Published in final edited form as:

Anal Chem. 2013 November 19; 85(22): 10812–10819. doi:10.1021/ac4021352.

SILAC Surrogates: Rescue of quantitative information for orphan analytes in spike-in SILAC experiments

Jason M. Gilmore¹, Jeffrey A. Milloy², and Scott A. Gerber^{1,2,3,*}

¹Department of Genetics, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA

²Norris Cotton Cancer Center, Lebanon, NH 03756, USA

³Department of Biochemistry, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA

Abstract

Super-SILAC enables the sensitive and accurate analysis of complex biological tissue and tumor samples by comparison of light peptides observed in biological samples to heavy peptides from SILAC cell culture spike-ins. However, despite the use of multiple cell lines for Super-SILAC spike-in standards, the full protein and peptide profiles of biological samples are not completely represented in these internal standards, leading to orphan analytes for which sample to standard ratios cannot be calculated. This problem is exacerbated in some biological systems, such as muscle tissue, which lack adequate cell culture lines to reflect their complex and idiosyncratic protein profiles, resulting in up to 40% of peptide analytes without heavy cognates. Furthermore, these unquantified orphan analytes may be among the most biologically interesting and significant species, since their presence is not common to cell lines cultured *in vitro*. Here, we report on the development of a surrogate analysis strategy to interpolate quantitative relationships between peptide species, observed across multiple biological samples, which lack representation within the spike-in standards. The precision and accuracy of this method was assessed by replicate experiments in which surrogate-derived ratios from defined mixtures of spike-in SILAC standard and tissue lysate were compared against traditional SILAC ratios for species where both light and heavy peptide cognates were observed. We demonstrate the robustness of our SILAC Surrogates strategy across a variety of murine tissues, including liver, spleen, brain and muscle. Our approach increases the quantitative coverage and precision within a biological sample by rescuing previously intractable peptide species and applying additional evidence to improve the precision of existing quantifications.

Keywords

Quantitative proteomics; SILAC; internal standard

INTRODUCTION

The comprehensive quantification of protein abundance differences in tissues or in tumors is a critical component in the study of cellular processes that are dysregulated in cancer and other diseases. In particular, an accurate and sensitive catalog of protein profile changes would contain cancer biomarkers for earlier detection and tumor-specific protein profiles that could be used in personalized medicine to inform therapy decisions. However, the

*Corresponding Author, Scott A. Gerber, PhD; scott.a.gerber@dartmouth.edu.

ASSOCIATED CONTENT

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>.

analysis of tissues and tumors is complicated by sample complexity, stoichiometric limitations, and technical challenges associated with the accurate quantification of patient-derived samples. Mass spectrometry based proteomics is a powerful analytical platform for these studies, however, differences in ionization efficiency and detection of peptides confounds routine absolute quantification. To address these limitations, several relative quantitative proteomic methods have been developed, including dimethyl labeling¹, isotope-coded affinity tags and/or tandem mass tags (iTRAQ/TMT reagents)^{2, 3}, and spike-in SILAC⁴, among others⁵.

Dimethyl labeling uses formaldehyde to globally label the N-terminus and lysine residues of peptides through reductive amination. TMT and iTRAQ utilize isobaric labeling which is able to resolve several conditions or states in a single experiment through multiplex reagents by tandem mass spectrometry. Unfortunately, the use of TMT and iTRAQ reagents becomes cost-prohibitive to implement on a routine basis for large-scale phosphorylation analyses, owing to the relatively large amount of input (5 – 15 milligrams protein) typically employed for phosphopeptide enrichment⁶. In contrast, stable isotope labeling with amino acids in cell culture (SILAC) is a low cost, highly accurate quantification technique in which metabolic incorporation of heavy amino acids, typically lysine and arginine, is achieved through cell culture⁷. SILAC has been demonstrated to have high labeling efficiency in general, although some cell lines fail to incorporate labeled amino acids efficiently, or will metabolize them to other amino acids^{8, 9}. Additionally, SILAC is not directly applicable to tissue and tumor analyses, where instead Super-SILAC spike-in standards must be employed. Like dimethyl labeling, SILAC has been shown to be accurate and amenable to quantifying states of posttranslational modifications, such as phosphorylation¹⁰. Thus, spike-in SILAC presents as a simple and cost-effective strategy for quantitative proteomic analyses of phosphorylation in tissues and tumors.

In the standard spike-in SILAC workflow, whole cell lysates of target tissues are mixed with multiple cell-type matched and heavy-labeled standards, digested by a protease, separated into fractions, and analyzed by LC-MS/MS (Figure 1A). However, despite confident qualitative identification of tissue peptides, any lack of a heavy cognate in the spike-in SILAC digest prevents the quantitative comparison of that peptide across biological samples. This can occur either due to low abundance or failure to complement the protein profile of the biological sample of interest; we refer to these peptides as “orphan analytes” (Figure 1B). The orphan frequency is low for ideal super-SILAC experiments, although the number of orphans increases as the diversity of the standard decreases (Figure 1C)⁴.

To achieve this diversity, Super-SILAC experiments rely on the availability of multiple heavy labeled cell-type matched cell lines to be included in the standard. In our experience, however, this requirement is not always feasible. One elegant solution to this problem is to heavy-label an entire organism, and then use lysate from tissues of interest from this labeled animal as a common spike-in reference standard when performing multiple comparisons on other test animals (stable isotope labeling in mammals, SILAM)¹¹. However, many labs conducting quantitative proteomics experiments may not have the tools necessary to generate these labeled animals, and even under ideal conditions, variable levels of incorporation are often observed, which can also reduce the number of spike-in proteins that are quantifiable^{12, 13}. To address these and other limitations, we conducted a spike-in experiment for murine liver tissue and evaluated the orphan frequency when using a single heavy cell line as a standard. As expected, we observed a relatively high frequency of orphans, which provided us with a dataset whose characteristics we could evaluate for the purposes of designing a recovery strategy (Figure 1D).

Here, we develop and test the use of a surrogate internal standard approach to recover quantitative information for orphan analytes in spike-in SILAC experiments, in which nonorphan internal standard peptides are used to generate analysis-specific correction factors that enable quantification of these orphan peptides. We show that this method both accurately and precisely agrees with the quantifications obtained by traditional super-SILAC where available, as well as providing quantification information for a substantially greater number of peptides and proteins. Based on these results, we test our hypothesis that a single, readily heavy-labeled and easy-to-grow murine fibroblast cell line can be used as a generalized spike-in standard in this workflow, and use it to quantify abundance differences in murine liver, spleen, brain and muscle tissues with excellent quantitative precision and accuracy. Taken together, our results describe a novel extension of the spike-in SILAC quantification strategy that provides greater depth of quantitative coverage for tissue analyses without expensive or time consuming alterations to existing methods.

EXPERIMENTAL DETAILS

Materials

Modified trypsin was from Promega (Madison, WI). Urea, Tris-HCl, ammonium bicarbonate (NH_4HCO_3), sodium chloride (NaCl), potassium chloride (KCl), potassium phosphate (KH_2PO_4), phosphoric acid, sodium orthovanadate, sodium fluoride, sodium molybdate, sodium tartrate, beta-glycerophosphate, dl-dithiothreitol (DTT), iodoacetamide were from Sigma-Aldrich (St. Louis, MO). Acetonitrile (ACN), trifluoroacetic acid (TFA) and HPLC-MS grade water were from Honeywell Burdick and Jackson (Morristown, NH). Methanol (MeOH) was from Fisher (Pittsburgh, PA). High-purity formic acid was from EMD (Gibbstown, NJ). 500 mg sorbent C_{18} solid-phase extraction cartridges were from Grace Davidson and Oasis IHLB vacuum extraction plates were from Waters Corporation (Milford, MA). Dulbecco's modified Eagle's medium (DMEM), PBS, penicillin and streptomycin were from Invitrogen (Carlsbad, CA). Fetal bovine serum (dialyzed and undialyzed; Hyclone) was purchased from ThermoFisher Scientific (Pittsburgh, PA). Isotopically labeled [$^{13}\text{C}_6$, $^{15}\text{N}_2$]lysine and [$^{13}\text{C}_6$, $^{15}\text{N}_4$]arginine were obtained from Cambridge Isotope Laboratories Inc (Andover, MA). Murine hepatocytes (TIB-75) were kindly provided by Dr. James Gorham (Geisel School of Medicine).

Cell culture, lysis and digestion

TIB-75 (murine hepatocyte) and 3T3 (fibroblast) cells were grown as adherent cultures in arginine- and lysine-free DMEM, with 10% FBS and penicillin and streptomycin. Labeling was achieved by supplementing this media with isotopically heavy lysine and arginine, both at 100 mg/liter, for at least six cell doublings. For harvesting, cells were collected, washed with PBS and snap-frozen in liquid nitrogen. For lysis, cells were thawed on ice and lysed in lysis buffer (8 M urea, 25 mM Tris-HCl, 150 mM NaCl), phosphatase inhibitors (2.5 mM beta-glycerophosphate, 1 mM sodium fluoride, 1 mM sodium orthovanadate, 1 mM sodium molybdate, 1 mM sodium tartrate) and protease inhibitors (1 mini-Complete EDTA-free tablet per 10 ml lysis buffer; Roche Life Sciences, Mannheim, Germany). Mouse tissues were homogenized first using a dounce homogenizer and then sonicated three times at 30 – 40% power for 15 sec each in lysis buffer with intermittent cooling on ice, followed by centrifugation at $15,000 \times g$ for 30 min at 4 °C to clarify the lysate. The lysates were then reduced with DTT at a final concentration of 5 mM and incubated for 30 min at 50 °C. Afterwards, lysates were thoroughly cooled to room temperature (~22 °C) and alkylated with 15 mM iodoacetamide at room temperature for 45 min. The alkylation was then quenched by the addition of an additional 5 mM DTT. After sixfold dilution with 25 mM Tris-HCl pH 8 and 1 mM CaCl_2 , the sample was digested overnight at 37 °C with 1% (w/w) trypsin. The next day, the digest was stopped by the addition of 0.25% TFA (final v/v),

centrifuged at 3,500×g for 30 min at room temperature to pellet precipitated lipids, and desalted on a C₁₈ cartridge (wash: MeOH; equilibration: 3% MeOH, 0.1% TFA; elution: 60% MeOH, 0.1% formic acid). Desalted peptides were lyophilized and stored at –80 °C until further use.

SCX Chromatography

Peptides from mouse liver were independently mixed at three dilutions (1:1, 1:4, and 4:1, all L:H) with either heavy labeled TIB-75 or 3T3 cells. The liver-to-TIB-75 mixing was performed with four separate, technical replicates; each replicate was independently separated by strong cation exchange (SCX) chromatography as described below. The other mouse tissues were mixed as before but with only 3T3 heavy standard. 250 micrograms of peptides mixed in SCX buffer A (7 mM KH₂PO₄, pH 2.65/30% ACN) were separated per injection on a SCX column (Luna SCX, Phenomenex; 150 × 2.0 mm, 5 μm 100 Å pore). We used a gradient of 0 to 11% SCX buffer B (350 mM KCl/7 mM KH₂PO₄, pH 2.65/30% ACN) over 11 min, 11% to 26% SCX buffer B over 11 min, 26% to 54% SCX buffer B over 7 min, 54% to 100% SCX buffer B over 1 min, holding at 100% SCX buffer B for 5 min, from 100% to 0% SCX buffer B over 2 min, and equilibration at 0% SCX buffer B for 65 min, all at a flow rate of 0.22 ml/min. After a full blank injection of the same program was run to equilibrate the column, a 250 microgram sample was injected on to the HPLC, and 24 fractions were collected from the onset of the void volume (2.2 min) until the elution of strongly basic peptides at 100% SCX buffer B (52 min), at 2.075-min intervals. After separation, the SCX fractions 12–17 were lyophilized and desalted using a OASIS μHLB C₁₈ 96-well desalting plate and manifold (wash: MeOH; equilibration: 3% MeOH, 0.1% TFA; elution: 60% MeOH, 0.1% formic acid). These contiguous fractions spanned the +2 solution charge regions of those chromatograms, were selected based on peptide abundance, and included less abundant flanking fractions (fractions 12 and 17). The liquid eluate from the OASIS plate (60 μl) was transferred to deactivated glass micro inserts (Agilent), dried by vacuum centrifugation directly in inserts and analyzed by LC-MS/MS.

LC-MS/MS Analysis

LC-MS/MS analysis was performed on a LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) equipped with an Agilent 1100 capillary HPLC, FAMOS autosampler (LC Packings, San Francisco, CA) and nanospray source (Thermo Fisher Scientific). Peptides were redissolved in 6% MeOH/1% formic acid and loaded onto an in-house packed polymer-fritted trap column at 2.5 μl/min (1.5 cm length, 100 μm inner diameter, ReproSil, C₁₈ AQ 5 μm 200 Å pore (Dr. Maisch, Ammerbuch, Germany)) vented to waste via a micro-tee. The peptides were eluted by split-flow at ~800–1,000 psi head pressure from the trap and across a fritless analytical resolving column (16 cm length, 100 μm inner diameter, ReproSil, C₁₈ AQ 3 μm 200 Å pore) pulled in-house (Sutter P-2000, Sutter Instruments, San Francisco, CA) with a 50 min gradient of 5–30% LC-MS buffer B (LC-MS buffer A: 0.0625% formic acid, 3% ACN; LC-MS buffer B: 0.0625% formic acid, 95% ACN).

An LTQ-Orbitrap (LTQ-Orbitrap MS control software v. 2.5.5, build 4 (06/20/08); previously tuned and calibrated per instrument manufacturer's guidelines using caffeine, MRFA, and UltraMark "CalMix") method consisting of one Orbitrap survey scan (AGC Orbitrap target value, 700 K; R = 60 K; maximum ion time, 800 ms; mass range, 400 to 1,400 m/z; Orbitrap "preview" mode enabled; lock mass set to background ion 445.120029) was collected, followed by ten data-dependent tandem mass spectra on the top ten most abundant precursor ions (isolation width, 1.6 m/z; CID relative collision energy (RCE), 35%; MS1 signal threshold, 12,500; AGC LTQ target value, 3,500; maximum MS/MS ion time, 125 ms; dynamic exclusion: repeat count of 1, exclusion list size of 500 (max), 24 s

wide in time, ± 20 ppm wide in m/z . Doubly- and triply charged precursors were selected for MS/MS and no neutral-loss dependent or multi-stage activation methods were employed).

Peptide spectral matching and bioinformatics

Raw data were searched using SEQUEST^{14–16} against a target-decoy (reversed) version of the murine proteome sequence database (UniProt; downloaded 6/2011; 92,042 total (forward and reverse) proteins) with a precursor mass tolerance of ± 1 Da¹⁷ and requiring fully tryptic with up to two mis-cleavages, carbamidomethylcysteine as a fixed modification and oxidized methionine, heavy Lys (+ 8.01420 Da) and heavy Arg (+ 10.00827) as variable modifications. The resulting peptide spectral matches were manually filtered to $< 1\%$ false discovery rate, based on reverse-hit counting (typical cutoffs include mass measurement accuracy within ± 2.5 ppm, a delta-XCorr (dCn) of greater than 0.08, and XCorr values of greater than 1.5 for +2-charge state peptides and greater than 2 for +3-charge state peptides). Peptide abundances were calculated using the LC-MS quantification software MassChroQ¹⁸. Data filtering and comparative analyses were performed using the R statistical programming language (<http://www.R-project.org>). Summary information for all peptide assignments can be found in Supporting Information online. All raw LC-Orbitrap data files can be found on our lab website at <http://proteomics.dartmouth.edu>.

RESULTS AND DISCUSSION

Orphan frequency in super-SILAC based quantitative proteomics

Spike-in SILAC experiments rely on protein profile complementarity between biological tissues of interest and a standard composed of isotopically labeled cell lines. When heavy cognates of tissue peptides are not observed, orphan analytes result. Importantly, these orphan peptides may be among the most biologically interesting and significant species in an experiment, since their expression is unique to an organism and is not present in cultured cells *in vitro*. We began by obtaining and analyzing raw LC-MS/MS data from the original spike-in SILAC publication⁴ from the online data repository Tranche¹⁹. In that experiment, human breast cancer tumor samples were compared to one another using either one or a mixture of five heavy labeled breast cancer cell lines. After database searching and filtering to a $< 1\%$ false discovery rate (FDR), we calculated the orphan frequency as the number of peptides for which the light peptide and not its heavy cognate were quantified, divided by the total number of unique quantified light peptides. When using all five breast cancer cell lines as the heavy standard, the orphan frequency in this canonical example was just less than 10% (Figure 1C); the number of orphans jumps to almost 33% when only one heavy breast cancer cell line is employed as the spike-in standard. In our own experience involving attempts at spike-in SILAC in mouse liver tissue, only one out of several attempts produced a mouse liver cell line (TIB-75; chemically immortalized murine hepatocytes) that readily incorporated heavy amino acids with minimal heavy proline artifacts (Figure S1, Supporting Information). After spiking in heavy TIB-75 into mouse liver digest in a 1:1 ratio based on protein abundance as measured by bicinchoninic acid (BCA) assay and analysis by LC-MS/MS, we observed a similar rate of orphans (36%; Figure 1D) as with a single heavy breast cancer cell line example above.

Surrogate analysis strategy

The calculation of a super-SILAC ratio for an observed light species fails when its heavy cognate is not quantified in either sample. One intuitive solution to address this latter point would be to increase the concentration of heavy peptides in the sample, in order to increase lower abundance heavy cognates to a quantifiable level relative to their light counterparts. To test this hypothesis, we prepared a dilution series where we mixed protein digests from mouse liver and heavy TIB-75 cells at 1:1, 4:1, and 1:4 dilutions (light:heavy), separated

those peptide mixtures by strong cation exchange (SCX) and collected six fractions across the +2 solution charge regions of those chromatograms, followed by LC-Orbitrap-MS/MS analysis and database searching for each fraction (Figure 2A). In total, the three different dilutions produced similar total numbers of qualitative peptide identifications (7417, 7017, and 6822 peptides identified in the 1:1, 4:1 and 1:4 dilutions, respectively) for the four middle SCX fractions. Quantification software was then used to generate light:heavy ratios for these peptides, and orphan analyte occurrences were calculated (Figure 2B). The number of peptides for which both light and heavy species were detected and quantified is indicated by the dark gray bars. Orphan peptides are indicated in medium gray and heavy peptides from the standard which did not correspond to any light species from the tissue sample are shown in light gray.

Although the 1:4 dilution sample produced fewer orphans than the 1:1 sample (1357 versus 2220 peptides, respectively; Figure 2B, medium grey), this was accompanied by a large reduction in the number of total spike-in SILAC quantifications in which both heavy and light species were observed (2854 versus 4579, respectively; Figure 2B, dark grey). This is likely due to a corresponding increase in the number of heavy-only quantitative observations (Figure 2B, light grey), with many light peptides falling off of the limit of quantification relative to the heavy standard. In contrast, the 4:1 sample produced a larger number of orphans (4087 peptides) but a similar number of successfully quantified SILAC ratios to that of the 1:4 (2617 peptides). Importantly, we noticed that the sum of successfully quantified species plus orphans (peptides observed only in their light form) for the 1:1 sample, was similar to the sum of these species for the 4:1 sample (6799 versus 6704, respectively; Figure 2B, dark plus medium grey). Our observations are consistent with at least two potential explanations: first, that although there are differences in specific protein abundances between TIB-75 cells and mouse liver tissue, the *distribution* of protein abundances between the two are relatively similar, and therefore have an optimum relative difference in that distribution; and second, that when additional heavy peptides are diluted into a sample, they outcompete light species in the stochastic ion selection scheme within the mass spectrometer, resulting in more time spent analyzing the standard and less time spent on the biological tissue of interest. In any case, for real-world quantitative analyses in practice, it is obviously the biological sample of interest that is to be quantified, and not the internal standard. Thus, there is clearly a point at which the addition of increasing amounts of internal standard reduces the number of tissue-specific analytes that could be quantified.

In a typical spike-in SILAC analysis, several hundred to thousands of heavy peptides will be detected in each fraction and in common across all biological samples. We therefore surmised that these readily and reliably quantified heavy peptides could serve as ‘surrogates’ to orphan peptides and, when taken together, could be used to interpolate the relationship for each orphan between samples. For this surrogate approach, we chose to calculate a correction factor based on the distribution of heavy/heavy peptide ratios between samples (Figure 3). We selected the median of the heavy/heavy ratios for this, so that the correction factor would be robust to outlier values. Using this correction factor, it is possible to assign a common surrogate ratio-of-ratios for every orphan analyte; if orphan analytes exist between two samples under comparison, then this surrogate ratio would allow for recovery of quantitative information for those orphans.

Stratification of surrogates

In developing our correction factor, we considered that perhaps not all heavy peptides should serve as surrogates. For example, we realized that a potential requirement for the use of a heavy peptide as a surrogate would be its unique observation in individual fractions when upstream fractionation is employed. In our SCX-based test experiment (Figure 2), any

heavy peptide whose elution profile is split across adjacent fractions may result in variations in the relative abundance of this peptide between two biological replicates that exhibit minor imprecision in the fractionation step. We tested this directly by calculating candidate surrogate ratios of heavy peptides that were either uniquely identified in a single fraction or identified in adjacent fractions in each dilution series (1:1, 4:1 and 1:4 light:heavy) relative to a replicate 1:1 dilution and plotting the median-corrected distributions of ratios for all SCX fractions on a log2 scale (Figure 2A). The percentage of heavy peptides that straddled fraction boundaries was roughly the same for each dilution series (~13%). Interestingly, for the 1:1 / 1:1 case, the distribution of fraction-unique heavy peptide ratios is relatively tight and only slightly better than those found in adjacent fractions (0.30 vs 0.34). However, for the 1:1 / 4:1 and 1:1 / 1:4 cases, heavy peptides found in adjacent fractions were much less consistently quantified between the two samples than those found uniquely in a single SCX fraction (1:1 / 4:1, 0.81 vs 0.87; 1:1 / 1:4, 0.71 vs 1.29). Further inspection of these ratio distributions on a per fraction basis demonstrates additional skewing of the medians of heavy peptide ratios that straddle fraction boundaries (Figure S2, Supporting Information). Based on these data, we chose to exclude this class of heavy peptides from further consideration as surrogates.

We also considered the possibility that certain peptides may exhibit greater tendency to generate imprecise ratios by virtue of other empirical or physicochemical properties. To explore this idea, we generated three additional 1:1 (light, liver tissue:heavy, TIB-75 lysate) protein digest dilutions, separated each of them by SCX chromatography, and collected six fractions at the same retention times. After desalting, all 18 (6 fractions/replicate, 3 replicates) additional fractions were analyzed by LC-Orbitrap-MS/MS exactly as was done before. We then applied the “unique fraction only” rule above to each replicate series within a dilution to filter out peptides observed in adjacent fractions. We also required that a heavy peptide feature be observed in all four replicate runs. This produced a list of 6088 heavy peptide sequences to be considered as candidate surrogates.

We evaluated these candidates by calculating the relative standard deviation (RSD) in heavy peptide peak area for each of them, and plotting their RSD versus various plausible features by which future surrogate candidates could be excluded. We rationalized the possibility that peptides that elute very early or very late during the online reverse-phase chromatographic separation could exhibit greater imprecision by virtue of the specific solvent composition during electrospray ionization, or due to other matrix effects that they experience at different elution profiles (Figure S3, Supporting Information). We also considered that peptides that display greater chromatographic peak tailing may result in greater imprecision in how these peaks were quantified by software. Finally, we surmised that peptides of greater length may be less precisely quantified, potentially due to a greater distribution of charge states or a combination of the two factors above. However, no significant trends in RSD were observed for any of these candidate surrogate features, although we did see a slight trend towards greater imprecision with increasing peptide length. Thus, in our experience to this point, each and every heavy peptide observation that i) is always observed in every replicate analysis of a given fractionated digest and ii) is never observed in more than one fraction can be considered a viable surrogate peptide.

Validation of SILAC Surrogates in murine liver

To evaluate the precision and accuracy of the surrogate strategy we utilized the dilution scheme outlined in Figure 2A, murine liver protein digest mixed with heavy labeled TIB-75cell protein digest. In this way, we were able to compare the population of surrogate-derived light/light ratios to both the observed spike-in SILAC ratios-of-ratios and to the expected light/light ratios for comparison of the 1:1 and 4:1 dilution samples. First, we determined correction factors on a fraction-specific basis for the four SCX fractions by

comparing the heavy peptide abundances (surrogates) between the two samples (Figure S4, Supporting Information); although these correction factors differed only modestly across all SCX fractions (~20% maximum deviation), they were not found to be identical. Next, we calculated spike-in SILAC ratios for all fully quantified peptides, i.e. those for which both light and heavy species were detected in both samples (non-orphan peptides). As expected for this specific dilution, these ratios-of-ratios were centered at -2 on a $\log_2$ scale (Figure 4A, first plot), although we observed a slight right skew to the data.

Non-orphan peptides can also be used to validate the surrogate method. The observed heavy quantifications for any light:heavy peptide pair can simply be ignored; these peptides can then be treated as orphans for the purposes of computing and evaluating surrogate ratios. Using this approach, the surrogate strategy recapitulated the distribution of spike-in ratios for non-orphans, and the right skew observed in the distribution of spike-in SILAC ratios of these peptides was greatly diminished when using the surrogate-derived correction factor (Figure 4A, second plot). Finally, the surrogate-derived ratios for the true orphan analytes in this dilution series were calculated and plotted, and displayed a similar distribution while providing several thousand additional quantifications (4019 beyond the 2559 for spike-in SILAC alone).

A second analysis, between SCX-based replicates of the 1:1 liver and TIB-75 dilution, further demonstrated the accuracy and utility of the surrogate strategy (Figure 4B). As previously shown, the orphan frequency of the 1:1 dilution is much lower than in the 4:1 mixing, and as a result there are many more successful spike-in SILAC ratios. However, across the four fractions there are still 2319 orphan peptides, for which application of the surrogate correction factor resulted in a 53% increase in total quantifications. These new quantifications are centered as expected and are evenly distributed in a manner very similar to the surrogate-derived non-orphan ratios. Together, these analyses suggest that the surrogate strategy is an accurate and precise supplement to the spike-in SILAC workflow and provides supporting evidence to existing quantifications as well as novel information for many otherwise unquantified peptides.

Generalization of surrogate strategy to additional tissues with a common internal standard

Having evaluated the surrogate strategy for the murine liver with a single cell-type matched heavy standard, we sought to both generalize our strategy and extend it to additional tissues. We recognized that the successful application of the surrogate correction factor was based primarily on the robust detection of many heavy peptide standards analysis-wide, in which an ensemble of heavy peptide quantifications produced a median correction factor that was robust to outliers, and not necessarily on the qualitative extent to which the heavy standard and the target tissue were matched in terms of protein composition. Indeed, deep proteomic analysis of multiple cancer cell lines derived from different organs and divergent cell types indicates that most cell lines express a similar complement of proteins, albeit at distinctly different levels²⁰. Based on these observations, we hypothesized that any easily cultured murine cell line that readily incorporates heavy amino acids without proline conversion would suffice, in combination with our surrogate analysis, to provide proteome-wide quantification for any mouse tissue. To test this hypothesis, we cultured Swiss mouse fibroblasts (3T3) for six doublings in heavy SILAC medium, and verified very high extent of labeling and minimal heavy proline conversion (Figure S1, Supporting Information). We then repeated the tissue analysis with 1:1 and 4:1 dilutions of liver with this heavy fibroblast cell lysate digest as an internal standard. The spike-in SILAC ratios for this analysis were centered as expected, and many orphan analytes (3,760) were not able to be quantified (Figure 5A). As with the TIB-75 spike-in experiment, the surrogate-derived ratios for non-orphan peptides showed a similar distribution to the spike-in SILAC ratios, validating our method for a heavy standard that was not cell-type matched. Furthermore, the orphan

recovery also produced quantifications centered at -2 on the $\log_2$ scale and distributed in a similar fashion with the rest of the data, supporting the use of a single generic standard for future experiments.

To test the robustness of the general heavy fibroblast spike-in standard in conjunction with the surrogate method, we extended our analysis to several additional tissue types. Murine tissue samples of spleen, brain and gastrocnemius muscle were also harvested, prepared as previously described and mixed with heavy labeled fibroblast peptides at 1:1 and 4:1 dilutions (Figure 5B–D). After separation by SCX, desalting and analysis by LC-Orbitrap-MS/MS, the total number of unique peptides quantified was higher for spleen and brain tissue (9933 and 9122 respectively), and lower for the muscle sample (7287). Within these experiments, the spleen sample was similar to the liver in the overall number of successful spike-in SILAC and rescued orphans while brain and muscle samples yielded fewer successful quantifications, even after surrogate rescue. For the brain and muscle samples, a greater percentage of the quantified species were from the heavy labeled fibroblast standard and not detected as light species in both dilution samples. Of the light species observed by tissue, orphan frequencies were 32%, 30%, and 39% for brain, spleen and muscle, respectively, as compared to 33% in the analogous liver analysis utilizing a 3T3-based heavy standard.

In our hands, the accuracy and precision of the liver sample mixed with 3T3 fibroblast standard was comparable to that of liver mixed with TIB-75 hepatocyte standard. Both showed surrogate-derived ratios which confirmed the spike-in SILAC quantifications for non-orphan peptides and rescued substantial numbers of orphan analytes with similar ratios-of-ratios (Figures 4B and 5A). Interestingly, the surrogate-derived ratios for non-orphans in the liver sample were more tightly distributed around the expected value on the $\log_2$ scale than the spike-in SILAC ratios. This is a result of the way the surrogate correction factor is calculated and applied to produce surrogate ratios. Since the correction factor is the median of many heavy/heavy peptide relationships between the samples it is unsurprising that it would be more robust to variability in these quantifications leading to a tighter distribution of ratios. The results for spleen were very similar to that of the liver though with a greater reliance on orphan rescue to obtain the final set of quantifications. The brain and muscle samples had lower yields than either of the other two tissues as well as wider distributions of peptide abundance ratios. In particular, the spike-in SILAC distribution for the muscle samples showed a stronger right skew than the other tissues and the number of peptides quantified was substantially lower. This may be due to a less diverse protein profile or an indication of protein degradation during the more difficult homogenization of the muscle tissue. Additional efforts to assess the utility of spike-in SILAC to the analysis of murine muscle tissues are currently ongoing.

CONCLUSIONS

Despite rapid technical advances and computational analysis techniques in proteomics, comprehensive quantification of tissues and tumors remains a challenging problem. Isotopic labeling strategies with spike-in standards provide powerful tools in pursuit of this problem, but robust, matched standards are often difficult, if not impossible, to generate. Some systems, such as muscle tissue, lack adequate standards while others require the combination and maintenance of complex heavy cell line mixtures to enable comparisons between biological samples. Still, portions of the proteome remain resistant to detection and quantification by classical approaches. Here we present data demonstrating that an interpolative strategy using quantification information from unrelated peptide sequences allows for precise and accurate rescue of relative peptide abundance ratios for orphan analytes. This is in principle similar to early reports of using single, standard protein digests

spiked in to tissue lysates as a common reference standard²¹, but extends that capability by providing thousands of peptide spike-ins that persist across a diverse and representative range of fractionation methods, and provide a robust median correction factor to improve the accuracy of differences in quantification. The heavy standard is of course useful outright, in that tissue analytes that straddle fraction boundaries that have heavy cognates can use them for quantification. Although the complexity of an additional proteome in its entirety to a target tissue digest could detract from the identification of tissue-specific peptide and protein identifications, we note here that mass spectrometer capabilities in sensitivity, dynamic range and sequencing speed continue to improve, suggesting that this issue may be resolved by other means.

Unsurprisingly, the number of orphan species increases as the mixing of sample and standard deviates from 1:1, but, perhaps unintuitively, we find that erring on the side of more sample versus more standard is better for absolute number of eventual quantifications. Notably, using the surrogate method, the amount of heavy standard does not need to be as precisely managed or even known between samples allowing for prospective analyses. Furthermore, we find that a generalized and minimal heavy standard, mouse fibroblast cells (3T3), is sufficient to allow quantification of peptide species abundance across biological samples when the surrogate strategy is employed. Using a single heavy cell line for all spike-in standard experiments reduces the complexity of the peptide mixtures sent to the mass spectrometer and improves the proportion of analysis time spent on biologically interesting peptide species relative to standards. Additionally, in contrast to the use of small sets of peptide standards, surrogate analysis leverages peptide species that are naturally distributed across SCX fractions. By using the median of heavy abundance relationships as a correction factor, the ultimate peptide ratios are less susceptible to errant measurements or increased deviation near the signal to noise. Taken together, the surrogate strategy represents a robust and sensitive supplement to existing spike-in SILAC workflows which taps previously inaccessible proteomic quantifications.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to thank the Israel Lab for access to mouse tissues and A.N. Kettenbach for assistance with mouse tissue harvesting, and acknowledge funding from the National Institutes of Health, National Cancer Institute (R01-CA155260) (to S.A.G.).

REFERENCES

1. Boersema PJ, Raijmakers R, Lemeer S, Mohammed S, Heck AJ. Nat Protoc. 2009; 4:484–494. [PubMed: 19300442]
2. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, Hochstrasser DF, Burkhard PR, Sanchez JC. Analytical Chemistry. 2008; 80:2921–2931. [PubMed: 18312001]
3. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, Purkayastha S, Juhasz P, Martin S, Bartlett-Jones M, He F, Jacobson A, Pappin DJ. Mol Cell Proteomics. 2004; 3:1154–1169. [PubMed: 15385600]
4. Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M. Nat Methods. 2010; 7:383–385. [PubMed: 20364148]
5. Altelaar AF, Munoz J, Heck AJ. Nat Rev Genet. 2013; 14:35–48. [PubMed: 23207911]
6. Villen J, Gygi SP. Nat Protoc. 2008; 3:1630–1638. [PubMed: 18833199]
7. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M. Mol Cell Proteomics. 2002; 1:376–386. [PubMed: 12118079]

8. Park SK, Liao L, Kim JY, Yates JR 3rd. *Nat Methods*. 2009; 6:184–185. [PubMed: 19247291]
9. Van Hoof D, Pinkse MW, Oostwaard DW, Mummery CL, Heck AJ, Krijgsveld J. *Nat Methods*. 2007; 4:677–678. [PubMed: 17762871]
10. Altelaar AF, Frese CK, Preisinger C, Hennrich ML, Schram AW, Timmers HT, Heck AJ, Mohammed S. *J Proteomics*. 2013
11. Liao L, McClatchy DB, Park SK, Xu T, Lu B, Yates JR 3rd. *J Proteome Res*. 2008; 7:4743–4755. [PubMed: 18823140]
12. Kruger M, Moser M, Ussar S, Thievensen I, Lubner CA, Forner F, Schmidt S, Zanivan S, Fassler R, Mann M. *Cell*. 2008; 134:353–364. [PubMed: 18662549]
13. Liao L, Sando RC, Farnum JB, Vanderklish PW, Maximov A, Yates JR. *J Proteome Res*. 2012; 11:1341–1353. [PubMed: 22070516]
14. Eng JK, McCormack AL, Yates JR Iii. *Journal of the American Society for Mass Spectrometry*. 1994; 5:976–989. [PubMed: 24226387]
15. Faherty BK, Gerber SA. *Anal Chem*. 2010; 82:6821–6829. [PubMed: 20684545]
16. Milloy JA, Faherty BK, Gerber SA. *J Proteome Res*. 2012; 11:3581–3591. [PubMed: 22640374]
17. Hsieh EJ, Hoopmann MR, Maclean B, Maccoss MJ. *J Proteome Res*. 2010; 9:1138–1143. [PubMed: 19938873]
18. Valot B, Langella O, Nano E, Zivy M. *Proteomics*. 2011; 11:3572–3577. [PubMed: 21751374]
19. Hill JA, Smith BE, Papoulias PG, Andrews PC. *J Proteome Res*. 2010; 9:2809–2811. [PubMed: 20356086]
20. Geiger T, Wehner A, Schaab C, Cox J, Mann M. *Mol Cell Proteomics*. 2012; 11 M111 014050.
21. Bondarenko PV, Chelius D, Shaler TA. *Anal Chem*. 2002; 74:4741–4749. [PubMed: 12349978]

Figure 1.

Orphan analytes in spike-in SILAC experiments. **(a)** In a standard super-SILAC workflow, proteolytically digested cell lysates are mixed with heavy standards and analyzed by LC-MS/MS. The abundance ratio of an unlabeled peptide to its heavy cognate is compared across samples to quantify peptide abundance over multiple conditions and/or biological replicates. **(b)** Peptides from unlabeled tissue samples with no detected heavy cognate can occur due to poor protein profile complementation or low abundance for specific proteins within the heavy standard, which we term “orphan analytes”. **(c)** Orphan frequency is less than 10% when many cell-type matched heavy cell lines are combined to form the heavy standard. However, such a rich standard pool is not always possible and orphan frequency increases substantially as the number of included cell lines in the standard decreases. **(d)** Orphan frequency for a model spike-in SILAC experiment with murine liver tissue and a single isotopically labeled murine hepatocyte cell line (TIB-75).

Figure 2.

Experimental design and evaluation of orphan frequency by tissue:heavy cell line (L:H) dilution. **(a)** Schematic diagram of murine tissue analysis. Light and heavy peptides were mixed at three dilutions and up to four technical replicates. Each peptide mixture was separated into 24 fractions by strong cation exchange (SCX), and six contiguous fractions were selected in the +2 solution charge state SCX region for analysis by LC-MS/MS. Peptide abundance was quantified by MassChroQ. **(b)** Although the total number of unique unlabeled peptides quantified was similar for the 1:1 and 4:1 dilutions, we observed fewer successful SILAC ratios as the L:H mixing ratio deviated from 1:1. Similarly, while there were fewer orphans in the 1:4 dilution samples, we also observed a sizable decrease in the total number of light peptides quantified.

Figure 3.

Surrogate-based correction factor. By definition, direct comparison to internal standards is not possible for orphan peptides. Instead, surrogates (heavy peptides) are selected which are reliably observed across all replicate analyses. By comparing the abundances of these surrogates across samples, we calculate a correction factor which scales the direct comparison of orphan peptide abundances. Surrogates were included irrespective of matched light sequences, and the correction factor is set to the median of the surrogate abundance ratios so that it is robust to outliers.

a liver + TIB75 (1:1 vs 4:1) **b** liver + TIB75 (1:1 vs 1:1) **Figure 4.**

Surrogate analysis of murine liver tissue. **(a)** Liver analysis with TIB-75 heavy standard. The first panel shows the distribution of spike-in SILAC ratios ($\log_2$ scale) and the total number of unique peptides quantified in fractions 13–16. The upper tail of the spike-in SILAC ratios, indicated by the arrow, shows a bias of spike-in SILAC ratios near unity. The second panel shows the surrogate-based quantification of these same peptides, calculated by ignoring their observed heavy cognates. The right skew is no longer present when the sample-wide heavy relationships are used in place of a single heavy observation. The third panel depicts results from the surrogate analysis on the set of orphan analytes alone, and the fourth panel combines all surrogate analysis quantifications for orphans and non-orphans. **(b)** Comparison of two replicates of the 1:1 dilution yielded similar total numbers of quantifiable peptides suggesting that our method is robust to a four-fold dilution. All ratios of ratios are displayed on a $\log_2$ scale.

Figure 5.

Surrogate analysis of multiple murine tissues with a common heavy mouse fibroblast cell line. We generalized our method by performing SILAC spike-in experiments on mouse **(a)** liver, **(b)** spleen, **(c)** brain and **(d)** muscle tissue using heavy 3T3 cells as our standard. In all cases the super-SILAC and surrogate derived ratios were as expected for the comparison of 1:1 to 4:1 (L:H) and the surrogate method increased the peptide and protein coverage. Here again, the upper tail of the spike-in SILAC ratios is mitigated by the surrogate method. All ratios of ratios are displayed on a $\log_2$ scale.