Research Article   **Open Access**

Microwave and Magnetic (M²) Proteomics of Childhood B-ALL

Mahesula S[1,2,7], Zhang X[5,6,7], Raghunathan R[1,2,7], Kotagiri V[1,2,7], Ruiz JAT[12], Horton T[11], Gelfond JAL[9,10] and WE Haskins[*1-8,10]

[1]Pediatric Biochemistry Laboratory, University of Texas at San Antonio, San Antonio, TX, USA
[2]Department of Biology, University of Texas at San Antonio, San Antonio, TX, USA
[3]Department of Chemistry, University of Texas at San Antonio, San Antonio, TX, USA
[4]RCMI Proteomics, University of Texas at San Antonio, San Antonio, TX, USA
[5]RCMI Protein Biomarkers Cores, University of Texas at San Antonio, San Antonio, TX, USA
[6]Center for Interdisciplinary Health Research, University of Texas at San Antonio, San Antonio, TX, USA
[7]Center for Research & Training in the Sciences, University of Texas at San Antonio, San Antonio, TX, USA
[8]Department of Medicine, Division of Hematology & Medical Oncology, University of Texas at San Antonio, San Antonio, TX, USA
[9]Department of Epidemiology & Biostatistics, University of Texas at San Antonio, San Antonio, TX, USA
[10]Cancer Therapy & Research Center, University of Texas at San Antonio, San Antonio, TX, USA
[11]Texas Children's Cancer and Hematology Centers, Baylor College of Medicine, Houston, TX, USA
[12]Ponce School of Medicine, Puerto Rico

[*]Corresponding author: Mahesula S and Haskins WE, Ph.D., Department of Chemistry, The University of Texas at San Antonio, One UTSA Circle, San Antonio, Texas, USA, Fax: (210)458-5658, Tel: (210)563-4492, Email: mahesulaswetha@gmail.com , WEH.scholar@gmail.com

Abstract

Background: We hypothesized that quantitative tandem mass spectrometry-based proteomics, incorporating rapid microwave and magnetic sample preparation (M² proteomics), might enable relative protein expression to be correlated to childhood B-precursor acute lymphoblastic leukemia (B-ALL) cytogenetic subtypes, corresponding to low-risk (ETV6-RUNX1) and high-risk (MLL-R) subtypes.

Procedure: To test our hypothesis, microwave-assisted reduction/alkylation/digestion of proteins from lymphoblast protein lysates bound to C8 magnetic beads and microwave-assisted isobaric tandem mass tag (TMT) labeling of released peptides were performed for 7 cryopreserved lymphoblast specimens from newly diagnosed pediatric B-ALL patients and pooled reference material from all 7 specimens. These samples were analyzed by M² proteomics in a blinded fashion.

Results: Decoding relative protein expression for each specimen (e.g., LGALS1, IGF2BP1, LMNB1 and PCNA), from 2,132 ± 41 top-ranked TMT-labeled peptides, revealed a strong statistical correlation to cytogenetic subtype. Indeed, hierarchical clustering using Euclidean normalization accurately distinguished 7/7 specimens by cytogenetic subtype. Results were confirmed by M² proteomics with isotopic super stable isotopic labeling of amino acids in culture (SILAC) labeling and by enrichment of differentially expressed proteins in key leukemia signaling pathways.

Conclusions: While studies in larger cohorts merit further investigation, this research supports the proof-of-concept that M² proteomics, incorporating orthogonal labeling strategies such as TMT and super SILAC, is a rapid method to quantify putative prognostic/predictive protein biomarkers and therapeutic targets of childhood B-ALL.

Keywords: Microwave; Magnetic; Proteomics; Leukemia; Proteins; Childhood

Introduction

Childhood leukemia, with relatively well-characterized patient populations, treatments and specimen collection protocols, is an ideal test case for quantitative-tandem mass spectrometry (MS/MS)-based proteomics. Acute lymphoblastic leukemia (ALL), despite an ~85% cure rate, remains the second most common cause of cancer-related mortality in children in the U.S. [1]. Subtle changes in protein expression, such as those reflecting the heterogeneity of dynamic socioeconomic and environmental factors (e.g., time, ethnicity, age, sex, nutrition, healthcare and exposure to drugs, chemicals, infectious diseases or radiation), cannot be determined based solely on analysis of the relatively static genome. Indeed, a well-recognized roadblock to clinical/translational research is a lack of molecular epidemiology data from which to select novel protein biomarkers and therapeutic targets. In humans, approximately 24,000 genes are translated into an estimated 2 million protein isoforms. Paradoxically, there are less than 100 pro-

tein biomarkers that are routinely measured in blood today [2,3]. There is also a poor correlation between the transcriptome and proteome due to alternative splicing, post-translational modifications, single nucleotide polymorphisms, limiting ribosomes available for translation, mRNA and protein stability and other factors affecting translation efficiency (e.g., microRNA) [4].

Masking of low abundance proteins by high abundance proteins in heterogeneous clinical specimens is a significant challenge that limits the dynamic range of proteomics, the large-scale study of protein expression. This is particularly true for studies seeking to identify novel protein biomarkers and therapeutic targets in blood, where proteins from hundreds of different cell types can be present in a relative abundance that may span up to twelve orders of magnitude. Masking problems can be partially overcome by numerous sample preparation strategies, including affinity depletion of high abundant proteins, enrichment of low abundance proteins, and fractionation of proteins by either two-dimensional electrophoresis [5-7] or liquid chromatography [8]. While improving selectivity, these strategies suffer from poor sample throughput and reproducibility due to lengthy and complex sample preparation times and poor dynamic range due to adsorptive losses at various steps. Quantitative MS/MS-based proteomics, in which peptides and proteins are assigned amino acid sequences by searching high-quality product ion spectra from protease-specific peptide precursor ions against spectra predicted from protein databases, place additional constraints on time and dynamic range. Consequently, previous studies have focused on qualitative or semi-quantitative methods for identifying large numbers of proteins in relatively small numbers of specimens. Nonetheless, numerous differentially regulated proteins have been reported in *adult* leukemia specimens and lymphoblast cell lines.

In contrast, there are only a few reports of proteomic analysis of childhood ALL [6,7,9-14]. Significantly, few previous studies have employed isobaric labeling techniques such as 'tandem mass tagging' (TMT) [15,16] or 'super' stable isotopic labeling of amino acids in culture (SILAC) [17,18] for quantitative MS/MS-based proteomics. In TMT, peptides from individual sample protein digests are modified with one of six (ten) isobaric labeling reagents prior to combining the labeled peptides into a single sample mixture. Quantification of TMT-labeled peptides, and by extrapolation their corresponding proteins, is achieved by computing the relative intensities of product (reporter) ions (e.g., TMT 126/TMT131) in MS/MS spectra. Super SILAC metabolically labels proteins from multiple cell lines with lysine and/or arginine containing "heavy" ^{13}C and ^{15}N isotopes (vs. light ^{12}C and ^{14}N) to form a 'heavy' internal standard that can be mixed with 'light' (unlabeled) clinical specimens. Quantification of super SILAC-labeled peptides, and by extrapolation their corresponding proteins, is achieved by computing the relative intensities of precursor ions (e.g., light/heavy) in MS spectra.

Our laboratory has pioneered quantitative MS/MS-based "Microwave & Magnetic (M^2) Proteomics" for rapid (90 second) and high-throughput (96-well) gel-free sample preparation with isobaric/isotopic-labeling for simultaneous sequence-specific quantification of thousands of proteins or selected low abundance proteins in large numbers of specimens [19,20]. Therefore, we hypothesized that quantitative tandem mass spectrometry-based proteomics, incorporating rapid microwave and magnetic sample preparation (M^2 proteomics), might enable relative protein expression to be correlated to childhood B-precursor acute lymphoblastic leukemia (B-ALL) cytogenetic subtypes, corresponding to low-risk (ETV6-RUNX1) and high-risk (MLL-R) subtypes. To test our hypothesis, microwave-assisted reduction/alkylation/digestion of proteins from lymphoblast protein lysates bound to C8 magnetic beads and microwave-assisted isobaric tandem mass tag (TMT) labeling of released peptides were performed for 7 cryopreserved lymphoblast specimens from newly diagnosed pediatric B-ALL patients and pooled reference material from all 7 specimens. These samples were analyzed by M^2 proteomics in a blinded fashion. Results were evaluated by M^2 proteomics with isotopic super SILAC labeling, in which protein lysates from 9 unlabeled 'light' specimens were individually mixed with 3 combinations of pooled reference materials from 4 'heavy'-labeled B-ALL cell lines representing distinct cytogenetic B-ALL subtypes. Results were also evaluated by mapping the expression of differentially expressed proteins, corresponding to top-ranked TMT-labeled peptides, to key leukemia signaling pathways.

Material and Methods

B-ALL Clinical Specimens

Nine de-identified cryopreserved pretreatment lymphoblast specimens from Texas Children's Cancer and Hematology Centers were obtained from newly diagnosed B-ALL patients. All specimens had either the ETV6-RUNX1 or mixed lineage leukemia-rearranged (MLL-R) fusion proteins. Bone marrow and peripheral blood specimens were analyzed by M^2 proteomics in a blinded fashion as described below. After assigning each specimen to one of two cytogenetic subtypes (I or II) by M^2 proteomics, clinical data for each specimen, including cytogenetic subtype, was provided to us (Table 1). Since we were blinded to the fact that two specimens were from the same patient in two cases, all specimens were treated as originating from separate patients. Due to the limited amount of protein extracted from some specimens, not all specimens were included in all analyses.

B-ALL Cell Lines

Commercially characterized childhood B-ALL cell lines were employed for these studies within six months of their receipt at UTSA, including: REH with the ETV6-RUNX1 fusion protein [ATCC RS4;11 with the MLL-AF4 fusion protein [t(4;11)(q21;q23) translocation] (DSMZ, Braunschweig, Germany); GM04154 (Corielle Cell Repositories, New Jersey); and cytokine-receptor-like factor 2-overexpressing (CRLF2+)-MHH-CALL4 [45,X,-Y,r(12)(p13q24.33),der(14)(t(Y;14)(p11;q32) translocation] (DSMZ, Braunschweig, Germany) [21-26]. *For isobaric TMT labeling,* cells were grown in RPMI1640 medium (ATCC) with 20% FBS (GibCo)/5% Penicillin-Streptomycin (MPBiomedicals) in 75 cm² T-flasks at 37 ºC, 5psi CO_2. These cells were used for method d-

evelopment (data not shown) and not as reference materials. *For isotopic super SILAC labeling,* cells were cultured in RPMI 1640 containing 10% dialyzed FBS containing 0.46mM [U-13C6] L-lysine (heavy media) (Invitrogen, Grand Island, NY). Cell lines were allowed to grow for approximately 7 cycles of DNA replication to achieve maximum incorporation of heavy-labeled amino acids into proteins. Cells were harvested by centrifugation at 1000 rpm for 5 min and washed twice with 1X PBS. These cells were centrifuged, washed, and made into lysate using RIPA Lysis Buffer kit (Santa Cruz Biotechnology, Santa Cruz CA) per manufacturer's protocol. Whole cell protein lysates were used as reference material (described below).

Specimen ID	Cytogenetics	Risk Group	Age	Gender	Ethnicity	Initial White Blood Cell Count (x10³)
3036	ETV6-RUNX1	I	4 yrs	F	Hispanic/Latino	35
3730	ETV6-RUNX1	I	10 yrs	F	Hispanic/Latino	43
BM1951 or 1953	ETV6-RUNX1	I	3 yrs	M	Hispanic/Latino	7
BM3957	ETV6-RUNX1	I	8 yrs	F	Hispanic/Latino	28
1398 or PB5058	MLL-R	II	6 mos	M	Non-Hispanic/Latino White	602
P186	MLL-R	II	4 mos	M	Hispanic/Latino	965
P214	MLL-R	II	9 yrs	F	Vietnamese	1132

Table 1: Clinical data for B-ALL specimens.

Reference Materials

For isobaric TMT labeling, protein was pooled from all 7 specimens by protein amount as reference material prior to sample preparation as described below. *For isotopic super SILAC labeling,* protein lysates from 9 unlabeled 'light' specimens were individually mixed with 3 combinations of pooled reference materials from 4 'heavy'-labeled B-ALL cell lines (described above) representing distinct cytogenetic B-ALL subtypes. Combination A contained REH + RS4;11, combination B contained GM04154 + MHH-CALL4, and combination C contained all 4 cell lines.

Microwave and Magnetic (M²) Sample Preparation

For isobaric TMT labeling, 50 mg of C8 magnetic beads (BcMg, Bioclone Inc.) were suspended in 1 mL of 50% methanol. Immediately before use, 100 μL of the beads were washed 3 times with equilibration buffer (200 mM NaCl, 0.1% trifluoroacetic acid (TFA)). Whole cell protein lysate (25-100 μg at 1μg/μL) was mixed with pre-equilibrated beads and 1/3rd sample binding buffer (800 mM NaCl, 0.4% TFA) by volume. The mixture was incubated at room temperature for 5 min followed by removing the supernatant. The beads were washed twice with 150 μL of 40 mM triethylammonium bicarbonate (TEAB), and then 150 μL of 10 mM dithiolthreitol (DTT) was added. The bead-lysate mixture underwent microwave heating for 10s. DTT was removed and 150 μL of 50 mM iodoacetamide (IAA) added, followed by a second microwave heating for 10s. The beads were washed twice and re-suspended in 150 μL of 40 mM TEAB. In vitro proteolysis was performed with 4 μL of trypsin in a 1:25 trypsin-to-protein ratio (stock = 1μg/μL in 50mM acetic acid) and microwave heated for 20 s in triplicate. The supernatant was used immediately or stored at -80 °C. Released tryptic peptides from digested protein lysates, including the reference materials described above, were modified at the N-terminus and at lysine residues with the tandem mass tagging (TMT)-6plex isobaric labeling reagents (Thermo scientific, San Jose, CA). Each individual specimen was encoded with one of the TMT-126-130 reagents, while reference material was encoded with the TMT-131 reagent: 41 μL of anhydrous acetonitrile was added to 0.8 mg of TMT labeling reagent for 25μg of protein lysate and microwave-heated for 10s. To quench the reaction, 8 μL of 5% hydroxylamine was added to the sample at room temperature. To normalize across all specimens, TMT-encoded cell lysates from individual specimens, labeled with the TMT-126-130 reagents, were mixed with the reference material encoded with the TMT-131 reagent in $1_{126}:1_{127}:1_{128}:1_{129}:1_{130}:1_{131}$ ratios. Since there were more than five specimens, two sample mixtures (a and b) were prepared with randomly selected specimens for each mixture and the same reference material. *For super SILAC labeling*, the same steps were performed without the TMT labeling and pooling steps. These sample mixtures, including all TMT- and super SILAC-encoded specimens, were stored at -80 °C until further use.

Peptide Fractionation

C18 ziptips (Millipore, Billerica, MA) were employed to fractionate peptides according to the manufacturer's protocol, where 5 fractions were generated by eluting with 10, 25, 35, 50 and 70% acetonitrile/0.1% formic acid prior to speed vacuuming to dryness and the addition of 20 μL of 0.1% formic acid.

Capillary Liquid Chromatography-Fourier-Transform-Tandem Mass Spectrometry (LC/FT/MS/MS) with Protein Database Searching

Capillary LC/FT/MS/MS was performed with a splitless nano LC-2D pump (Eksigent, Livermore, CA), a 50 μm-i.d. column packed with 7 cm of 3 μm-o.d. C18 particles, and a hybrid linear ion trap-Fourier-transform tandem mass spectrometer (LTQ-ELITE; ThermoFisher, San Jose, CA) operated with a lock mass for calibration. The reverse-phase gradient was 2 to 62% of 0.1% formic acid (FA) in acetonitrile over 60 min at 350 nL/min. *For unbiased analyses,* the top 6 most abundant eluting ions were fragmented by data-dependent HCD with a mass resolution of 120,000 for MS and 15,000 for MS/MS. *For targeted analyses,* only ions corresponding to peptides observed for selected proteins were fragmented by HCD (data not shown). *For isobaric TMT* labeling, probability-based protein database searching of MS/MS spectra against the IPI_human protein database (release 2010_jan10; 87,061 sequences) was performed with a 10-node MASCOT cluster (v. 2.3.02, Matrix Science, London, UK) with the follo-

wing search criteria: peak picking with Mascot Distiller; 10 ppm precursor ion mass tolerance, 0.8 Da product ion mass tolerance, 3 missed cleavages, trypsin, carbamidomethyl cysteines as a static modification, oxidized methionines and deamidated asparagines as variable modifications, an ion score threshold of 20 and TMT-6-plex for quantification. *For isotopic super SILAC labeling*, probability-based and protein database searching of MS/MS spectra against the IPI_human protein database (release 2010_jan10; 87061 sequences) were performed with a MaxQuant (vr. 1.2.2.5) with 12 CPU threads with the following search criteria: peak matching between runs with 2 min windows; 10 ppm precursor ion mass tolerance, 0.02 Da product ion mass tolerance, 3 missed cleavages, trypsin, carbamidomethyl cysteines as a static modification, oxidized methionines and deamidated asparagines as variable modifications, false-discovery ratios of 1.0 and SILAC K+6 quantification.

Signaling Pathway Analysis

Mapping the expression of proteins, corresponding to top-ranked TMT-labeled peptides, to key leukemia signaling pathways was performed with Ingenuity Pathways Analysis (IPA, Ingenuity® Systems) according to the manufacturer's suggestions. A vertical bar plot, showing the percentage of proteins quantified in each canonical signaling pathway, was visualized to investigate pathway enrichment across all specimens, where *P*-values for enrichment were assigned by IPA.

Statistical Analysis

The M^2 proteomics technical replicates estimated protein expression for individual specimens, TMT- or SILAC-encoded in sample mixtures, relative to pooled reference materials. Relative protein expression levels were transformed to log base 2 for quantile normalization. Outlier arrays were removed based upon the following quality control procedures: 1) overall intensity histograms of normalized expression were compared with kernel smoothed density plots, and 2) non-supervised hierarchical clustering of sample profiles was performed to assess the consistency of technical and biological variation. We tested the association between relative protein expression and cytogenetic subtype using a linear mixed-effect while treating cytogenetic subtype as a continuous predictor. We treated the cytogenetic subtype effect on relative protein expression singly, as a univariate predictor. We tested all the pairwise differences in relative protein expression between all specimens using an unpaired, unequal variance t-test on the replicate averages. We examined the relationship between the overall expression profile for all specimens using a non-supervised hierarchical clustering display based upon Euclidean distance and complete linkage. For clustering analyses of relative protein expression profiles, we considered the subset of proteins that were most variable by selecting the proteins in the top quartile (top 25%) by their standard deviation ranking. All statistical analysis was performed with R v2.13 (R-Project, Vienna, Austria).

Results

Decoding relative protein expression for each specimen, from 2,132 ± 41 top-ranked TMT-labeled peptides, revealed a strong statistical correlation to cytogenetic subtype. Indeed, heirarchical clustering using Euclidean normalization accurately distinguished 7/7 specimens by cytogenetic subtype (Figure 1 and Table 2). For example, the protein Galectin I (**LGALS1**) (IPI00219219) was observed to be differentially expressed, where the TMT-labeled peptide SFVLNLGK of LGALS1 was down-regulated 1.7 fold (P=9.9E-03) for risk group I vs. II. Likewise, the TMT-labeled peptide, FNAHGDANTIVCNSK of LGALS1 was down-regulated 1.4 fold (P=2.8E-04) for risk group I vs. II (Figure 2). Also consider insulin-like growth factor 2 mRNA-binding protein 1 (**IGF2BP1**) (IPI00008557), where the TMT-labeled peptide LLVPTQYVGAIIGK was up-regulated 1.0 fold (P=5.5E-05) for risk group I vs. II. Another excellent example is lamin B1 (**LMNB1**) (IPI00217975), where the TMT-labeled peptide KIGDTSVSYK was down-regulated 1.2 fold (P=9.1E-05) for risk group I vs. II. Lastly, consider proliferating cell nuclear antigen (**PCNA**) (IPI00021700), where the TMT-labeled peptide LMDLDVEQLGIPEQEYSCVVK was up-regulated 0.6 fold (P=8.1E-02) for risk group I vs. II (Supplementary Figure 1). Pilot M^2 proteomics experiments with protein standards spanning a dynamic range of 1:250, performed in triplicate, confirmed quantitative isobaric labeling of N-termini and lysine residues and technical reproducibility (data not shown).

Results from M^2 proteomics with TMT labeling were evaluated by M^2 proteomics with super SILAC labeling (Figure 1), where "heavy" super SILAC-labeled reference material was comprised of various combinations (A-C) of B-ALL cell lines with ETV6-RUNX1, MLL-R, CLRF2+, and other cytogenetic subtypes. Decoding relative protein expression for individual specimens revealed 1230 ± 110 top-ranked super SILAC-labeled proteins (see Supplementary Material). Shown in Table 2 is the cytogenetic subtype assignment of risk group I vs. II with M^2 proteomics for each specimen and labeling strategy. M^2 proteomics with super SILAC labeling accurately distinguished 7/8, 5/8, and 5/9 specimens, respectively. Results from M^2 proteomics were also evaluated by mapping the expression of differentially expressed proteins, corresponding to top-ranked TMT-labeled peptides, to key leukemia signaling pathways (e.g., B-cell receptor signaling, mTOR signaling, granzyme signaling, NFκB signaling, etc.). As shown in Figure 3, the "molecular mechanisms of cancer" pathway represents one of many enriched pathways from our preliminary studies. More than 121 of the 377 (32.1%) known proteins in this pathway were differentially expressed in our dataset (P =4.1E-20).

Discussion

Relatively few pediatric proteomic studies have been performed using patient specimens with ETV6-RUNX1, MLL-R, or CRLF2+ cytogenetic subtypes of childhood B-ALL. Previous studies have included measurements using genomic, transcriptomic [27-30] and proteomic [9-14] approaches. This study utilizes a new approach, M^2 proteomics, to examine differences in protein expression between cytogenetic subtypes of childhood B-ALL. Below, we provide a brief discussion of selected differentially expressed proteins revealed by M^2 proteomics and comparisons to previous studies and other proteomics approaches.
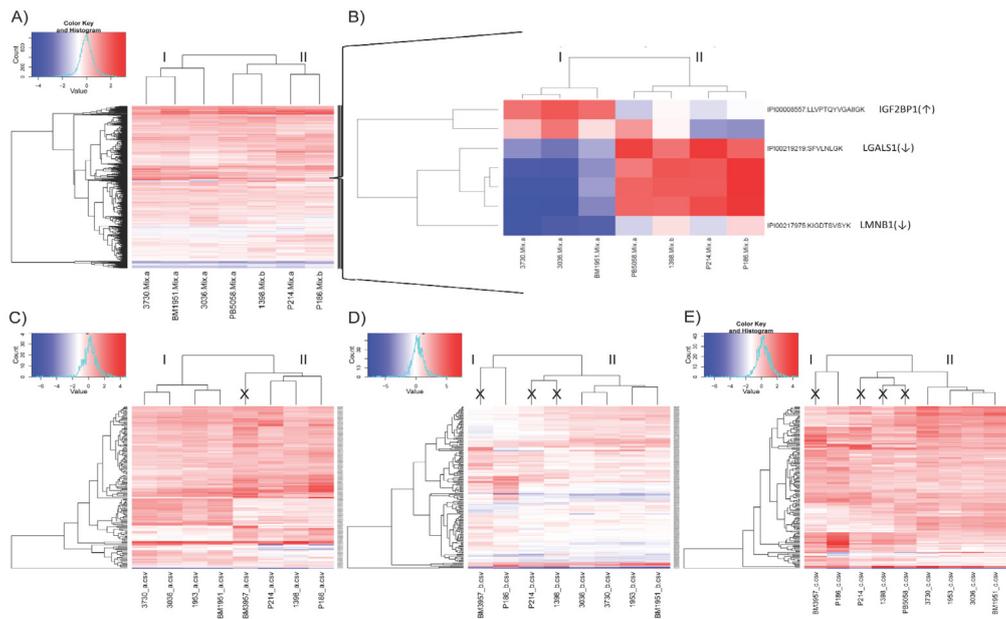
Figure 1: M^2 proteomics accurately distinguished cytogenetic subtypes of pediatric B-ALL. **A)** Isobaric TMT labeling and hierarchical clustering of M^2 proteomics signatures using Euclidean normalization accurately distinguished 7/7 specimens by cytogenetic subtype [ETV6-RUNX1 (I) or MLL-R (II)] with 2132 ± 41 top-ranked TMT-labeled peptides and reference material pooled from all 7 specimens. **B)** A zoomed-in region that contains selected TMT-labeled peptides of interest [from IGF2BP1(↑)1.0-fold (P=5.5E-05), LGALS1(↓) 1.7-fold (P=9.9E-03), and LMNB1(↓)1.2-fold (P=9.1E-05)] is also shown. **C-E)** Super SILAC labeling with "light"-unlabeled patient specimens and "heavy" SILAC-labeled reference materials comprised of various combinations of blast cell lines (A-C, left to right) supported the TMT labeling results. Super SILAC distinguished 7/8, 5/8, and 5/9 specimens, respectively, where (X)s are shown for misclassified specimens. Misclassifications, including nearest neighbors to the correct classification, are most simply explained by masking of key low abundance proteins. Super SILAC combination A (MLL-AF4 + ETV6-RUNX1) most accurately distinguished the two cytogenetic subtypes (risk groups) of patient specimens (MLL-R or ETV6-RUNX1), compared to the other super SILAC combinations (B = GM04154 + MHH-CALL4 and C = all 4 B-ALL cell lines), as expected. In other words, the super SILAC combination that most closely matches the TMT labeling results was also the most representative of the patient specimens.

M^2 Proteomics Labeling Strategy	Risk Group II	Risk Group I	No. of Specimens Accurately Distinguished
	(MLL-R)	(ETV6-RUNX1)	
TMT	3036	1398	7/7
	3730	P186	
	BM1951	P214	
		PB5058	
Super SILAC Combination A (RS4;11 + REH)	1953	1398	7/8
	3036	*BM3957*	
	3730	P186	
	BM1951	P214	
Super SILAC Combination B (MHH-CALL4 + GM04154)	*1398 (nearest neighbor to group II)*	*BM3957*	5/8
	1953	P186	
	3036		
	3730		
	BM1951		
	P214 (nearest neighbor to group II)		
Super SILAC Combination C (RS4;11 + REH+ MHH-CALL4 + GM04154)	*1398 (nearest neighbor to group II)*	*BM3957*	5/9
	1953	P186	
	3036		
	3730		
	BM1951		
	P214 (nearest neighbor to group II)		
	PB5058 (nearest neighbor to group II)		

Table 2: Cytogenetic subtype assignment (group I or II) by M^2 proteomics for each specimen and labeling strategy with different combinations of reference materials. Heirarchical clustering using Euclidean normalization clearly distinguishes at least two risk groups (I and II) that correspond to the ETV6-RUNX1 (I) and MLL-R (II) cytogenetic subtypes of childhood B-ALL. Misclassified specimens (Xs) are shown in italics (albeit all of these were nearest neighbors to the correct classification as described in Figure 1).
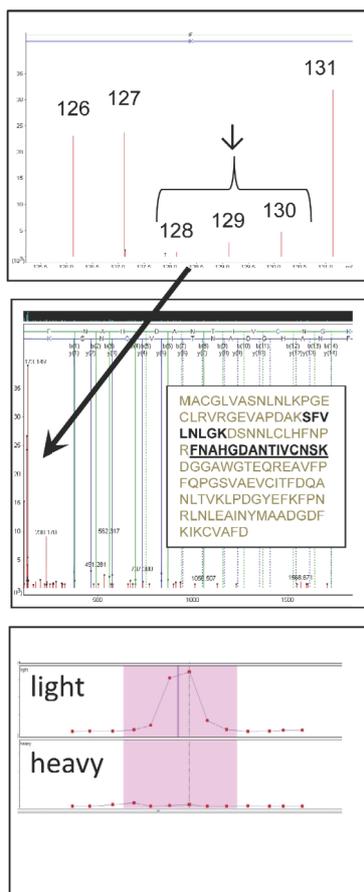
Figure 2: M^2 proteomics results for the top-ranked tryptic peptide FNAHGDANTIVCNSK, spanning amino acid residues 50-64 of LGALS1, quantified by both TMT- and super SILAC-labeling. Top) TMT reporter ions encoding 5 clinical specimens (TMT1$_{126-130}$) and pooled reference material (TMT$_{131}$) are shown. This TMT-labeled peptide was down-regulated (↓) by 1.4 fold (P=2.8E-04) in risk group I (ETV6-RUNX1) vs. II (MLL-R). Specimens PB5058 and P214 from risk group II were encoded with the TMT$_{126}$ and TMT$_{127}$ reagents, while specimens BM1951, 3036, and 3730 from risk group I were encoded with the TMT$_{128-130}$ reagents in this sample mixture. Middle) The amino acid sequence assignment of product ions for a representative MS/MS spectrum. Other TMT- and SILAC-labeled peptides that were observed are shown in bold. Bottom) The extracted ion chromatograms for co-eluting SILAC-labeled precursor ions encoding clinical specimens P214 ('light') and super SILAC reference material combination A ('heavy') are shown for comparison.
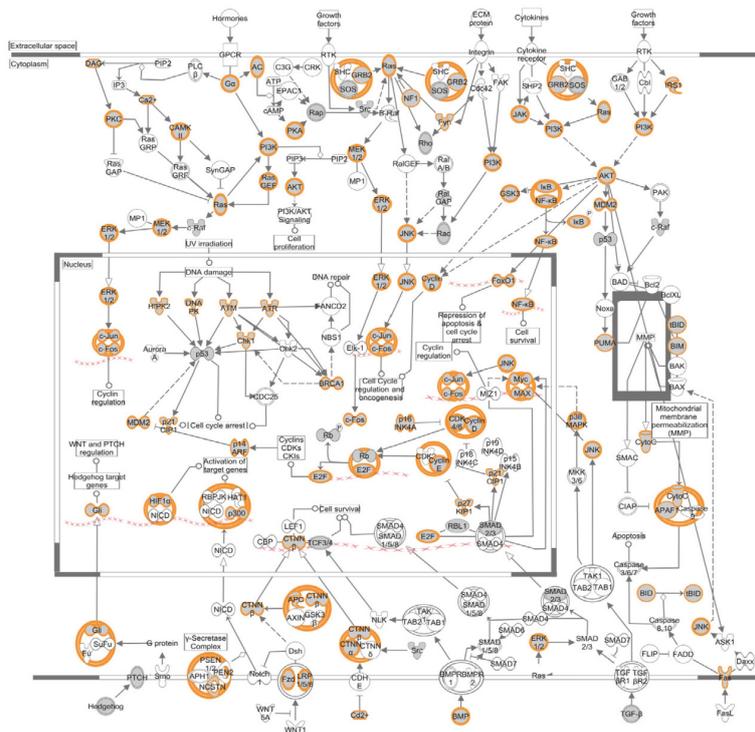
Figure 3: The "molecular mechanisms of cancer" pathway represents one of many enriched pathways from our studies where a total of 2,799 identified protein isoforms were mapped. More than 121 of the 377 (32.1%) known proteins in this pathway were observed in our dataset (P=4.1E-20).

LGALS1 was down-regulated in risk group I vs. II. LGALS1 is a carbohydrate-binding protein defined by its affinity for β-galactosides [31]. LGALS1 has been shown to be abundantly expressed in IgM (+) memory B cells, inhibiting Akt phosphorylation and up-regulating the pro-apoptotic BH3-only protein Bim during B-cell receptor signaling [32]. Down-regulation of LGALS1 has been associated with hyper-methylation of its promoter [33] and cytarabine resistance [34]. Previous work on childhood B-ALL showed that LGALS1 was up-regulated in MLL-R vs. other cytogenetic subtypes [35], in agreement with our observations.

IGF2BP1 was up-regulated in risk group I vs. II. Insulin-like growth factor 2 mRNA-binding proteins are involved in tissue- and cell-specific localization, stability, and translational control of IGF2 by binding to the 5' untranslated region of IGF2 mRNA [36-38]. IGF2BP1 might also play a pivotal role in mediating the mitogenic response of leukemia cells, as suggested by breast cancer studies [39,40]. Lastly, gene expression analysis showed that IGF2BP1 was up-regulated in risk group I vs. II, in excellent agreement with our observations [36].

LMNB1 was down-regulated in risk group I vs. II. LMNB1 belongs to the family of lamins, proteins that reside within the nuclear lamina and nucleoplasm. Lamins play a major role in DNA replication and repair, transcription and nuclear restructuring [41]. Previous studies have shown that LMNB1 plays a particular role in cell proliferation and senescence through p53 and reactive-oxygen species signaling pathways [42]. Lamins might also directly bind to PCNA [43]. Lastly, proteomic analyses have revealed LMNB1 as a putative protein biomarker for hepatocellular carcinoma [44] and down-regulation of LMNB1 in rapamycin-treated T-ALL blasts [45].

PCNA was up-regulated in risk group I vs. II. PCNA is a prominent cancer-associated protein involved in DNA replication and repair. In addition to protein amount, differential and combinatorial expression of post-translational modifications (e.g., ubiquitylation) modulate repair of drug-induced DNA damage. This might explain why PCNA is highly expressed in ETV-RUNX1 leukemia, which are considered to be more responsive to chemotherapy than leukemias containing the MLL-R high-risk translocation [46]. Two-dimensional gel electrophoresis based proteomic analysis revealed down-regulation of PCNA in prednisone-treated RS4 cells containing the MLL-AF4 fusion protein [13], and flow cytometry analysis of PCNA in 43 paired patient specimens (pre-treatment vs. 7 days of prednisolone treatment) showed that PCNA down-regulation was a promising biomarker of prednisone response. We investigated pretreatment lymphoblast specimens in this work. Therefore, future M^2 proteomics studies will be needed to determine whether PCNA is up- or down-regulated in prednisone-treated specimens from risk group I vs. II.

In this work, we have shown that M^2 proteomics with isotopic super SILAC labeling is a valuable tool for confirming results from M^2 proteomics with isobaric TMT labeling. M^2 proteomics with super SILAC labeling accurately distinguished 7/8, 5/8, and 5/9 specimens, respectively, for the 3 super SILAC-labeled reference material combinations. As shown in Figure 1 and Table 2, the 'heavy'-super SILAC labeled reference material made by combining protein lysates from B-ALL cell lines with cytogenetic characteristics that are the most representative of the 'light'-unlabeled patient specimens in question (combination A = MLL-AF4 + ETV6-RUNX1) was superior to the two other 'heavy'-super SILAC labeled material combinations (combination B = GM04154 + MHH-CALL4; combination C = all 4 cell lines) for accurately distinguishing cytogenetic subtypes. Three and four misclassifications [(X) s in Figure 1 and italics in Table 2] were observed for reference material combinations B and C, respectively, albeit all of these were nearest neighbors to the correct classification. Misclassifications with super SILAC are most simply explained by masking of key low abundance proteins. Not surprisingly, we observed no misclassifications with TMT labeling, where reference material was pooled from all 7 specimens.

The main advantage of SILAC- (over TMT-) labeling is that encoding is done at the protein (rather than peptide) level to enable the addition of internal standards prior to downstream sample preparation steps (i.e., separation, digestion and labeling) and more accurate relative quantification. For example, SILAC-labeling of B-cells differentiated *in vitro* enabled the downstream separation of proteins with SDS-PAGE prior to MS/MS [47]. More recently, cell lines representing activated B-cell-like or germinal-center B-cell-like subtypes of diffuse large B-cell lymphoma were distinguished by super SILAC [48]. The main disadvantage of SILAC- (vs. TMT-) labeling is that it is a low-throughput and computationally intensive method that does not allow for extensive multiplexing, often resulting in statistically underpowered proteomic studies. Neutron-encoded SILAC-labeling is expected to dramatically improve sample throughput [49] and novel computer algorithms [50,51] are expected to reduce computational burden. However, our results suggest that TMT-labeling studies can be more accurate than super SILAC-labeling studies, depending on the reference materials employed.

The proteins corresponding to TMT-labeled peptides, rather than the TMT-labeled peptides themselves, were compared to SILAC-labeled proteins for validation of results. A different subset of peptides was expected for M^2 proteomics with TMT- vs. super SILAC-labeling. This precludes direct comparison at the peptide and/or protein level in many cases. For example, PCNA, LGALS1, IGF2BP1, and LMNB1 were quantified with TMT labeling but only LGALS1 and IGF2BP1 were quantified with super SILAC labeling (Supplementary Table 1). The protein expression ratios for each specimen relative to reference materials are expected to differ across the different reference materials. However, the direction of differentially expressed proteins (i.e., up- vs. down-regulation) in risk group I vs. II revealed with TMT- and super SILAC-labeling are in excellent agreement with each other. PCNA and LMNB1 were only identified, not quantified, with super SILAC labeling, possibly due to weak peptide signals. Consider specimen P214 and super SILAC reference material combination A (= MLL-AF4 + ETV6-RUNX1). The peptides SFVLNLGK and FNAHGDANTIVCNSK from LGALS1 were quantified by both TMT- and SILAC- labeling. Likewise, peptides LLVPTQYVGAIIGK and MVIITGPPEAQFK from IGF2BP1 were quantified by both TMT- and SILAC- labeling. Examples of peptides that were iden-

tified but not quantified include SILAC-labeled peptides ISYSGQFIVK and QQQVDIPIR and TMT-labeled peptides GQHIKQLSR and TVNELQNLTAAEVVVPR from IGF2BP1. This shows that TMT- and super SILAC-labeling are complementary strategies for examining the pediatric B-ALL proteome.

Conclusion

While studies in larger cohorts merit further investigation, this research supports the proof-of-concept that M^2 proteomics, incorporating orthogonal labeling strategies such as TMT and super SILAC, is a rapid method to quantify putative prognostic/predictive protein biomarkers and therapeutic targets of childhood B-ALL. From these preliminary studies, we calculate that M^2 proteomics can reveal subtler changes in protein expression than those due to differences in cytogenetic subtype, such as those reflecting the heterogeneities of dynamic socioeconomic and environmental factors. Therefore, M^2 proteomics is expected to become increasingly important to accurately predict clinical outcome and improve risk group stratification and therapy for pediatric B-ALL patients.

Acknowledgment

Supplementary Data

References

1. Pui C H, Jeha S (2007) New therapeutic strategies for the treatment of acute lymphoblastic leukaemia. Nat Rev Drug Discov 6: 149-65.

2. Rifai N, Gillette MA, Carr SA (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. Nat Biotechnol 24: 971-83.

3. Anderson NL, Anderson NG (2002) The human plasma proteome: history, character, and diagnostic prospects. Mol Cell Proteomics 1: 845-67.

4. Schotte D, Chau JC, Sylvester G, Liu G, Chen C, et al. (2009) Identification of new microRNA genes and aberrant microRNA profiles in childhood acute lymphoblastic leukemia. Leuk 23: 313-22.

5. Haskins WE, Eedala S, Jadhav YL, Labhan MS, Pericherla VC, et al. (2012) Insights on neoplastic stem cells from gel-based proteomics of childhood germ cell tumors. Pediatr blood cancer 58: 722-8.

6. Wang D, Lv YQ, Liu YF, Du XJ, Li B (2013) Differential protein analysis of lymphocytes between children with acute lymphoblastic leukemia and healthy children. Leuk lymphoma 54: 381-6.

7. Braoudaki M, Lambrou GI, Vougas K, Karamolegou K, Tsangaris GT, et al. (2013) Protein biomarkers distinguish between high- and low-risk pediatric acute lymphoblastic leukemia in a tissue specific manner. J Hematol Oncol 6: 52.

8. Ottens AK, Kobeissy FH, Wolper RA, Haskins WE, Hayes RL, et al. (2005) A multidimensional differential proteomic platform using dual-phase ion-exchange chromatography-polyacrylamide gel electrophoresis/reversed-phase liquid chromatography tandem mass spectrometry. Anal chem 77: 4836-45.

9. Braoudaki M, Stathopoulou TF, Anagnostopoulos AK, Papathanassiou C, Vougas K, et al. (2011) Proteomic analysis of childhood de novo acute myeloid leukemia and myelodysplastic syndrome/AML: correlation to molecular and cytogenetic analyses. Amino Acids 40: 943-51.

10. Verrills NM, Liem NL, Liaw TY, Hood BD, Lock RB, et al. (2006) Proteomic analysis reveals a novel role for the actin cytoskeleton in vincristine resistant childhood leukemia--an in vivo study. Proteomics 6: 1681-94.

11. Hegedus CM, Gunn L, Skibola CF, Zhang L, Shiao R, et al. (2005) Proteomic analysis of childhood leukemia. Leukemia 19: 1713-8.

12. Accordi B, Espina V, Giordan M, VanMeter A, Milani G, et al. (2010) Functional protein network activation mapping reveals new potential molecular drug targets for poor prognosis pediatric BCP-ALL. PLoS One 5: e13552.

13. Jiang N, Kham SK, Koh GS, Suang Lim JY, Ariffin H, et al. (2011) Identification of prognostic protein biomarkers in childhood acute lymphoblastic leukemia (ALL). J Proteomics 74: 843-57.

14. Hunsucker SW, Accurso FJ, Duncan MW (2007) Proteomics in pediatric research and practice. Adv Pediatr 54: 9-28.

15. Thompson A, Schafer J, Kuhn K, Kienle S, Schwarz J, et al. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. Anal Chem 75: 1895-904.

16. Dayon L, Hainard A, Licker V, Turck N, Kuhn K, et al. (2008) Relative quantification of proteins in human cerebrospinal fluids by MS/MS using 6-plex isobaric tags. Anal Chem 80: 2921-31.

17. Geiger T, Cox J, Ostasiewicz P, Wisniewski JR, Mann M (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. Nature methods 7: 383-5.

18. Neubert TA, Tempst P (2010) Super-SILAC for tumors and tissues. Nature methods 7: 361-2.

19. Raphael I, Mahesula S, Kalsaria K, Kotagiri V, Purkar AB, et al. (2012) Microwave and magnetic (M(2)) proteomics of the experimental autoimmune encephalomyelitis animal model of multiple sclerosis. Electrophoresis 33: 3810-9.

20. Mahesula S, Raphael I, Raghunathan R, Kalsaria K, Kotagiri V, et al. (2012) Immunoenrichment microwave and magnetic proteomics for quantifying CD47 in the experimental autoimmune encephalomyelitis model of multiple sclerosis. Electrophoresis 33: 3820-9.

21. Lazarus H, Barell EF, Oppenheim BS, Krishan A (1974) Divergent properties of two human lymphocytic cell lines isolated from a single specimen of peripheral blood. In Vitro 9: 303-10.

22. Matsuo Y, Drexler HG (1998) Establishment and characterization of human B cell precursor-leukemia cell lines. Leuk Res 22: 567-79.

23. Miyoshi I, Hiraki S, Tsubota T, Kubonishi I, Matsuda Y, et al. (1977) Human B cell, T cell and null cell leukaemic cell lines derived from acute lymphoblastic leukaemias. Nature 267: 843-4.