

SysQuan: Repurposing SILAC Mice for the Cost-Effective Absolute Quantitation of the Human Proteome

Authors

Yassene Mohammed, Vincent R. Richard, M. Immanuel Reyes Madlangsakay, Ying Lao, Victor Spicer, Robert Popp, Claudia Gaither, Laura Hennecken, Wolfgang Kleinekoft, René P. Zahedi, and Christoph H. Borchers

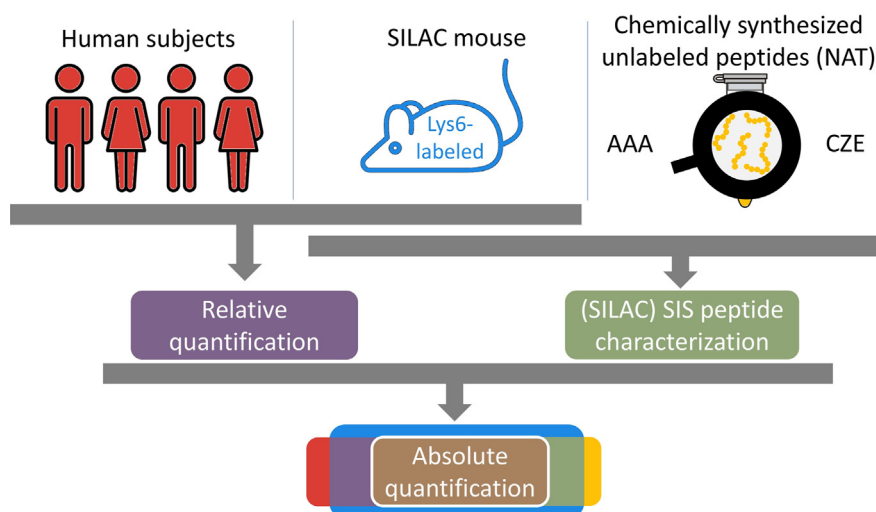
Correspondence

Rene.Zahedi@umanitoba.ca;
Christoph.borchers@mcgill.ca

Graphical Abstract

In Brief

SysQuan is a novel approach that enables cost-effective absolute quantitation in MS-based proteomics by repurposing SILAC mouse tissues as system-wide internal standards for human samples. This method allows quantification of up to two-thirds of the human proteome, enhancing data comparability across labs, studies, and time points. SysQuan overcomes cost barriers associated with traditional SIL standards, making large-scale absolute quantitation feasible and improving data reusability in population-wide and longitudinal studies.



Highlights

- SysQuan enables large-scale absolute quantitation using SILAC mouse tissues.
- SysQuan quantifies 70% of liver proteome and 31% of plasma proteome in human samples.
- We demonstrate that absolute quantification is possible using 14 metabolic proteins.
- Cost-effective method enables quantifying 1000s of targets in 1000s of samples.
- SysQuan enhances cross-study comparability and data reusability at scale.



SysQuan: Repurposing SILAC Mice for the Cost-Effective Absolute Quantitation of the Human Proteome

Yassene Mohammed^{1,2,3}, Vincent R. Richard¹, M. Immanuel Reyes Madlangsakay⁴, Ying Lao⁴ , Victor Spicer⁴, Robert Popp⁵, Claudia Gaither^{5,6}, Laura Hennecken⁷, Wolfgang Kleinekoft⁷ , René P. Zahedi^{4,8,9,10,*} , and Christoph H. Borchers^{1,2,11,12,*}

Relative quantitation, used by most mass spectrometry-based proteomics laboratories to determine protein fold-changes, requires samples being processed and analyzed together for best comparability through minimizing batch differences. This limits the adoption of mass spectrometry-based proteomics in population-wide studies and the detection of subtle but relevant changes in heterogeneous samples. Absolute quantitation circumvents these limitations and enables comparison of results across laboratories, studies, and over time. However, high cost of the essential stable isotope labeled (SIL) standards prevents widespread access and limits the number of quantifiable proteins. Our new approach, called “SysQuan”, repurposes SILAC mouse tissues/biofluids as system-wide internal standards for matched human samples to enable absolute quantitation of, theoretically, two-thirds of the human proteome using 157,086 shared tryptic peptides, of which 73,901 with lysine on the c terminus. We demonstrate that SysQuan enables quantification of 70% and 31% of the liver and plasma proteomes, respectively. We demonstrate for 14 metabolic proteins that abundant SIL mouse tissues enable cost-effective reverse absolute quantitation in, theoretically, 1000s of human samples. Moreover, 10,000s of light/heavy doublets in untargeted SysQuan datasets enable unique postacquisition absolute quantitation. SysQuan empowers researchers to replace relative quantitation with affordable absolute quantitation at scale, making data comparable across laboratories, diseases and tissues, enabling completely novel study designs and increasing reusability of data in repositories.

Proteins play a crucial role in health and disease and comprise the majority of disease biomarkers and drug targets. Mass spectrometry (MS) enables proteome-wide identification and (relative) quantitation of proteins with high sensitivity, precision, and unmatched specificity in virtually any type of sample, from mummies to human tissue biopsies. Consequently, MS-based proteomics has become the most important tool for biological discovery, with applications ranging from systems biology to biomarker candidate discovery.

Although, in recent years throughput and depth of MS-based proteomics approaches have considerably increased while concurrently cost per sample has been reduced (1), (pre) clinical studies continue to rely mainly on genomics approaches to study larger patient sample cohorts. One major reason why the impact of MS-based proteomics on large-scale studies is clearly lagging behind its unmatched potential to reflect actual disease phenotypes is that most laboratories exclusively employ relative quantitation to determine fold-changes in protein abundance between samples. This is because of the ease of use of relative quantification methods, but also due to it being cost-effective. However, relative quantitation works best for a limited number of samples, which ideally are processed and analyzed together on the same LC-MS system. Once the number of samples increases and sample collection, preparation, measurement batches are introduced relative quantification suffers from considerable batch-effects that contradict the goals of low-abundance biomarker detection and target identification in heterogeneous patient populations.

From the ¹Segal Cancer Proteomics Centre, Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, Quebec, Canada; ²Gerald Bronfman Department of Oncology, Jewish General Hospital, Montreal, Quebec, Canada; ³Center for Proteomics and Metabolomics, Leiden University Medical Center, Leiden, The Netherlands; ⁴Manitoba Centre for Proteomics and Systems Biology, University of Manitoba, Winnipeg, Manitoba, Canada; ⁵MRM Proteomics, Inc, Montreal, Quebec, Canada; ⁶Faculty of Veterinary Medicine, Department of Clinical Sciences, University of Montreal, Saint-Hyacinthe, Quebec, Canada; ⁷Department of Engineering, Hochschule RheinMain, Rüsselsheim am Main, Germany; ⁸Department of Internal Medicine, University of Manitoba, Winnipeg, Manitoba, Canada; ⁹Department of Biochemistry and Medical Genetics, University of Manitoba, Winnipeg, Manitoba, Canada; ¹⁰Paul Abrechtsen Research Institute, CancerCare Manitoba, Winnipeg, Manitoba, Canada; ¹¹Division of Experimental Medicine, McGill University, Montreal, Quebec, Canada; ¹²Department of Pathology, McGill University, Montreal, Quebec, Canada

* For correspondence: Christoph H. Borchers, Christoph.borchers@mcgill.ca; René P. Zahedi, Rene.Zahedi@umanitoba.ca.

Large-scale studies with 100s to 1000s of samples require quantitative reproducibility that can eventually cover data generation across different laboratories, instruments, as well as longitudinal measurements. To date, these requirements can only be met by absolute quantitation where protein concentrations are determined in individual samples using stable isotope labeled (SIL) internal standards (2). Spiking defined amounts of SIL peptides into samples enables quality control of the entire analytical workflow, including monitoring changes in instrument performance, and thus determining protein concentrations with maximum specificity and precision that can be compared across laboratories and platforms, between studies, and over time (3).

Targeted MS using multiple reaction monitoring (MRM)—also known as selected reaction monitoring—on triple-quadrupole mass spectrometers is the “gold standard” for absolute quantitation, particularly in clinical settings (4, 5). However, having at least one SIL peptide for each target protein that is individually synthesized, purified, and characterized (for purity and quantity), drastically increases cost. The high cost of SIL peptides confine this powerful technology to a few applications and a few expert labs that mostly measure small protein panels, even though the feasibility of precise quantitation of 100s of proteins using highly multiplexed MRM has been demonstrated (6, 7). Consequently, MS-based proteomics is clearly dominated by more cost-efficient relative quantitation, which, apart from the abovementioned limitations, also lacks the precision to identify small but biologically relevant changes in protein abundance.

We present SysQuan—a novel strategy for cost-effective system-wide absolute quantitation of human proteins that we anticipate to reshape quantitative proteomics field. SysQuan enables determining precise concentrations for 1000s of proteins by exploiting the genetic proximity between humans and mice, with 99% of all human genes having homologs in the mouse genome. Homologous proteins in mice and humans may not have identical amino acid sequences, but they often share identical sequence stretches that are essential to protein function. We hypothesized that this proteomic proximity enables replacing costly SIL peptides with system-wide internal standards derived from SIL tissues that can be extracted in large quantities from SILAC (stable isotope labeling by amino acids in cell culture) mice (8). Matched “light” human and “heavy” SIL mouse tissues can be pooled, and after codigestion peptides that are shared between humans and mice will generate light/heavy peptide pairs with defined mass shifts (doublets), just like with spiked-in SIL peptides or in conventional SILAC experiments.

We believe that the approach we are presenting here with SysQuan as a proteome-wide internal standard paves the way to MS-based large-scale proteomics studies with their unmatched precision and specificity to complement genomics and address the genotype phenotype disconnect (9–11).

EXPERIMENTAL PROCEDURES

Experimental Design and Statistical Rationale

We chose one human liver and one human plasma sample to experimentally confirm the in-silico predicted utility of SysQuan. The goal was to identify SILAC doublets representing human proteins that can be quantified by using corresponding SILAC mouse standards. To demonstrate the reproducibility of the approach, we performed the entire workflow for liver in two different laboratories, using different sample preparation protocols and different LC-MS instrumentation for the analysis. The obtained human:mouse (light:heavy) SILAC ratios of 10,000s of human peptides showed a good agreement between sites. This is in line with studies comparing targeted MS workflows across laboratories, but on a much wider scale.

To demonstrate that SysQuan enables absolute quantitation of human proteins, we first determined the concentrations of 14 proteins in the SILAC mouse liver using classical targeted MRM with characterized light synthetic peptide standards, and then used this characterized SIL mouse liver standard to quantify the same proteins in the light human tissue.

Human and SILAC Mouse Materials

Human plasma was purchased from BioIVT (Westbury) and obtained in K₂-EDTA vials and stored at –80 °C until use (Western Institutional Review Board–WIRB number 20161665). Plasma samples were from healthy donors (aged 18–50 years), who provided informed consent, in accordance with the Declaration of Helsinki. Fresh-frozen normal-adjacent liver tissue from an adult human donor with hepatocellular carcinoma (male, 53 years old) was also obtained from BioIVT. Informed consent for the use of the samples in research was provided by the patients. Isotopically labeled mouse plasma and fresh-frozen liver samples were acquired from Silantes GmbH, product numbers 252923900 and 252923905. The labeling was performed in C57BL6 mice fed ¹³C-lysine (K + 6) containing chow over six generations and was confirmed by the vendor to have 97% incorporation of K + 6 (Silantes GmbH). Silantes has stated that the mouse tissues “comply with German and European ethical guidelines for animal research, including EU Directive 2010/63/EU and the German Animal Welfare Act (Tierschutzgesetz). The tissues also adhere to the regulations set by the relevant German authorities such as the Regierungspräsidium and local animal welfare committees.”

Plasma Sample Preparation

Human and mouse plasma samples were depleted of the top 14 most abundant plasma proteins using High Select Top14 abundant protein depletion columns (Thermo Fisher Scientific), in order to reduce the dynamic range of the plasma proteome, and maximize the depth of proteome coverage. Depleted mouse and human plasma samples were pooled 1:1 and subsequently diluted in protein denaturation buffer (5% SDS, 100 mM Tris pH 8.5, and 10 mM tris(2-carboxyethyl)phosphine [TCEP]) and heated to 60 °C for 30 min. Proteins were reduced in 10 mM TCEP, alkylated with 20 mM iodoacetamide (IAA), and digested with trypsin (Sigma-Aldrich) at a 1:10 enzyme:substrate ratio (approximately 5 µg of trypsin to 50 µg of depleted plasma protein) at 37 °C for 16 h using S-TRAP micro cartridges (ProtiFi LLC). Proteolysis was stopped with formic acid (FA; 0.5% v/v), and the resulting tryptic peptides were lyophilized to dryness prior to offline fractionation (detailed below).

Liver Tissue Homogenization

Human (JGH) and mouse (Silantes) fresh-frozen liver samples were cryohomogenized using a prechilled (liquid nitrogen) mortar-and-pestle. Homogenates were transferred to protein

low-binding Eppendorf tubes containing protein extraction buffer (5% SDS, 100 mM Tris pH 8.5, and 10 mM TCEP) and subjected to probe-based ultrasonication (Thermo Sonic Dismembrator) and heating to 95 °C for 10 min. Liver lysates were clarified by centrifugation (21000g, 5 min) and approximately 5% of the sample was reserved for protein concentration determination by bicinchoninic acid (BCA) assay—(NAT) (Thermo/Pierce).

Liver SysQuan/University of Manitoba (U of M)

Light human and heavy mouse liver lysate were mixed 1:1 to generate a 100- μ g human/mouse protein mix. Disulfide bonds were reduced in 10 mM DTT for 30 min at 60 °C followed by alkylation in 30 mM IAA for 45 min at room temperature in the dark. Proteolytic digestion was performed using single-pot solid-phase-enhanced sample preparation (SP3) (12). Prior to SP3, two types of carboxylate-modified Sera-Mag Speed beads (GE Life Sciences) were combined 1:1 (v/v), rinsed, and reconstituted in water at a concentration of 20 μ g solids/ μ l. Ten microliters of the prepared bead mix were added to the lysate, and samples were adjusted to pH 7.0 using Hepes buffer. To promote proteins binding to the beads, acetonitrile (ACN) was added to a final concentration of 70% (v/v) and samples were incubated at room temperature on a tube rotator for 18 min. Subsequently, beads were immobilized on a magnetic rack for 1 min. The supernatant was discarded, and the pellet was rinsed twice with 200 μ l of 70% ethanol and once with 200 μ l of 100% ACN while on the magnetic rack. The rinsed beads were resuspended in 115 μ l of 50 mM Hepes buffer (pH 8.0) supplemented with trypsin/LysC (Promega) at an enzyme-to-protein ratio of 1:25 (w/w) and incubated for 16 h at 37 °C. Peptide concentration was determined using Pierce Quantitative Fluorometric Peptide Assay (Thermo Fisher Scientific).

Offline High pH Reversed-phase Fractionation—An Agilent 1100 series LC system with UV detector (214 nm) and a 1 mm \times 100 mm Gemini C18, 5 μ m column (Phenomenex) were used for reversed-phase fractionation at pH 10. Both eluents (a) water and b (1:9 water: ACN) contained 20 mM ammonium formate at pH 10. Both eluents — (a) water and b (1:9 water: ACN) — contained 20 mM ammonium formate pH 10. A linear gradient from 0.1 to 34% eluent B in 100 min was delivered at flow rate of 150 μ l/min. A 65- μ g aliquot of the liver digest was injected and 80 1-min fractions were collected (between 10 and 90 min) and combined into 40 fractions to provide optimal separation orthogonality. These fractions were lyophilized and resuspended in 0.1% FA.

Liquid Chromatography-Tandem Mass Spectrometry—Per fraction, 1 μ g of total peptide was analyzed on an Orbitrap Exploris 480 (Thermo Fisher Scientific) coupled to an Easy-nLC 1000 (Thermo Fisher Scientific) equipped with a C18 (Luna C18(2), 3 μ m particle size (Phenomenex)) column packed in-house in PicoFrit(100 μ m \times 30 cm) capillaries (New Objective). Peptides were separated using a binary gradient with (a) 0.1% FA and (b) 0.1% (v/v) FA in 80% ACN (LC-MS grade), ramping from 0 to 5% B over 3 min, 5 to 7% B over 2 min, 7 to 25% B over 84 min, 25 to 60% B over 15 min, 60 to 95% B over 1 min at a flow rate of 300 nL/min. The Orbitrap Exploris 480 instrument was operated in data-dependent acquisition (DDA) mode. Spray voltage was set to 2.4 kV, funnel RF level at 40, and the heated capillary at 275 °C. Survey scans covering the m/z 380 to 1500 were acquired at a resolution of 90,000 (at m/z 200), with a maximum ion injection time of 50 ms, and a normalized automatic gain control target of 300%. This was followed by MS2 acquisition at a resolution of 30,000, selected ions were fragmented at 30% normalized collision energy, with intensity threshold kept at 2e4. Automatic gain control target value for fragment spectra was set to 100%, with a maximum ion injection time

set to auto and an isolation width set at 1.6 m/z. Dynamic exclusion of previously selected ions was enabled for 30 s, charge state filtering was limited to 2 to 6, peptide match was set to preferred, and isotope exclusion was on.

Liver SysQuan/McGill

Fresh-frozen human and mouse liver samples were cryo-homogenized using a prechilled mortar and hammer pestle. Homogenates were transferred to protein low binding Eppendorf tubes containing protein extraction buffer (5% SDS, 100 mM Tris pH 8.5, and 10 mM TCEP) and subjected to probe-based ultrasonication (Thermo Sonic Dismembrator) and heated to 95 °C for 10 min. Liver lysates were then clarified by centrifugation (21Kg, 5 min) and approximately 5% of the sample was reserved for protein concentration determination by BCA assay. Protein extracts were reduced and alkylated with 10 mM TCEP and 40 mM IAA, respectively and an equivalent of 200 μ g of liver lysate was digested with trypsin (Sigma-Aldrich) at a 1:10 enzyme to substrate ratio at 37 °C for 16 h using S-TRAP micro cartridges (ProtiFi LLC). Proteolysis was stopped with FA (0.5% v/v) and resulting tryptic peptides were lyophilized to dryness prior to offline fractionation. Subsequently, 50 μ g of resultant protein digests were then pooled 1:1 (100 μ g total) and subjected to offline high pH reversed phase separation.

Offline High pH Reversed-Phase Fractionation—To further increase the depth of coverage, we performed offline high-pH reversed-phase fractionation using an Agilent 1290 fraction collection system. Peptides were separated using a Waters XBridge peptide C18 column (2.1 \times 150 mm and 2.1 μ m particle), and peptides were separated using a binary gradient of (a) 10 mM ammonium formate (pH 10) and (b) acetonitrile with 10 mM ammonium formate from 0% to 80% B over 48 min at flow rate of 400 μ l/min with fractions collected every 30 s. Wells corresponding to fractions separated by 24 min or 48 fractions (i.e., well positions 1 + 49, 2 + 50, 3 + 51, and so on from left to right) were pooled and vacuum concentrated for a total of 48 fractions.

LC-MS/MS—Fractions from offline high-pH separation were reconstituted in 0.1% FA and analyzed by DDA using an Evosep One LC system (30SPD, EV1137 column at 45 °C) coupled to a Bruker timsTOF HT mass spectrometer. Briefly, samples were acquired in DDA mode using 10 parallel accumulation serial fragmentation (PASEF) ramps scanning between 100 to 1700 m/z (method cycle time was approximately 1.1 s) (13). The capillary voltage was set to 1.6 kV, dry gas was 3.0 L/min, and temperature was set to 180 °C. +2 and +3 charged peptides were selected based on their ion mobility values (0.6–1.6 1/k0). Precursors selected for ms/ms were dynamically excluded for 0.4 min.

Data Processing and Analysis for Untargeted Acquisition

For each individual sample set (plasma, liver McGill, liver University of Manitoba (UofM)) all raw data were searched using MSFragger embedded within the FragPipe (v21.1) interface (<https://fragpipe.nesvilab.org/>). In order to exclusively detect the shared peptides, the data were searched against the canonical human proteome downloaded from UniProtKB (UP000005640, downloaded on May 2024 containing 20,467 protein sequences). The FASTA file was supplemented with reverse decoy sequences and common contaminants using FragPipe. Searches were performed using default settings with labeled Lys (K + 6.020129 Da) set as a variable modification and including precursor quantitation based on SILAC with the following label types: light (K + 0) and heavy (K + 6.020129 Da). Other modifications included variable oxidation on methionine (+15.9949 Da), variable protein N-terminal acetylation (+42.0106 Da), and fixed

carbamidomethylation on cystine (+57.02146 Da). Precursor mass lower and upper tolerances were set to -20 and 20 ppm, respectively. Fragment mass tolerance was set to ± 20 ppm. Peptide and protein false discovery rates were set to $\leq 1\%$. Further processing and figure generation were performed in R using standard analysis and plotting functions.

Absolute Quantitation of Metabolomic Pathway Proteins

S-TRAP-digests of light human and SIL mouse liver were mixed 1:1 (w/w) based on BCA protein concentrations. A calibration curve was generated by injecting 0.1, 1, 10, 100, and 1000 fmol, of unlabeled (NAT) peptides on column for each of the 14 target proteins in constant SIL mouse liver background. The 14 proteins corresponded to the following genes: ACO1, ASS1, COX7C, DLAT, GAPDH, GOT1, HADH, NDUFA4, NDUFS4, NDUFV1, NDUFV2, SLC25A3, SUCLA2, and SUCLG1 (see also Table 1 and Supplemental Table S1). Next, human/mouse liver mix and the calibration curve were measured by LC-MRM on an Agilent 6495C mass spectrometer on-line coupled to an Agilent 1290 HPLC. Peptides were separated on an Agilent Zorbax Eclipse Plus RP-UHPLC column (2.1×150 mm, $1.8 \mu\text{m}$) using a 47-min binary gradient (A: 0.1% FA; B: 100% ACN) ramping from 2 to 7% B in 2 min, 7%–27.6% B in 43 min, and 27.6 to 80% B in 2 min. A total of 930 MRM transitions were acquired for the 14 light and 14 SIL peptides using dynamic MRM and 2-min retention time windows, with minimum and maximum dwell times of 1.62 ms and 71.85 ms, respectively; the maximum cycle time was 555.59 ms (see Supplemental Data Table S2). Data were analyzed using Skyline. Calibration curves were used to calculate target analyte concentrations (fmol analyte/ μg of liver tissue) in the human liver sample for

a given light:heavy peak area ratio (see Supplemental Data Fig. S1). The absolute quantitation performed corresponds to CPTAC Tier 1 (5).

RESULTS

Over the past 10 years, transformative advancements in MS-based proteomics have led to a significant boost in acquisition speed at improved proteome coverage, accompanied by a reduction in both required sample input and cost per sample. While, in the past, the considerable increase in sample complexity arising from mixing mouse and human proteomes had been a major challenge, studies involving “mixed proteomes” such as PDX models (14–16) and metaproteomics (17) studies with dozens of microorganisms have become feasible. It can be expected that the remarkable progress in proteomics technologies is going to continue, increasingly driven by single-cell proteomics with its huge demands for higher throughput, greater sensitivity (18, 19), and lower cost per sample. To assess the reproducibility and platform-independence of SysQuan, we performed in-depth analysis of human/mouse liver in two different laboratories using different sample preparation workflows and different LC-MS instruments (McGill and the University of Manitoba (UofM)) (Fig. 1). As recent reports indicate that a novel hybrid mass spectrometer enables the analysis of 5000 to 6000 proteins in below 5 min (20), we believe that a similar coverage

TABLE 1
Absolute quantitation of metabolomic pathway protein components

UniProtKB accession	Gene name	Protein name	Peptide	SILAC protein amount (fmol/ μg tissue)	Endogenous concentration in human tissue (fmol/ μg tissue)
Q9P2R7	SUCLA2	Succinate–CoA ligase [ADP-forming] subunit beta	ALIADSGLK	3.78	0.74
O00483	NDUFA4	Cytochrome c oxidase subunit NDUFA4	FYSVNVDSYK	24.70	8.09
P15954	COX7C	Cytochrome c oxidase subunit 7C	NLPFSVENK	5.16	1.46
P00966	ASS1	Argininosuccinate synthase	QHGIPIVTPK	65.44	21.49
Q43181	NDUFS4	NADH dehydrogenase [ubiquinone] iron-sulfur protein 4	LDITTLTGVP EEHIK	2.44	1.02
P19404	NDUFV2	NADH dehydrogenase [ubiquinone] flavoprotein 2	DIEEIIDELK	2.57	1.17
P49821	NDUFV1	NADH dehydrogenase [ubiquinone] flavoprotein 1	EAYEAGLIGK	2.55	1.56
P10515	DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex	ISVNDFIHK	3.44	1.94
P53597	SUCLG1	Succinate–CoA ligase [ADP/GDP-forming] subunit alpha	LVGGTTPGK	6.79	4.23
P04406	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	QASEGPLK	72.86	97.03
Q00325	SLC25A3	Solute carrier family 25 member 3 (Phosphate carrier protein)	FGFYEVFK	9.91	11.48
P21399	ACO1	Cytoplasmic aconitate hydratase	YQQAGLPLIVLAGK	3.24	3.50
Q16836	HADH	Hydroxyacyl-coenzyme A dehydrogenase	AGDEFVEK	10.73	16.41
P17174	GOT1	Aspartate aminotransferase	NLDYVATSIHEAVTK	18.75	46.55

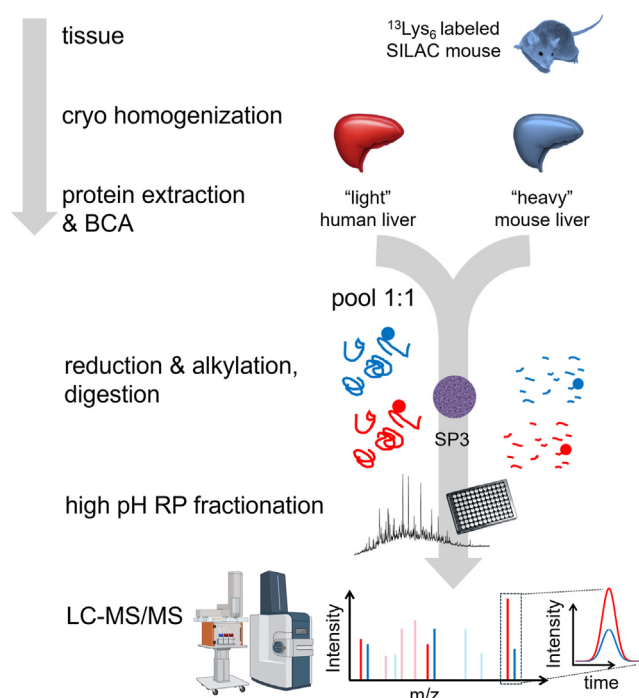


FIG. 1. SysQuan feasibility workflow exemplified for liver tissue. Heavy mouse (blue) and light human (red) liver tissue were homogenized, proteins extracted, and pooled 1:1 (wt/wt). Proteolytic digestion with trypsin was performed using SP3, followed by offline reversed phase fractionation at pH 10.0. Fractions were analyzed by LC-MS/MS and light/heavy doublets representing SysQuan-quantifiable peptides were identified. LC-MS/MS, liquid chromatography-tandem mass spectrometry.

as achieved here with fractionation will be feasible 50-100x faster in single-shot experiments.

Human-Mouse Proteomic Homology

Mus musculus has high similarity with humans at the molecular level with 99% of human genes having homologs in the mouse genome. Furthermore, many proteins in human and mouse share large part of their sequences. We calculated that approximately 68% of all human proteins have shared tryptic peptides in mouse proteome, and 60% with c-terminal lysine that can be labeled in SILAC mouse (Fig. 2). Due to the arginine catabolism pathway in which arginine is converted to glutamate and proline, labeling with arginine is not desirable as it can introduce quantification errors in SILAC experiments. Despite available solutions to this problem (21–23), these are impractical, especially on a whole mouse. Thus only heavy lysine is typically used for SILAC labeling in mouse. Approximately 157k out of the 455k human tryptic peptides are also present in mouse, considering 7 to 25 amino acid length. Of these 73k have lysine on the c terminus. SysQuan exploits this overlap by using SILAC mouse bio-fluids and tissues to quantify human biofluids and tissues.

The SysQuan-Quantifiable Proteome: Liver

Mouse and human liver tissues were cryogenically homogenized, proteins extracted, and heavy mouse and light human protein extracts were mixed 1:1 (wt/wt), followed by reduction and alkylation, and proteolytic digestion with trypsin. The mouse/human peptide mix was then fractionated using high pH reversed phase chromatography. This workflow was performed independently with slight deviations at two different laboratories where individual fractions were analyzed by either DDA on an Orbitrap Exploris 480 (40 fractions) or PASEF-DDA on a timsTOF HT (48 fractions). Both datasets were processed with identical workflows using the Fragpipe software suite and SILAC settings.

In total, we identified 59,734 and 115,762 peptide-spectrum matches in the Exploris and timsTOF datasets, respectively, corresponding to 50,494 and 76,049 unique human peptide sequences, and 8323 and 8035 proteins. In the Exploris data, approximately 50% (23,727) of the unique peptides had light/heavy pairs corresponding to 5990 human proteins, and 19,004 nonmodified mouse/human peptide pairs corresponded to 5560 unique proteins.

In the timsTOF data, 2/3 (48,760) of the unique peptides had light/heavy pairs, covering 6987 unique human proteins, and 35,226 nonmodified mouse/human peptide pairs corresponded to 6305 unique proteins.

Focusing on one of the main liver function, metabolism, out of the 1562 proteins involved in the metabolic pathways (24), 1187 were identified in the liver proteome that we acquired, of which 1002 were quantifiable with heavy/light pairs.

The SysQuan-Quantifiable Proteome: Plasma

Mouse and human plasma were depleted of the top 14 most abundant proteins, pooled 1:1, and processed as described for the liver/McGill dataset. A total of 48 high pH reversed phase fractions were analyzed by PASEF-DDA on a timsTOF HT. Fragpipe with SILAC settings yielded 23,501 peptide-spectrum matches corresponding to 15,302 unique peptides and 2279 unique proteins. More than 1/3 (5769) peptides corresponding to 1431 proteins had light/heavy pairs and 4020 nonmodified mouse/human peptide pairs corresponded to 1209 proteins.

Comparison to Reference Proteomes

To contextualize the depth of our SysQuan-quantifiable proteomes, we compared our data with the current reference proteomes from the proteomics community. The latest build of the Human Plasma Proteome Project (HPPP 2023–04) of the Human Proteome Organization (14), combines results from 313 experiments and comprises 4608 human plasma proteins. In our mouse/human plasma mix we identified 2279 human proteins, 1431 of which can be quantified using SysQuan. Thus, with a single workflow and proteolytic enzyme, SysQuan enables quantitation of approximately 31% of the

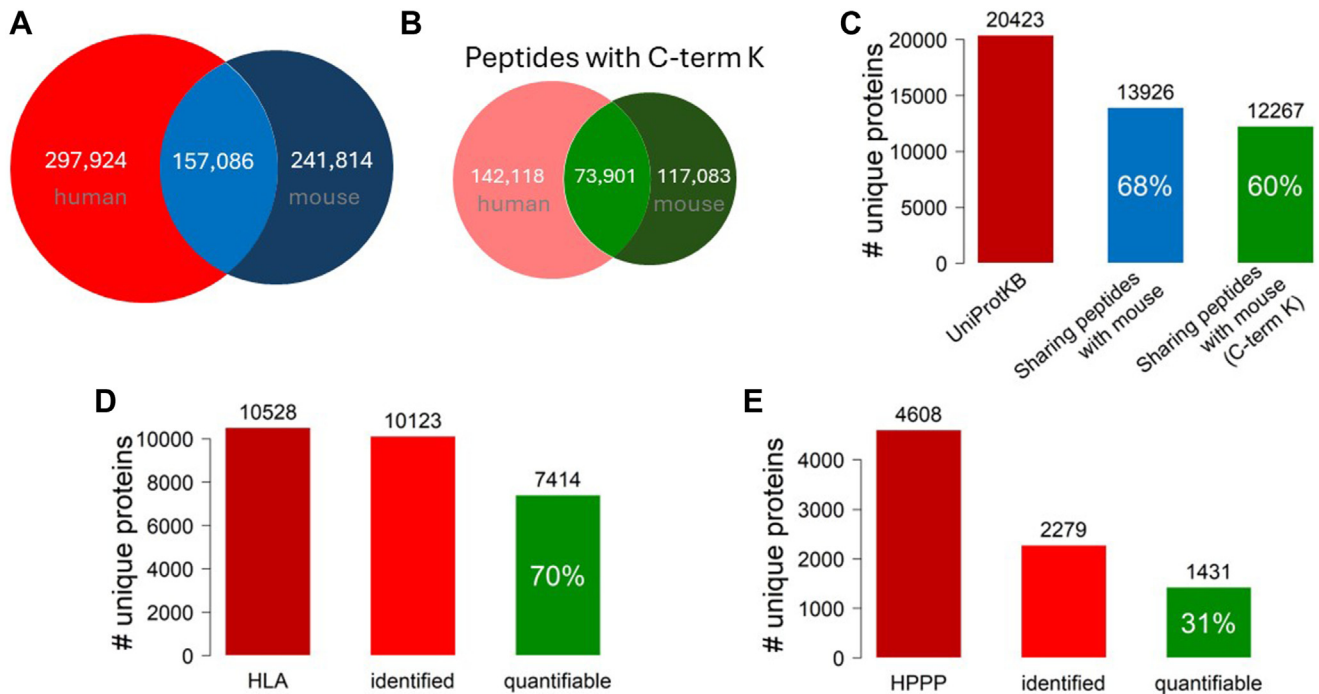


FIG. 2. **SysQuan human proteome coverage.** A, overlap of *in-silico*-generated tryptic human and mouse peptides. B, overlap of *in-silico*-generated tryptic human and mouse peptides with C-terminal lysine. C, based in this overlap, 68% of all human proteins in UniProtKB have shared peptides with mouse proteins, and 60% share peptides with C-terminal lysine and are therefore theoretically quantifiable using SysQuan. D, coverage and quantifiable share of the Human Liver Atlas (HLA), and (E) the Human Plasma Proteome Project (HPP) reference proteomes.

known plasma proteome (Fig. 2). Diversifying the proteomics workflow to better reflect the 313 experiments in HPP may allow quantifying up to 60% of the plasma proteome in accordance with our *in silico* predictions. Integrating SysQuan into recent strategies that have considerably increased the depth of plasma proteomics (25, 26) may help reaching this ambitious goal.

The human liver atlas contains 10,528 unique proteins (27), identified through the use of diverse immortalized liver cell lines, primary cells, and human biopsies, amounting to a total of 34 different sample types. Using mixed human/mouse liver samples, we were able to identify 96% of the known human liver proteome, and our data indicate that 70% of this known liver proteome is quantifiable using SysQuan, which aligns well with our *in silico* predictions.

Labeling Efficiency

Krüger *et al* reported average SILAC mouse labeling efficiencies of >95% after four generations of labeling (8). To determine our labeling efficiency, we analyzed the SILAC mouse liver as described above for the McGill mouse/human mixes after extensive prefractionation (48 fractions, timsTOF HT). For each identified mouse peptide, we used the SILAC light and heavy channel intensities to determine the % labeling efficiency. The average labeling efficiency over 70,287 peptides representing 7703 mouse proteins was 96.0% (see

Supplemental Data Fig. S2), which is well in-line the data from Krüger *et al*. Even for synthetic SIL peptides, a labeling efficiency >95% is widely considered sufficient for absolute quantitation. Importantly, the knowledge on the exact labeling efficiency for individual mouse peptides in a given SysQuan tissue can be leveraged to correct human protein concentrations and further increase quantitative accuracy—similar to correction factors for isobaric stable isotope labeling methods.

Translating SysQuan to Targeted Proteomics

We envision that once mouse protein concentrations have been determined, SysQuan enables proteome-wide absolute quantitation of human proteins. While nontargeted proteomics methods enable in-depth proteome analyses, targeted MS with scheduling and SIL peptides has been used for absolute quantitation of 100s of proteins in a single LC-MS run with unmatched precision, dynamic range, and sensitivity. Targeted MS can provide access to proteins that are below the limit of detection/quantitation of nontargeted methods.

To demonstrate the potential of SysQuan for absolute quantitation of human samples, we focused on a few liver proteins that map to metabolic pathways. From 1002 quantifiable proteins associated with various metabolic pathways, we selected 14 that cover various aspects of metabolism including electron transfer cycle, mitochondrial function,

tricarboxylic acid cycle, glycolysis, amino acid metabolism, beta-oxidation, and mitochondrial transport (Supplemental Data Table S1). These included proteins corresponding to the following genes: ASS1, COX7C, DLAT, GAPDH, GOT1, HADH, NDUFA4, NDUFS4, NDUFV1, NDUFV2, SLC25A3, SUCLA2, and SUCLG1, for which we purchased light synthetic standard peptides (NAT; from MRM Proteomics Inc). NAT purity was determined using capillary zone electrophoresis and NAT quantity was determined using amino acid analysis. For all 14 targets, MRM conditions were optimized on an Agilent 6495D using the inexpensive NAT peptides. Next, we mixed heavy mouse/light human liver digests 1:1 (w/w) and measured the corresponding heavy/light pairs by MRM, followed by a calibration curve with 0.1, 1, 10, 100, and 1000 fmol of each NAT in a constant SIL mouse liver background. Light/heavy peak ratios were determined to calculate human liver protein concentrations in fmol per μg liver tissue for all 14 target proteins (see Table 1, Supplemental Data Table S3), covering more than two orders of magnitude in dynamic range.

Next, we compared these results of the determined human liver protein absolute concentrations with data from proteomicsDB.org (28), where protein abundance in individual cells or tissues is estimated based on intensity-based

absolute quantitation (29) distributions over dozens of datasets. Although not accurate and not absolute in terms of referencing to standards, such estimates of protein abundance based on spectral counts or intensities have proven useful to get an overall idea of the composition of a given proteome (30, 31). Interestingly, and despite the completely different analytical workflows, protein concentrations determined by SysQuan correlated well with an intensity-based absolute quantitation estimates from proteomicsDB.org, with an R^2 of 0.89, which is an excellent confirmation of both methodologies (Fig. 3).

These results confirm that absolute quantitation using SysQuan is feasible. The same SILAC tissue reference standard can be used for characterizing 100s-1000s of samples, thus representing a cost-effective and concurrently more comprehensive alternative to current SIL peptide-based absolute quantitation. A recent study using SIL peptides and guided data-independent acquisition (DIA) parallel accumulation and serial fragmentation (g-dia-PASEF) reported absolute quantitation of 579 human proteins over a dynamic range of 4 orders of magnitude in depleted plasma samples (32). We have recently tested that state-of-the-art LC and triple quadrupole instrumentation allow reducing the run time for multiplexed MRM of plasma protein quantification by almost four

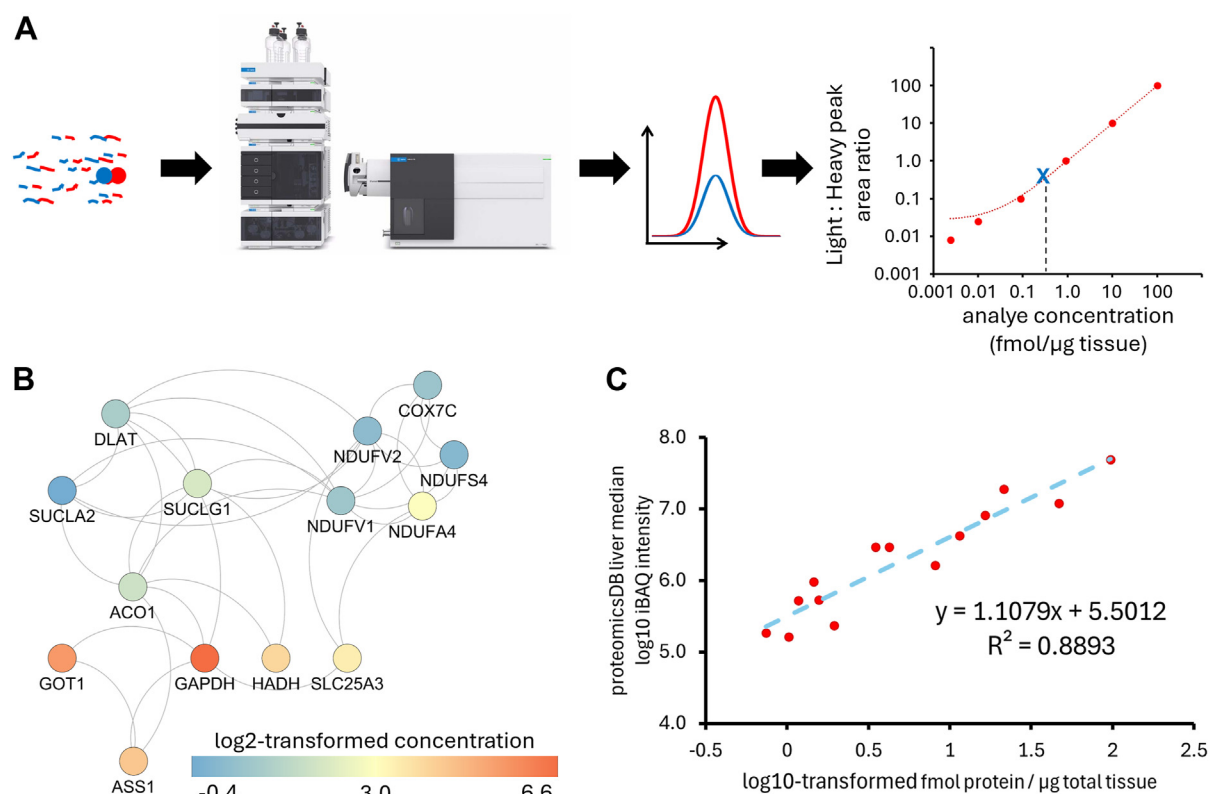


FIG. 3. **Absolute quantitation of metabolomic pathway proteins.** A, overall workflow. B, STRING protein/protein interaction network. C, correlation of SysQuan-derived concentrations and iBAQ-derived estimates for protein levels in proteomicsDB. iBAQ, intensity-based absolute quantitation.

times while maintaining depth and quantitative precision. Based on these results, we believe that the available LC-MS technology already enables quantifying 1000 proteins by targeted MS in 30 to 60 min (i.e., targeting both light and heavy forms of one peptide per protein and having a fairly even distribution of retention time). While this is not comparable to the proteome depth of nontargeted approaches, targeted MS provides improved precision (and accuracy) and unlocks access to a wider dynamic range, thus particularly enabling the quantitation of very low-abundance targets that are otherwise not accessible. This is particularly important as in many studies, precise measurement of a smaller but defined set of proteins across a large number of samples and without missing values is preferable over proteome-wide analysis with less precision and more incomplete data.

Notably, SysQuan-based absolute quantitation can be transferred to nontargeted methods, such as DDA or DIA-based label free quantitation. Interestingly, using nontargeted SysQuan for absolute quantitation provides a unique feature: all light/heavy doublets present in an acquired dataset can be retroactively absolutely quantified once the concentrations of the same mouse reference standard have been determined, i.e., even years post acquisition.

Precision and Accuracy

Quantifying proteins using proteotypic peptides as surrogates is widely accepted as precise, i.e. reproducible, and is the basis of relative quantitation in bottom-up proteomics. Achieving accuracy, i.e. determining protein concentrations that are close to the true value, however, is considerably more challenging and in the absence of reference values virtually not possible. It requires additional effort, including optimizing and determining digestion efficiency associated with the target proteotypic peptide release, as well as determining recovery of the protein and the surrogate peptides. While all of this can be easily performed in surrogate matrices, for example by spiking a recombinant protein in a matrix devoid of endogenous protein, it is difficult to perform in patient samples with often unknown levels of the protein-of-interest. Standard addition using certified authentic material of known concentration enables accurate quantitation (33), but such certified materials are, unfortunately, not available for most proteins, and it is often impossible or impractical to obtain them (34).

What matters most in practice and for preclinical applications is precision, since it encompasses repeatability and reproducibility. For most applications, from discrimination between treatment groups to identification of biomarker candidates, precision is sufficient as it enables comparison of data, if internal standards are being used also across platforms and laboratories. Traditional targeted proteomics using internal standards provides the required precision, regardless of whether these internal standards are synthesized or SILAC mouse-derived SIL peptides. Notably, the use of SIL proteins

to generate SIL peptide internal standards out of sequence stretches that are common between humans and mice, adds an inherent digestion control, addressing one of the main challenges of absolute quantitation using SIL peptides. This has been partially addressed by strategies such as QconCAT (35) where stable isotope labeled proteins are expressed and used as internal standards.

Reproducibility in SysQuan

The use of SIL mouse tissues and biofluid as internal standard stocks improves precision of relative quantitation and enables comparing large sample numbers even measured in different labs, as human proteins can be referenced to the same internal standard. Once the concentrations of a set of mouse proteins are determined in a given reference SIL tissue stock, this standard enables absolute quantitation of those proteins across 10,000s of samples. Concurrently, this overcomes two major challenges that impair the precision and accuracy of current absolute quantitation methods: matrix effects and variability in digestion efficiency (10, 36), both of which cannot be controlled using SIL peptides (9).

To test this, we analyzed the liver samples at two different laboratories, with different sample preparation, different pre-fractionation procedures, different reverse LC systems and different mass spectrometers. We found a very good agreement of protein quantification between the two acquisitions with two main insights (Fig. 4). First, in both measurements the \log_2 ratio between human and SILAC SIL mouse peptides were centered around zero, indicating that the concentration of the human proteins and released peptides by digestion are very similar to their SILAC SIL mouse counterparts. This reinforces the idea that matched SILAC mouse tissue can be used effectively to quantify human tissue “out-of-the-box” with minimal matrix effect. Alternative methods of generating SILAC peptides in cell culture (37–41) cannot account for this since proteome composition of a specific cell line does not accurately reflect the dynamic range and abundance distribution in differentiated tissues. Second, we assessed the agreement between the two quantification measurements using Bland–Altman analysis (42). We calculated that 95.35% of all quantifications were within the upper and lower limit of agreement—defined by mean of quantification difference \pm standard deviation of quantification difference. This robustness across different laboratory experimental conditions suggests that SysQuan quantification is reproducible, reliable, and independent of laboratory, sample preparation procedure, or instrumentation. This is a crucial aspect for translational research applications and should enable application of proteomics in longitudinal and population-wide studies.

DISCUSSION

High-throughput molecular phenotyping methods, such as RNA sequencing, metabolomics, and proteomics, are the

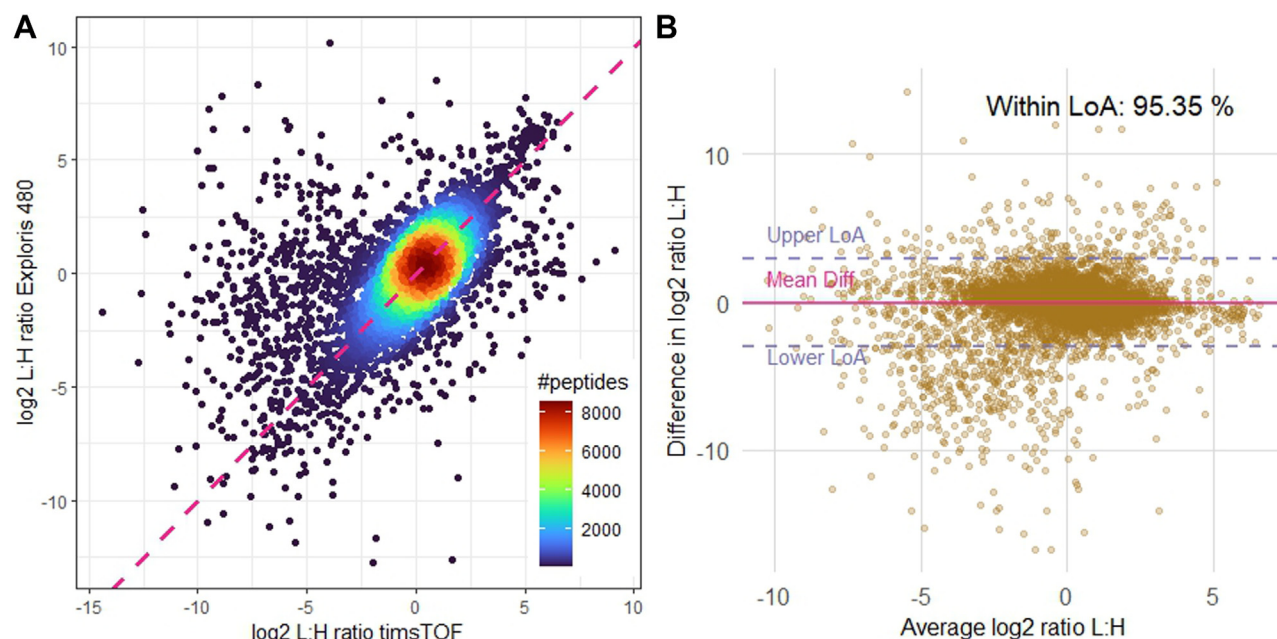


FIG. 4. **Agreement of quantitation between two sites of measurement with different sample preparation and instrumentation.** A, regression between the two analysis on two different liquid chromatography systems as well as mass spectrometers showing that most of the Log₂ ratios are centered around 0. B, Bland–Altman analysis showing an agreement of 95.35% between two measurements. LoA, limit of agreement; Mean Diff, mean difference.

cornerstones of modern biological and clinical studies, providing crucial mechanistic insights at the molecular level. Despite technological advancements, however, a key challenge remains—the affordability of these methods as precise and selective tools for determining and quantifying system-wide molecular changes. In contrast, the remarkable achievements in genomics can largely be attributed to efforts focused on making sequencing both rapid and reproducible at reduced cost.

MS-based proteomics has made significant strides in recent years, transitioning from qualitative to quantitative approaches. Label-free quantification techniques have emerged as cost-effective methods for biomarker discovery studies and hypothesis generation. However, despite the considerable increase in throughput accompanied by an increase in proteome coverage and concurrent decrease in cost per sample, experimental designs are increasingly facing constraints. These constraints arise from the need that samples need to be processed and analyzed together to avoid considerable batch effects that impair statistical analysis and consequently the identification of relevant biological changes. The only way to address these limitations is through determining the *amount* of a protein relative to a spiked-in internal standard, which allows subtle changes in protein levels can be detected. This enables more reliable comparisons between sample sets even across different batches and regardless of whether these batches have even been analyzed on the same platform. However, the high cost

of chemical synthesis of heavy labeled internal standards becomes prohibitive when quantifying large numbers of proteins, creating a significant barrier to widespread adoption. SysQuan can overcome this barrier by making quantitative proteomics using internal standards—and even system-wide absolute quantitation—more accessible and affordable. This is exemplified by the high coverage of the plasma and liver proteomes achieved in this study. This coverage can be further increased by the use of alternative enzymes: *in silico* digestions with LysC and GluC indicate that an additional ~1000 human proteins may be quantifiable compared to trypsin. Even the use of low-specificity enzymes such as subtilisin and thermolysin might be feasible (43), however, needs empirical testing as it is currently impossible to reliably predict their cleavage products.

Importantly, SysQuan also enables post acquisition absolute quantitation from untargeted proteomics datasets. Using NAT peptides to 1) quantify SIL proteins in SILAC mouse tissue 2) to be used as internal standard for reverse quantitation of human proteins in 100s of samples is significantly more cost-effective than the current gold-standard using synthetic SIL peptides (44).

Obtaining SILAC-labeled proteins can however also be done using SILAC-labeled cell lines instead of SILAC mouse tissues. One can argue against scaling up using cell cultures from two points of view: proteome coverage and cost. Proteomes obtained from cell lines compared to tissue or biofluid are different. For example, the HeLa proteome as identified

from a largest collection (39) represent less than half, i.e. 46% of the human plasma proteome as present in PeptideAtlas (40). C2C12 cells, used to study many aspects of muscle biology, are not well-reflective of the complexity of mouse muscle tissue—neither in terms of qualitative nor quantitative composition (41). The same is true for many other tissues and cell lines. Importantly, cell lines also do not account for sex-specific differences in protein composition of tissues that can be dramatic, especially in disease context. Regarding cost-effectiveness, it can be easily demonstrated that once we use various SILAC mouse tissues for tissue specific proteomes the costs will be lower compared to obtaining the same amount of labeled proteins from SILAC cells. Only the liver of a SILAC mouse contains approximately 170 mg of protein; it requires around 14 SILAC kits to obtain a similar amount of labeled protein from HeLa cells. Because a SILAC mouse has multiple labeled organs and also plasma, it can provide internal standards for more than a million samples—considering 100 ng of protein standard per sample which makes it easily more cost-effective than SILAC cells.

We believe that SysQuan can lead to a paradigm shift, as it enables researchers to replace relative quantitation with affordable absolute quantitation at scale, thus making every single proteomics study and individual sample considerably more informative and valuable for the community. This will be particularly relevant in the study of rare diseases, typically suffering from limited availability of patient samples. The ability to directly compare individual patients and studies, even when single cases have been analyzed and published years earlier, will transform rare disease research. We envision a utility for SysQuan as companion diagnostic and even diagnostic tool, as beyond classical absolute quantitation of biomarkers it will provide direct and clear evidence for the dysregulation of key metabolic pathways—without the need to run patients samples in batches with controls. The ability to compare protein concentrations across different human tissues and different diseases will enable novel study designs that are impossible with current technology.

SysQuan can be used for both targeted proteomics (MRM and parallel reaction monitoring) and nontargeted proteomics (DDA and DIA). SysQuan will, for the first time, enable researchers to retroactively quantify protein concentrations in DDA/DIA datasets, even if mouse SIL standard protein concentrations are determined years post acquisition. This unique feature opens new avenues for the reanalysis and repurposing of existing datasets within the scientific community (11, 45).

Similarly, the capability to universally compare protein concentrations across datasets will drive entirely new, community-driven studies based on the reanalysis and repurposing of datasets in public repositories such as PRIDE (46) and MassIVE (47), further enhancing their value and impact on biomedical and fundamental research. The ability to retroactively determine protein concentrations in existing

datasets, once the corresponding SIL mouse protein concentrations have been determined, opens up virtually unlimited possibilities for discovery.

In conclusion, we believe that SysQuan has the potential to transform fundamental and translational biomedical research, addressing the reproducibility crisis in life science (48) by making standardized quantitative proteomics broadly accessible (49).

DATA AVAILABILITY

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE, MassIVE, and PanoramaWeb partner repositories. The plasma and liver data acquired at McGill are available in PRIDE under the dataset identifier PXD057266. The liver data acquired at UofM is available in MassIVE under the dataset identifier MSV000096198. The targeted proteomics quantification data is available from PanoramaWeb using the following link: <https://panoramaweb.org/jJdH00.url>.

Supporting Information—This article contains [supporting information](#).

Acknowledgments—C. H. B. is grateful for the support from the Segal McGill Chair in Molecular Oncology at McGill University. C. H. B. is grateful for support from the Warren Y. Soper Charitable Trust for the Warren Y. Soper Clinical Proteomics Centre at the Jewish General Hospital (Montréal, Quebec, Canada). C. H. B. is also grateful for the support from the Terry Fox Research Institute and the Alvin Segal Family Foundation for the Segal Cancer Proteomics Centre at the Jewish General Hospital (Montréal, Quebec, Canada). R. P. Z. is grateful for support from the University of Manitoba, Shared Health Manitoba, and support from the H. E. Sellar Research Chair in Internal Medicine.

This work was done under the auspices of a Memorandum of Understanding between McGill and the U.S. National Cancer Institute's International Cancer Proteogenome Consortium (ICPC). ICPC encourages international cooperation among institutions and nations in proteogenomic cancer research in which proteogenomic datasets are made available to the public. This work was also done in collaboration with the U.S. National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC). [Figures 1 and 3](#) instrument cartoons were created in BioRender by Richard, V. (2024), <https://BioRender.com/d21u030>.

Author Contributions—Y. M., V. R. R., R. P. Z., and C. H. B. writing—review and editing; Y. M., V. R. R., and R. P. Z. writing—original draft; Y. M., R. P. Z., and C. H. B. conceptualization; Y. M. and R. P. Z. visualization; Y. M. data curation; Y. M. software; V. R. R., M. I. R. M., Y. L., V. S., R. P., C. G., and L. H. formal analysis; W. K., R. P. Z., and C. H. B. supervision; V. R.

R. investigation; R. P. Z. and C. H. B. funding acquisition; R. P. Z. and C. H. B. resources.

Funding and additional information—C. H. B. and R. P. Z. are also grateful to Genome Canada and Genome Quebec for the funding via the MutaQuant GAPP (#6567 Borchers_Zahedi_APf). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of Interest—CHB is the Scientific Advisor of MRM Proteomics, Inc, and the VP of Proteomics at Molecular You. R. P. Z. is the Scientific Advisor of MRM Proteomics, Inc. R. P. is the Chief Operating Officer of MRM Proteomics, Inc. Y. M. is the Chief Bioinformatics Officer of MRM Proteomics, Inc. C. G. is a Senior Scientist at MRM Proteomics, Inc. The other authors declare no competing interests. A patent application has been filed for SysQuan.

Abbreviations—The abbreviations used are: BCA, bicinchoninic acid; DDA, data-dependent acquisition; DIA, data-independent acquisition; FA, formic acid; IAA, iodoacetamide; MRM, multiple reaction monitoring; MS, mass spectrometry; PASEF, parallel accumulation serial fragmentation; SIL, stable isotope labeled; TCEP, tris(2-carboxyethyl)phosphine.

Received November 7, 2024, and in revised form, March 4, 2025
Published, MCPRO Papers in Press, April 18, 2025, <https://doi.org/10.1016/j.mcpro.2025.100974>

REFERENCES

- Messner, C. B., Demichev, V., Wang, Z., Hartl, J., Kustatscher, G., Müller, M., et al. (2023) Mass spectrometry-based high-throughput proteomics and its role in biomedical studies and systems biology. *Proteomics* **23**, e2200013
- Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W., and Gygi, S. P. (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6940–6945
- Addona, T. A., Abbatiello, S. E., Schilling, B., Skates, S. J., Mani, D. R., Bunk, D. M., et al. (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **27**, 633–641
- Lange, V., Picotti, P., Domon, B., and Aebersold, R. (2008) Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222
- Carr, S. A., Abbatiello, S. E., Ackermann, B. L., Borchers, C., Domon, B., Deutsch, E. W., et al. (2014) Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Mol. Cell. Proteomics* **13**, 907–917
- Mohammed, Y., Goodlett, D., and Borchers, C. H. (2023) Absolute quantitative targeted proteomics assays for plasma proteins. *Methods Mol. Biol.* **2628**, 439–473
- Wang, Q., Zhang, M., Tomita, T., Vogelstein, J. T., Zhou, S., Papadopoulos, N., et al. (2017) Selected reaction monitoring approach for validating peptide biomarkers. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 13519–13524
- Krüger, M., Moser, M., Ussar, S., Thievensen, I., Luber, C. A., Forner, F., et al. (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353–364
- Shuford, C. M., Walters, J. J., Holland, P. M., Sreenivasan, U., Askari, N., Ray, K., et al. (2017) Absolute protein quantification by mass spectrometry: not as simple as advertised. *Anal. Chem.* **89**, 7406–7415
- Proc, J. L., Kuzyk, M. A., Hardie, D. B., Yang, J., Smith, D. S., Jackson, A. M., et al. (2010) A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *J. Proteome Res.* **9**, 5422–5437
- Vaudel, M., Burkhart, J. M., Zahedi, R. P., Oveland, E., Berven, F. S., Sickmann, A., et al. (2015) PeptideShaker enables reanalysis of MS-derived proteomics data sets. *Nat. Biotechnol.* **33**, 22–24
- Sielaff, M., Kuharev, J., Bohn, T., Hahlbrock, J., Bopp, T., Tenzer, S., et al. (2017) Evaluation of FASP, SP3, and iST protocols for proteomic sample preparation in the low microgram range. *J. Proteome Res.* **16**, 4060–4072
- Meier, F., Beck, S., Grassl, N., Lubeck, M., Park, M. A., Raether, O., et al. (2015) Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. *J. Proteome Res.* **14**, 5378–5387
- Schwenk, J. M., Omenn, G. S., Sun, Z., Campbell, D. S., Baker, M. S., Overall, C. M., et al. (2017) The human plasma proteome draft of 2017: building on the human plasma PeptideAtlas from mass spectrometry and complementary assays. *J. Proteome Res.* **16**, 4299–4310
- Tabb, D. L., Wang, X., Carr, S. A., Clauser, K. R., Mertins, P., Chambers, M. C., et al. (2016) Reproducibility of differential proteomic technologies in CPTAC fractionated xenografts. *J. Proteome Res.* **15**, 691–706
- Uzozie, A. C., Ergin, E. K., Rolf, N., Tsui, J., Lorentzian, A., Weng, S. S. H., et al. (2021) PDX models reflect the proteome landscape of pediatric acute lymphoblastic leukemia but divert in select pathways. *J. Exp. Clin. Cancer Res.* **40**, 96
- Dumas, T., Martinez Pinna, R., Lozano, C., Radau, S., Pible, O., Grenga, L., et al. (2024) The astounding exhaustiveness and speed of the Astral mass analyzer for highly complex samples is a quantum leap in the functional analysis of microbiomes. *Microbiome* **12**, 46
- Kelly, R. T. (2020) Single-cell proteomics: progress and prospects. *Mol. Cell. Proteomics* **19**, 1739–1748
- Rosenberger, F. A., Thielert, M., and Mann, M. (2023) Making single-cell proteomics biologically relevant. *Nat. Methods* **20**, 320–323
- [Preprint] Xiong, Y., Tan, L. I., Chan, W.-K., Yin, E. S., Donepudi, S. R., Ding, J., et al. (2024) Ultra-fast multi-organ proteomics unveils tissue-specific mechanisms of drug efficacy and toxicity. *bioRxiv*. <https://doi.org/10.1101/2024.09.25.615060>
- Park, S. K., Liao, L., Kim, J. Y., and Yates, J. R. (2009) A computational approach to correct arginine-to-proline conversion in quantitative proteomics. *Nat. Methods* **6**, 184–185
- Bicho, C. C., de Lima Alves, F., Chen, Z. A., Rappsilber, J., and Sawin, K. E. (2010) A genetic engineering solution to the "arginine conversion problem" in stable isotope labeling by amino acids in cell culture (SILAC). *Mol. Cell. Proteomics* **9**, 1567–1577
- Van Hoof, D., Pinkse, M. W., Oostwaard, D. W., Mummery, C. L., Heck, A. J., and Krijgsvel, J. (2007) An experimental correction for arginine-to-proline conversion artifacts in SILAC-based quantitative proteomics. *Nat. Methods* **4**, 677–678
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361
- Blume, J. E., Manning, W. C., Troiano, G., Hornburg, D., Figa, M., Hesterberg, L., et al. (2020) Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nat. Commun.* **11**, 3662
- [Preprint] Wu, C. C., Tsantilis, K. A., Park, J., Plubell, D., Sanders, J. A., Naicker, P., et al. (2024) Mag-Net: rapid enrichment of membrane-bound particles enables high coverage quantitative analysis of the plasma proteome. *bioRxiv*. <https://doi.org/10.1101/2023.06.10.544439>
- Niu, L., Geyer, P. E., Gupta, R., Santos, A., Meier, F., Doll, S., et al. (2022) Dynamic human liver proteome atlas reveals functional insights into disease pathways. *Mol. Syst. Biol.* **18**, e10947
- Lautenbacher, L., Samaras, P., Muller, J., Grafberger, A., Shraideh, M., Rank, J., et al. (2022) ProteomicsDB: toward a FAIR open-source resource for life-science research. *Nucleic Acids Res.* **50**, D1541–D1552
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., et al. (2011) Global quantification of mammalian gene expression control. *Nature* **473**, 337–342
- Burkhart, J. M., Vaudel, M., Gambaryan, S., Radau, S., Walter, U., Martens, L., et al. (2012) The first comprehensive and quantitative analysis of human platelet protein composition allows the comparative analysis of structural and functional pathways. *Blood* **120**, e73–e82

31. Geiger, J., Burkhart, J. M., Gambaryan, S., Walter, U., Sickmann, A., and Zahedi, R. P. (2013) Response: platelet transcriptome and proteome-relation rather than correlation. *Blood* **121**, 5257–5258
32. Lesur, A., Bernardin, F., Koncina, E., Letellier, E., Kruppa, G., Schmit, P. O., et al. (2023) Quantification of 782 plasma peptides by multiplexed targeted proteomics. *J. Proteome Res.* **22**, 1630–1638
33. Mohammed, Y., Pan, J., Zhang, S., Han, J., and Borchers, C. H. (2018) ExSTA: external standard addition method for accurate high-throughput quantitation in targeted proteomics experiments. *Proteomics Clin. Appl.* **12**, 1600180
34. Cobbaert, C. M., Althaus, H., Begcevic Brkovic, I., Ceglarek, U., Coassin, S., Delatour, V., et al. (2021) Towards an SI-traceable reference measurement system for seven serum apolipoproteins using bottom-up quantitative proteomics: conceptual approach enabled by cross-disciplinary/cross-sector collaboration. *Clin. Chem.* **67**, 478–489
35. Pratt, J. M., Simpson, D. M., Doherty, M. K., Rivers, J., Gaskell, S. J., and Beynon, R. J. (2006) Multiplexed absolute quantification for proteomics using concatenated signature peptides encoded by QconCAT genes. *Nat. Protoc.* **1**, 1029–1043
36. Burkhart, J. M., Schumbrutski, C., Wortelkamp, S., Sickmann, A., and Zahedi, R. P. (2012) Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *J. Proteomics* **75**, 1454–1462
37. Pinho, J. P. C., Bell-Temin, H., Liu, B., and Stevens, S. M. J. (2017) Spike-In SILAC Approach for Proteomic Analysis of Ex Vivo Microglia. In: *Neuroproteomics. Methods in Molecular Biology*, Springer Nature: 295–312
38. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* **7**, 383–385
39. Webel, H., Perez-Riverol, Y., Webel, H., Perez-Riverol, Y., Nielsen, A. B., Rasmussen, S., et al. (2024) Mass spectrometry-based proteomics data from thousands of HeLa control samples. *Scientific data* **11**, 112
40. Desiere, F., Deutsch, E. W., King, N. L., Nesvizhskii, A. I., Mallick, P., Eng, J., et al. (2006) The PeptideAtlas project. *Nucleic Acids Res.* **34**, D655–D658
41. Deshmukh, A. S., Murgia, M., Nagaraj, N., Trebak, J. T., Cox, J., and Mann, M. (2015) Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol. Cell. Proteomics* **14**, 841–853
42. Bland, J. M., and Altman, D. G. (1986) Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* **1**, 307–310
43. Jiang, X., Yeung, D., Liu, Y., Spicer, V., Afshari, H., Lao, Y., et al. (2024) Accelerating proteomics using broad specificity proteases. *J. Proteome Res.* **23**, 1360–1369
44. Ong, S., and Mann, M. (2005) Mass Spectrometry-based proteomics turns quantitative. *Nat. Chem. Biol.* **1**, 252–262
45. Perez-Riverol, Y. (2022) Proteomic repository data submission, dissemination, and reuse: key messages. *Expert Rev. Proteomics* **19**, 297–310
46. Vizcaino, J. A., Cote, R. G., Csordas, A., Dianes, J. A., Fabregat, A., Foster, J. M., et al. (2013) The PRoteomics IDentifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41**, D1063–D1069
47. Choi, M., Carver, J., Chiva, C., Tzouros, M., Huang, T., Tsai, T. H., et al. (2020) MassIVE.quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods* **17**, 981–984
48. Baker, M. (2015) Blame it on the antibodies. *Nature* **521**, 274–276
49. Hollmann, S., Kremer, A., Baebler, Š., Trefois, C., Gruden, K., Rudnicki, W. R., et al. (2020) The need for standardisation in life science research - an approach to excellence and trust. *F1000Res* **9**, 1398