610–1621.
27. Gorr, T.A. and Vogel, J. (2015) Western blotting revisited: critical perusal of underappreciated technical issues. *Proteomics*, **9**, 396–405.
28. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.*, **3**, doi:10.1186/gb-2002-3-7-research0034.
29. Wiśniewski, J.R. and Mann, M. (2016) A proteomics approach to the protein normalization problem: selection of unvarying proteins for MS-based proteomics and western blotting. *J. Proteome Res.*, **15**, 2321–2326.
30. Eisenberg, E. and Levanon, E.Y. (2013) Human housekeeping genes, revisited. *Trends Genet.*, **29**, 569–574.
31. Ölander, M., Wiśniewski, J.R., Mattsson, P., Lundquist, P. and Artursson, P. (2016) The proteome of filter-grown Caco-2 cells with a focus on proteins involved in drug disposition. *J. Pharm. Sci.*, **105**, 817–827.
32. Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A. and Narechania, A. (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res.*, **13**, 2129–2141.
33. Jadot, M., Boonen, M., Thirion, J., Wang, N., Xing, J., Zhao, C., Tannous, A., Qian, M., Zheng, H. and Everett, J.K. (2017) Accounting for protein subcellular localization: A compartmental map of the rat liver proteome. *Mol. Cell. Proteomics*, **16**, 194–212.
34. Settembre, C. and Ballabio, A. (2014) Lysosome: regulator of lipid degradation pathways. *Trends Cell Biol.*, **24**, 743–750.
35. Blomen, V.A., Májek, P., Jae, L.T., Bigenzahn, J.W., Nieuwenhuis, J., Staring, J., Sacco, R., van Diemen, F.R., Olk, N. and Stukalov, A. (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science*, **350**, 1092–1096.
36. Zecha, J., Meng, C., Zolg, D.P., Samaras, P., Wilhelm, M. and Kuster, B. (2018) Peptide level turnover measurements enable the study of proteoform dynamics. *Mol. Cell. Proteomics*, **17**, 974–992.
37. Kozłowski, L.P. (2016) Proteome-pI: proteome isoelectric point database. *Nucleic Acids Res.*, **45**, D1112–D1116.
38. Giurgiu, M., Reinhard, J., Brauner, B., Dunger-Kaltenbach, I., Fobo, G., Frishman, G., Montrone, C. and Ruepp, A. (2019) CORUM: the comprehensive resource of mammalian protein complexes—2019. *Nucleic Acids Res.*, **47**, D559–D563.
39. Mathieson, T., Franken, H., Kosinski, J., Kurzawa, N., Zinn, N., Sweetman, G., PoECKel, D., Ratnu, V.S., Schramm, M. and Becher, I. (2018) Systematic analysis of protein turnover in primary cells. *Nat. Commun.*, **9**, 689.
40. Gutiérrez-Sacristán, A., Bravo, A., Centeno, E., Sanz, F., Piñero, J., García-García, J., Deu-Pons, J., Queralt-Rosinach, N. and Furlong, L.I. (2016) DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.*, **45**, D833–D839.
41. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S. and Eppig, J.T. (2000) Gene Ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–29.
42. Gregor, M.F. and Hotamisligil, G.S. (2011) Inflammatory mechanisms in obesity. *Annu. Rev. Immunol.*, **29**, 415–445.
43. Geyer, P.E., Kulak, N.A., Pichler, G., Holdt, L.M., Teupser, D. and Mann, M. (2016) Plasma proteome profiling to assess human health and disease. *Cell Syst.*, **2**, 185–195.
44. Zhao, M., Li, M., Yang, Y., Guo, Z., Sun, Y., Shao, C., Li, M., Sun, W. and Gao, Y. (2017) A comprehensive analysis and annotation of human normal urinary proteome. *Sci. Rep.*, **7**, 3024.
45. Plevy, S., Silverberg, M.S., Lockton, S., Stockfish, T., Croner, L., Stachelski, J., Brown, M., Triggs, C., Chuang, E., Princen, F. *et al.* (2013) Combined serological, genetic, and inflammatory markers differentiate non-IBD, Crohn's disease, and ulcerative colitis patients. *Inflamm. Bowel Dis.*, **19**, 1139–1148.
46. Danhof, M. (2016) Systems pharmacology – Towards the modeling of network interactions. *Eur. J. Pharm. Sci.*, **94**, 4–14.
47. Lin, L., Yee, S.W., Kim, R.B. and Giacomini, K.M. (2015) SLC transporters as therapeutic targets: emerging opportunities. *Nat. Rev. Drug Discovery*, **14**, 543.
48. Vasilio, V., Vasilio, K. and Nebert, D.W. (2009) Human ATP-binding cassette (ABC) transporter family. *Hum. Genomics*, **3**, 281.
49. Williams, J.A., Hyland, R., Jones, B.C., Smith, D.A., Hurst, S., Goosen, T.C., Peterkin, V., Koup, J.R. and Ball, S.E. (2004) Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUC<sub>i</sub>/AUC) ratios. *Drug Metab. Dispos.*, **32**, 1201–1208.
50. Zanger, U.M. and Schwab, M. (2013) Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacol. Therap.*, **138**, 103–141.
51. Wessely, F., Bartl, M., Guthke, R., Li, P., Schuster, S. and Kaleta, C. (2011) Optimal regulatory strategies for metabolic pathways in *Escherichia coli* depending on protein costs. *Mol. Syst. Biol.*, **7**, 515.
52. Buttgerit, F. and Brand, M.D. (1995) A hierarchy of ATP-consuming processes in mammalian cells. *Biochem. J.*, **312**, 163–167.
53. Khan, Z., Ford, M.J., Cusanovich, D.A., Mitrano, A., Pritchard, J.K. and Gilad, Y. (2013) Primate transcript and protein expression levels evolve under compensatory selection pressures. *Science*, **342**, 1100–1104.
54. Liu, Y., Beyer, A. and Aebersold, R. (2016) On the dependency of cellular protein levels on mRNA abundance. *Cell*, **165**, 535–550.
55. Gabrielsson, B.G., Olofsson, L.E., Sjögren, A., Jernäs, M., Elander, A., Lönn, M., Rudemo, M. and Carlsson, L.M.S. (2005) Evaluation of reference genes for studies of gene expression in human adipose tissue. *Obes. Res.*, **13**, 649–652.
56. Välikangas, T., Suomi, T. and Elo, L.L. (2016) A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.*, **19**, 1–11.
57. Kmiec, Z. (2001) Cooperation of liver cells in health and disease. *Adv. Anat. Embryol. Cell Biol.*, **161**, 1–151.
58. Tortora, G.J. and Derrickson, B.H. (2008) *Principles of Anatomy and Physiology*. John Wiley & Sons, Hoboken.
59. Ding, C., Li, Y., Guo, F., Jiang, Y., Ying, W., Li, D., Yang, D., Xia, X., Liu, W., Zhao, Y. *et al.* (2016) A Cell-type-resolved liver proteome. *Mol. Cell. Proteomics*, **15**, 3190.
60. Lown, K.S., Kolars, J.C., Thummel, K.E., Barnett, J.L., Kunze, K.L., Wrighton, S.A. and Watkins, P.B. (1994) Interpatient heterogeneity in

- expression of CYP3A4 and CYP3A5 in small bowel. Lack of prediction by the erythromycin breath test. *Drug Metab. Dispos.*, **22**, 947–955.
61. Graham, M.F., Diegelmann, R.F., Elson, C.O., Lindblad, W.J., Gotschalk, N., Gay, S. and Gay, R. (1988) Collagen content and types in the intestinal strictures of Crohn's disease. *Gastroenterology*, **94**, 257–265.
  62. Hounnou, G., Destrieux, C., Desme, J., Bertrand, P. and Velut, S. (2002) Anatomical study of the length of the human intestine. *Surg. Radiol. Anat.*, **24**, 290–294.
  63. Drozdziak, M., Busch, D., Lapczuk, J., Müller, J., Ostrowski, M., Kurzawski, M. and Oswald, S. (2018) Protein abundance of clinically relevant Drug-Metabolizing enzymes in the human liver and intestine: a comparative analysis in paired tissue specimens. *Clin Pharmacol Ther.*, **104**, 515–524.
  64. Wiśniewski, J.R., Wegler, C. and Artursson, P. (2019) Multiple-Enzyme-Digestion strategy improves accuracy and sensitivity of Label- and Standard-Free absolute quantification to a level that is achievable by analysis with stable Isotope-Labeled standard spiking. *J. Proteome Res.*, **18**, 217–224.
  65. Moore, S.M., Hess, S.M. and Jorgenson, J.W. (2016) Extraction, enrichment, solubilization, and digestion techniques for membrane proteomics. *J. Proteome Res.*, **15**, 1243–1252.
  66. Moor, A.E., Harnik, Y., Ben-Moshe, S., Massasa, E.E., Rozenberg, M., Eilam, R., Bahar Halpern, K. and Itzkovitz, S. (2018) Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell*, **175**, 1156–1167.
  67. Alemu, E.Y., Carl, J.W. Jr, Corrada Bravo, H. and Hannenhalli, S. (2014) Determinants of expression variability. *Nucleic Acids Res.*, **42**, 3503–3514.