Journal of proteome research

Article

pubs.acs.org/jpr

Quantitative Analysis of Single Amino Acid Variant Peptides Associated with Pancreatic Cancer in Serum by an Isobaric Labeling Quantitative Method

Song Nie,[†] Haidi Yin,[†] Zhijing Tan,[†] Michelle A. Anderson,[‡] Mack T. Ruffin,[§] Diane M. Simeone,[†] and David M. Lubman*[,†]

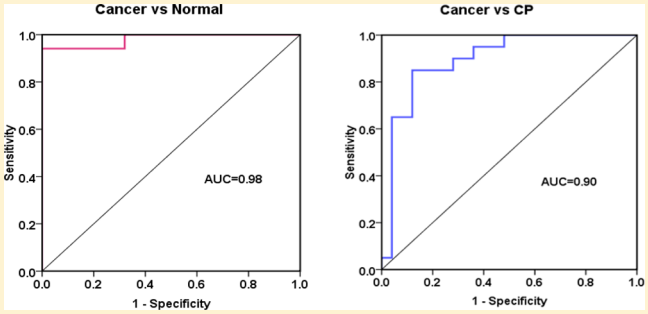
[†]Department of Surgery, University of Michigan, Ann Arbor, Michigan 48109, United States

[‡]Division of Gastroenterology, Department of Internal Medicine, University of Michigan, Ann Arbor, Michigan 48109, United States

[§]Department of Family Medicine, University of Michigan, Ann Arbor, Michigan 48109, United States

Supporting Information

ABSTRACT: Single amino acid variations are highly associated with many human diseases. The direct detection of peptides containing single amino acid variants (SAAVs) derived from nonsynonymous single nucleotide polymorphisms (SNPs) in serum can provide unique opportunities for SAAV associated biomarker discovery. In the present study, an isobaric labeling quantitative strategy was applied to identify and quantify variant peptides in serum samples of pancreatic cancer patients and other benign controls. The largest number of SAAV peptides to date in serum including 96 unique variant peptides were quantified in this quantitative analysis, of which five variant peptides showed a statistically significant difference between pancreatic cancer and other controls (p-value < 0.05). Significant differences in the variant peptide SDNCEDTPEAGYFAVAVVK from serotransferrin were detected between pancreatic cancer and controls, which was further validated by selected reaction monitoring (SRM) analysis. The novel biomarker panel obtained by combining α-1-antichymotrypsin (AACT), Thrombospondin-1 (THBS1) and this variant peptide showed an excellent diagnostic performance in discriminating pancreatic cancer from healthy controls (AUC = 0.98) and chronic pancreatitis (AUC = 0.90). These results suggest that large-scale analysis of SAAV peptides in serum may provide a new direction for biomarker discovery research.



KEYWORDS: *single amino acid variant, serum, isobaric labeling, proteomics, pancreatic cancer, biomarker*

INTRODUCTION

Since the completion of the Human Genome Project in 2003, a large number of genetic variations have been uncovered in humans, including single nucleotide variations, insertions and deletions.[1] Among these genetic variations, single nucleotide variants (SNVs) are recognized as the most common type of genetic variations in the human genome, occurring on average on the order of one per 860 base pairs where these variants are often associated with particular diseases and differential drug response.[2,3] Increasing evidence has shown that a small fraction of these variants are responsible for the initiation or progression of cancer.[4,5] The genomic variations of cancer are expressed through aberrant proteins and their altered functions. Therefore, the direct detection of these variant peptides may provide unique opportunities for biomarker development.

Pancreatic cancer is a dismal disease, with a poor long-term outcome of less than 5% for the 5 year-survival rate.[6−8] However, the possibility of patient's survival is considerably improved when the patient is diagnosed at an early stage of the disease and treated by surgical resection.[9] The current pancreatic cancer biomarker used in clinical diagnosis, glycoprotein CA 19−9, cannot satisfy this need because of its low sensitivity and specificity.[10,11] Finding a novel biomarker or biomarker panel with higher sensitivity and specificity for pancreatic cancer diagnosis is an urgent task. This situation requires new proteomics strategies applied in pancreatic cancer biomarker discovery. A genome-wide association study identified some susceptibility loci of pancreatic cancer on chromosomes.[12] Therefore, SAAV peptide analysis in serum provides a new alternative for pancreatic biomarker discovery.

Although single nucleotide polymorphisms (SNPs) in humans have been extensively studied over the past decade, little attention has been paid to single amino acid polymorphisms (SAPs) of proteins at the proteome level, especially direct analysis in serum samples. With the development of mass spectrometry techniques and bioinformatics, there has been a limited effort to identify single amino acid variant (SAAV)

Received: September 5, 2014
Published: November 13, 2014

   dx.doi.org/10.1021/pr500934u | J. Proteome Res. 2014, 13, 6058−6066

Table 1. Characteristics of the Study Patients[a]

		normal	PDAC	type II DM	Cyst	CP	OJ
Gender	Total	27	26	25	27	25	11
	M/F	12/15	11/15	15/10	8/19	14/11	5/6
Age, y	Mean	60.00	60.00	61.00	61.00	58.00	58.00
	SEM	2.44	2.12	2.10	2.60	2.03	2.84
	Range	28−89	28−80	37−82	24−86	30−83	46−78
Race	White	27	24	10	26	24	11
	Other[b]	0	2	15	1	2	0
Stage			T1, NO = 2; T2, NO = 1	>10 y	IPMN = 15		
		n/a	T2, N1 = 2; T3, NO = 3		MCN = 5	n/a	n/a
			T3, N1 = 5; T4, N1 = 3		Unkown = 7		
			IV = 10				

[a]Abbreviations: PDAC, pancreatic ductal adenocarcinoma; DM, diabetes mellitus; CP, chronic pancreatitis; OJ, obstructive jaundice; n/a, not applicable; M/F, male/female. [b]Other races include Black/African Amerian; Bi/multi racial/Hispanic and unknown.

peptides in tissues. A mass spectrometry-friendly modified database was created by adding variant sequences in the original International Protein Sequence (IPI) database.[13] Fifteen variant peptides were identified using this database for variant peptides analysis in lysosomes isolated from MCF7 breast cancer cells.[13] To improve the efficiency of variant peptide analysis, a new variant database and an integrated data analysis workflow were developed for variant peptide identification in three colorectal cancer cell lines, where a total of 204 variant peptides were detected.[14,15] Recently, a more comprehensive variant-associated database was constructed by combining the unique variant sequences from the MS-CanProVar database (http://bioinfo.vanderbilt.edu/canprovar) and a humansavar database (http://www.uniprot.org/docs/humsavar). This was applied to variant peptide analysis in hepatocellular carcinoma (HCC) and normal human liver tissues, where 282 unique SAAV sites were identified and quantified.[16]

Recently, a novel strategy developed by the Zeng group may provide a new direction for variant peptide analysis using a de novo sequencing method.[17] The de novo sequencing strategy is independent of variant databases and may identify some cancer specific somatic mutations. However, it is difficult to analyze thousands of MS/MS spectra in a high-throughput mode. Moreover, the false positive ratio control is another important challenge for large-scale variant peptide analysis. To identify mutant proteins as cancer-specific biomarkers, selected reaction monitoring (SRM) approaches have been used to quantify a few variant peptides from the mutant Ras protein in cancer cell lines and tissues.[18] In one study, absolute concentrations of three selected SAAV peptides among 290 serum samples were investigated using a selected reaction monitoring (SRM) approach, and the results revealed that heterozygous and homozygous proteins with various SAPs (single amino acid polymorphisms) have different associations with particular traits in the population.[14,19] Although some progress has been achieved in SAAV peptide analysis, the application of large-scale quantitative analysis of variant peptides in serum is still limited.

The complexity and the wide dynamic range of serum samples make comprehensive analyses of the serum proteome a challenging task. However, screening serum glycoproteins as a powerful means to find novel markers in serum biomarker discovery has been demonstrated in previous studies.[19−21] Alterations in the glycosylation profile are now known to be associated with many diseases. The discovery and detailed characterization of glycoprotein disease biomarkers is a primary interest of biomedical research. Therefore, we have focused on variant peptides from glycoproteins in this study. An isobaric labeling quantitative strategy was applied to identify and quantify variant peptides in serum samples of pancreatic cancer patients and other benign controls. The largest number of variant peptides to date in serum were identified and quantified, where 96 unique variant peptides were quantified in this quantitative analysis, of which five variant peptides showed a statistically significant difference between pancreatic cancer and other controls. The SAAV peptide which showed a significant difference between pancreatic cancer and controls was further validated by SRM analysis.

2. MATERIALS AND METHODS

2.1. Serum Sample process

This study was approved by the Institutional Review Board of the University of Michigan Medical School. All serum samples used in this study were from age and sex-matched patients with signed informed consent. All the serum samples were collected before any treatment. In total, serum samples from 141 individual patients were investigated in this study, of which cyst, chronic pancreatic, diabetes and obstructive jaundice served as benign controls and 27 normal serum samples were used as healthy controls. Twenty-six serum samples from patients of pancreatic cancer were investigated to determine the significantly different proteins against controls. The detailed sample information is presented in Table 1.

Serum samples were processed according to previous procedures.[19] Briefly, all the serum samples were aliquoted into different tubes and stored in a −70 °C freezer. The serum samples were randomized into different sets before depleting high-abundance proteins. Fourteen high-abundance proteins were depleted in this study using a IgY-14 LC10 column (Sigma, St. Louis, MO). Pooled samples were obtained by mixing 30 normal serum samples and used as an internal standard. All other procedures were the same as before.[19] None of the samples were thawed more than twice before analysis.

2.2. TMT Labeling and Affinity Fractionation

Amine-reactive, 6-plex Tandem Mass Tag Reagents with a +229.1629 mass shift were used to label the serum sample at the protein-level as described previously.[19,22] Briefly, an equal amount of processed serum protein (100 μg) from the different disease groups and internal standard (pooled samples) were labeled by TMT reagent. After the labeled samples from the same set were combined, the buffer was exchanged into lectin affinity binding buffer for affinity fractionation. To reduce the

dx.doi.org/10.1021/pr500934u | J. Proteome Res. 2014, 13, 6058−6066

complexity of the serum sample, *Aleuria aurantia* lectin (AAL) affinity chromatography was applied to enrich the glycoprotein fraction. Labeled samples were diluted to 1 mL using binding buffer (20 mM Tris, 0.15 M NaCl, pH = 7.5, protease inhibitor 1:100), and then were loaded onto an affinity column packed with 600 μL of agarose-bound AAL lectin. After incubating for 15 min twice, binding proteins were eluted with four volumes of elution buffer (200 mM fucose in binding buffer). The eluted fraction was exchanged into 50 mM NH_4HCO_3 for digestion.

2.3. Enzymatic Digestion

Trypsin and Asp-N enzymes (Promega, Madison, WI) were used to digest the TMT labeled samples in parallel. Then, PNGase F (New England Biolabs, Ipswich, MA) was added and incubated at 37 °C for 16 h. The digested samples were desalted using C_{18} ZipTips (Millipore, Billerica, MA) and dried using a SpeedVac concentrator (Thermo Savant, Milford, MA) before LC−MS/MS analysis.

2.4. LC−MS/MS Analysis

The LC−MS/MS analysis was performed on an Eksigent Nano 2D LC System (AB SCIEX) and Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The digested peptides were loaded on a commercial New Objective ProteoPepID trap column (150 μm × 25 mm), and then separated on an C18 analytical column (75 μm × 100 mm, 5 μm, 300 A) using a 100 min linear gradient from 2 to 32% solvent B (0.1% formic acid in acetonitrile) at a flow rate of 300 nL/min. The 10 most intense peaks from the MS spectrum were selected to run MS/MS analysis using the HCD fragmentation mode. The range of full scans is from m/z 400.0 to 1800.0 with a resolution of 30 000. The MS/MS scan was fixed at starting from m/z 100.00 with a resolution of 7,500. The collision energy was set as 45% NCE. A ± 1.5 Da isolation window was applied to isolate precursor ions with dynamic exclusion of a 10 ppm exclusion window. Every precursor ion was repeated twice within a duration time of 20 s and was excluded for 20 s. The maximum injection time for FTMS full scan was set as 250 ms reaching an AGC target value of 100 000, and the maximum injection time for FT MSn scan was set as 200 ms reaching an AGC target value of 40 000.

All the collected raw data were analyzed by searching the Swiss-CanSAAVs[16] database using SEQUEST in Proteome Discoverer 1.1 (Thermo). A total of 87 745 amino acid variant sequences and 73 910 UniProtKB/Swiss-Prot canonical protein entries are included in the Swiss-CanSAAVs database (download from http://bioanalysis.dicp.ac.cn/proteomics/Publications/SAAV/SAAV-Database.htm on 11/21/2013), of which 87,745 amino acid variant sequences were constituted by combining Data set "humsavar.txt" containing human polymorphisms and disease mutations (downloaded from www.uniprot.org/docs/humsavar on 11/25/2011) and MS-CanProVar database (downloaded from http://bioinfo.vanderbilt.edu/canprovar on 12/14/2011).[16] The m/z tolerance of precursor ion and fragment ion were set as ±10 ppm and ±0.05 Da, respectively. Enzyme was allowed two missed cleavages. Carbamidomethylation (+57.02146 Da, C) and TMT 6-plex (+219.163 Da) of lysines and protein N-termini were set as fixed modifications. Oxidation (+15.99492 Da, M) and deamidation (+0.98402 Da, N) were set as dynamic modifications. The identified results were filtered using a strict standard at 1% peptide-level false discovery rate (FDR). TMT quantification analysis used the same parameters as other work in our group.[19]

2.5. Selected Reaction Monitoring (SRM) Analysis and ELISA Assay

The SRM analysis was performed using a TSQ Quantum Ultra AM (Thermo-Finnigan, San Jose, CA) coupled with a Surveyor HPLC system (ThermoFisher, San Jose, CA). For each serum sample, about 0.7 μL of serum digest spiked with 4 pmol stable isotope labeled standard peptides was injected for analysis. The LC buffer system was: buffer A, 0.1% formic acid/water and buffer B, 0.1% formic acid/acetonitrile. The peptides were separated and eluted at a flow rate of 250 μL/min, with a LC gradient from 5% to 35% B in 25 min on a 2.1 × 150 mm C18 analytical column (3 μm, Atlantis T3, Waters). The samples were analyzed using a spray voltage at 4000 V and a capillary temperature of 300 °C. The SRM analyses were performed with a scan time of 0.05 s and a scan width of 0.002 m/z, using a unit resolution of 0.7 Da for both Q1 and Q3. The collision energy for each transition and tube lens voltage for each peptide were experimentally optimized by infusing each isotopically labeled peptide solution. The optimum transitions for each targeted peptide were determined empirically using the synthetic reference peptides. Before running each batch of samples, a mixture containing 4 pmol of each isotopically heavy peptide in 0.1% formic acid was injected to establish the retention time and chromatographic peak characteristics of each reference peptide. SRM quantification analysis was performed using Xcalibur 2.1 (Thermoscientific) software. For each targeted peptide, the precursor peak and coelution of the associated transitions were manually inspected and confirmed.

The ELISA kits of protein AACT, THBS1, and CA19−9 were purchased from Abcam, GenWay and Abnova, respectively. On the basis of the manufacturer's instructions, the duplicate readings for each standard and sample were averaged, followed by subtraction of the average density of the blank. Good linear calibration curves were generated to determine the amount of each protein.

2.6. Preparation of Stable Isotope-Labeled Heavy Peptides as Internal Standards

Isotopically labeled heavy forms (containing either a C-terminal [^{13}C6^{15}N2] Lys residue or [^{13}C6^{15}N] Leu residue) of peptides were synthesized by AnaSpec, Inc. as absolute quantification peptides with known quantities. All the synthetic peptides have a purity of ≥95%. All peptides were resuspended in 50% acetonitrile (ACN), 0.1% formic acid, and then diluted into 5% ACN containing 0.1% formic acid during the preparation of dilution series.

2.7. Statistical Analysis

All statistical analyses were performed using SPSS 16. The AUC value was used to evaluate the biomarker's performance by constructing receiver operating characteristic (ROC) curves. Multivariate analysis was done by logistic regression to obtain the best-fitting multivariate model for each biomarker panel. The Student's t test and one-way analysis of variance (ANOVA) were applied to calculate a p-value to determine statistical difference.

■ RESULTS AND DISCUSSION

Large-Scale Quantification of Pancreatic Cancer Related SAAV Peptides in Serum Samples

In this study, we investigate SAAV peptides associated with pancreatic cancer in serum using a quantitative proteomics

dx.doi.org/10.1021/pr500934u | J. Proteome Res. 2014, 13, 6058−6066

strategy. Recently, several studies of SAAV proteomics analyses conducted in cells or tissue samples have been reported.[23−26] Few SAAV proteomics analysis have been performed in serum samples. However, serum is widely used in clinical diagnosis because of its noninvasive advantage. Therefore, identification of SAAV peptides in serum may be a useful strategy to find potential markers.

As shown in Figure 1, the serum samples from different disease groups and healthy controls were used to perform

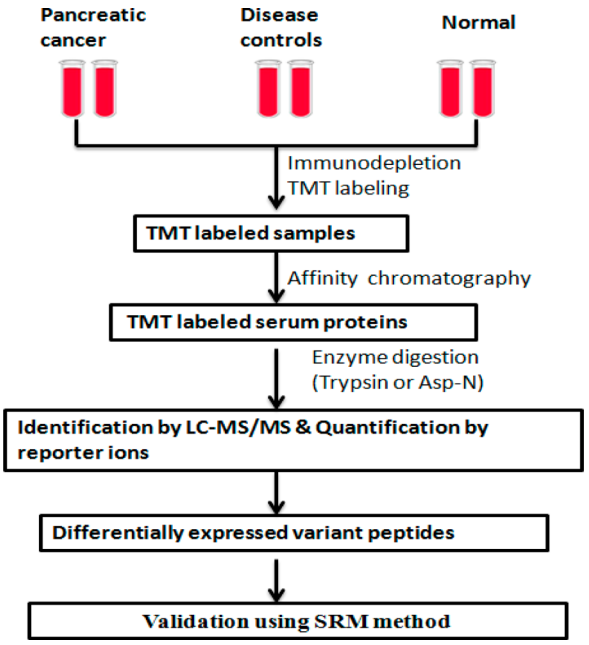


Figure 1. Experimental workflow for quantitative analysis of single amino acid variant peptides associated with pancreatic cancer by isobaric labeling strategy.

SAAV peptide analysis, where cyst, chronic pancreatic, diabetes and obstructive jaundice served as disease controls. The patients' characteristics are presented in Table 1. In this work, all serum samples were first depleted of 14 high abundance proteins, then labeled at protein-level using TMT 6 plex isobaric labeling reagents. A labeled mixture from different sample groups was fractionated by affinity chromatography. Trypsin and Asp-N were applied to digest labeled samples, respectively. The digested peptides were analyzed using an Orbitrap Velos Pro mass spectrometer in duplicate. Variant peptides were quantified based on the intensity of reporter ions. The ratio was calculated by comparing with the internal standard. Significantly different variant peptides were further validated by the SRM analysis.

The SAAV peptides cannot be identified using standard databases because the variant sequences are typically absent from generic proteomic databases. As a result, a considerable portion of high-quality tandem mass spectra collected is left unassigned. To identify SAAV peptides, a database containing 87 745 amino acid variant sequences and 73 910 UniProtKB/Swiss-Prot canonical protein entries was constructed for database search,[27] which enables one to analyze SAAVs on a large-scale. Therefore, a database combining amino acid variant sequences and UniProtKB/Swiss-Prot canonical protein entries was applied to search the raw data. After filtering with a 1% FDR, 96 quantified SAAV peptides were obtained in total from 141 serum samples, including 83 single amino acid variant sites (Supporting Information Table S1). After statistical analysis, five SAAV peptides were determined to be significantly different variant peptides (p value < 0.05, Table 2). This is probably the largest scale quantification data set of pancreatic cancer associated variant peptides in serum, which is larger than that performed in a recent study of diabetic serum samples.[14]

To increase the identification of SAAV peptides, trypsin and Asp-N were used to digest the serum samples. In the total of 83 SAAV sites, 47 SAAV sites were identified from trypsin digests and 36 SAAV sites were detected from Asp-N digests. The result suggests different enzymes increase the identification of SAAV peptides in serum. In this analysis, some variant peptides and their corresponding canonical peptides were identified simultaneously. The matched results of their MS/MS spectra clearly confirmed the variant sites. For example: MS/MS spectra of the variant peptide GGGAGFISGL***T***YLELDN-PAGNKR and canonical peptide GGGAGFISGL***S***YLELDN-PAGNKR from Complement component C7 (CO7) are shown in Figure 2. A +14 Da mass shift corresponding to the substitution of threonine (Thr) for serine (Ser) at position 389 was indicated by a series of y13, y14, y15, y16 and y17 ions, which confirms the variant site. The MS/MS spectra of SAAV peptide DEGKLQHL***E***NELTH and corresponding canonical peptide also clearly indicate variant site (E → V) from alpha-1-antitrypsin (Supporting Information Figure S1). To validate the identification results of SAAV peptides, all the MS/MS spectra were manually checked based on the following criteria: (1) the fragment ion peaks have a high signal-to-noise ratio, (2) and show sequential members of the b- or y-ion series, and (3) the major peaks were matched. Furthermore, two variant sites were further confirmed by synthetic peptides containing variant sites (Supporting Information Figure S2). As a result, highly confident and large-scale SAAV peptide data were identified in serum samples of pancreatic cancer.

Validation of SAAV Peptides by SRM Analysis

These SAAV peptides were further validated by SRM analysis according to the following rules:[19,28,29] (1) These peptides were

Table 2. Variant Peptides with Significantly Differential Abundance Identified in Pancreatic Cancer Serum Samples

variant peptides[a]	accession number	mutant site	protein name	number of samples detected	mean difference	*P*-value
TLLVFEVQQPFLF***M***LWDQQHkFPVFMGR	P05155	V → M	Plasma protease C1 inhibitor	12	7.01	0.001
DEGkLQHL***V***NELTH	P01009	E → V	Alpha-1-antitrypsin	7	3.15	0.002
DLkYLSFTLTkLS***H***	Q96MN2	R → H	NACHT, LRR and PYD domains-containing protein 4	8	1.95	0.016
DPkASLLT***K***AFLnGAL	Q96PD5	M → K	*N*-acetylmuramoyl-L-alanine amidase	83	0.76	0.030
DTPEAGYFA***V***AVVkkSAS	P02787	I → V	Serotransferrin	78	0.678	0.039

[a]Bold and italic letter denote mutant site; n denotes deaminated modification.

dx.doi.org/10.1021/pr500934u | J. Proteome Res. 2014, 13, 6058−6066

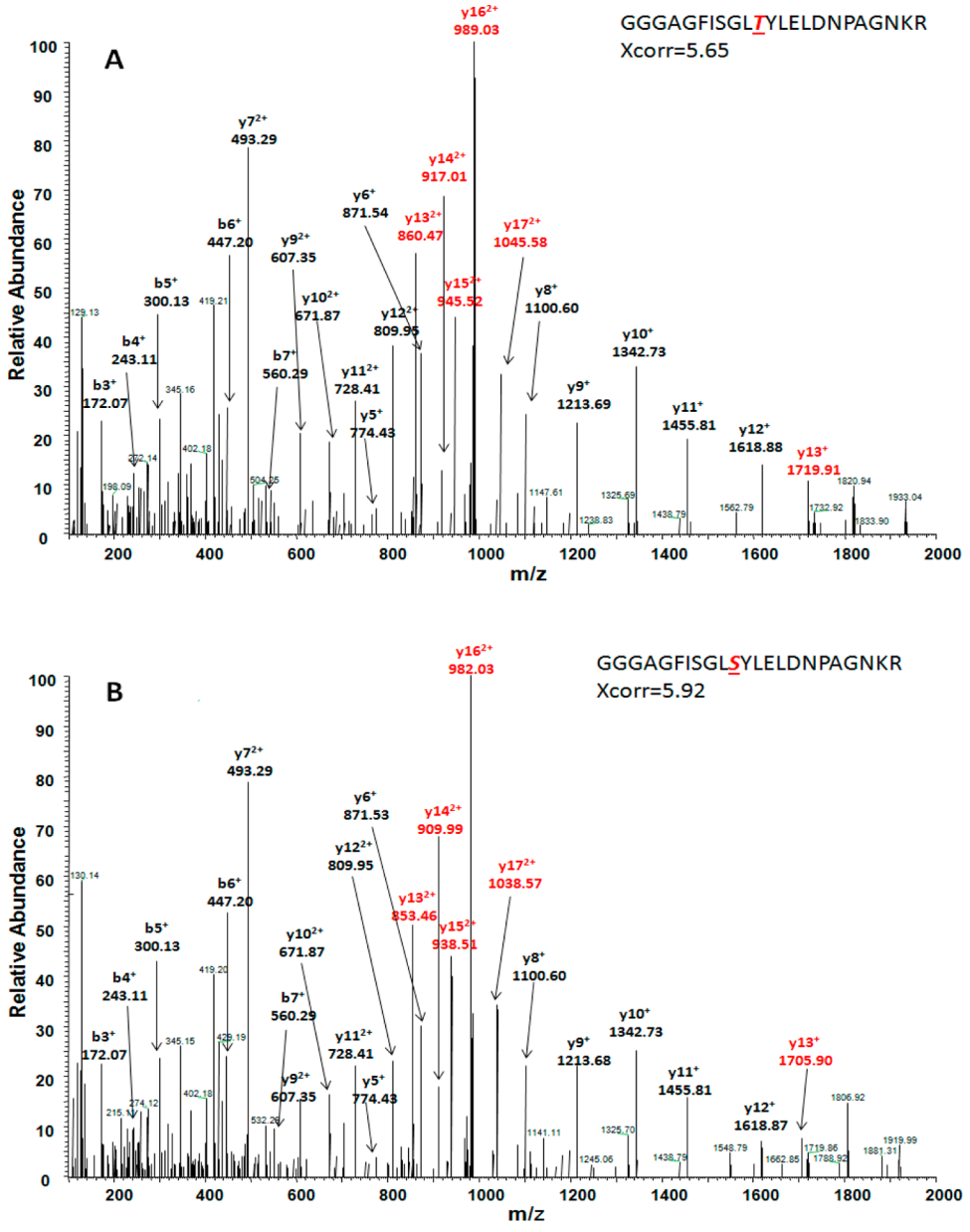


Figure 2. MS/MS spectra of the variant peptide GGGAGFISGL*T*YLELDNPAGNKR (A) and canonical peptide GGGAGFISGL*S*YLELDNPAGNKR (B) from Complement component C7 (CO7). The underlined and italic letter indicates the variant site. A +14 Da mass shift corresponding to the substitution of threonine (Thr) for serine (Ser) at position 389 was indicated by the series of y13, y14, y15, y16 and y17 ions, which confirms the variant site.

Table 3. Transitions and Collision Energy for SRM Experiments

protein		peptide sequence	peptide molecular mass (Da)	precursor ion (m/z)	Transition (m/z)	fragment ion	CE (v)
Serotransferrin	light	SDNCEDTPEAGYFA*V*AVVK	2013.9	1036.5	416.3	y4	40
					822.2	b7	38
					1250.7	y12	42
	Heavy	SDNCEDTPEAGYFA*V*AVVK[a]	2021.9	1040.5	424.2	y4	40
					822.2	b7	38
					1258.8	y12	42

[a]Denotes isotope labeling amino acid residue (8 Da shift). Retention time: 19.32 min.

quantified in more than half of the analyzed samples in the experiments; (2) no methionine (M) amino acid (AA) exists in the peptide; (3) the length of peptides is in the range of 7−25 AA. In the discovery phase, we used trypsin and Asp-N enzyme to digest serum samples in order to increase the identification of SAAV peptides. In the SRM analysis, original serum samples were applied. On the basis of the high cost of Asp-N consideration, only the trypsin enzyme was used to perform digestion. Finally, the SAAV peptide DTPEAGYFA*V*AVVkk-SAS from serotransferrin with a substitution of valine (Val) for

dx.doi.org/10.1021/pr500934u | *J. Proteome Res.* 2014, 13, 6058−6066

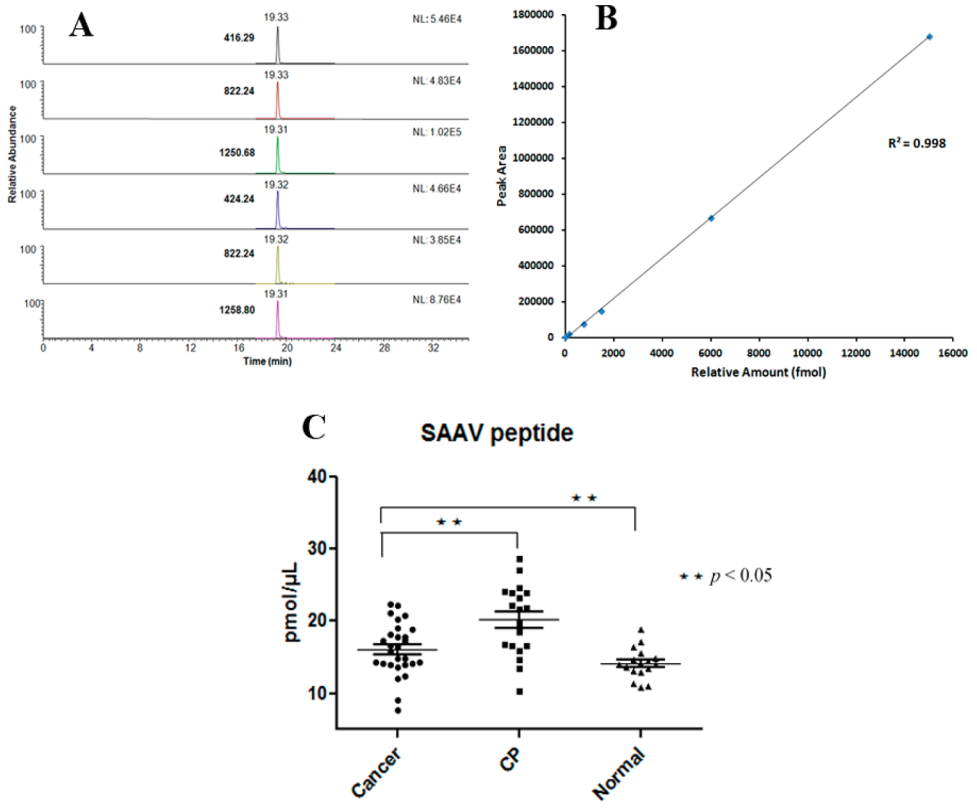


Figure 3. Variant peptide (SDNCEDTPEAGYFA*V*AVVK) from serotransferrin was verified and quantified by spiking heavy labeling peptide into serum samples. The TIC of the variant peptide (A), standard curve (B) and scatter plot of the amount of variant peptide in different disease groups (C).

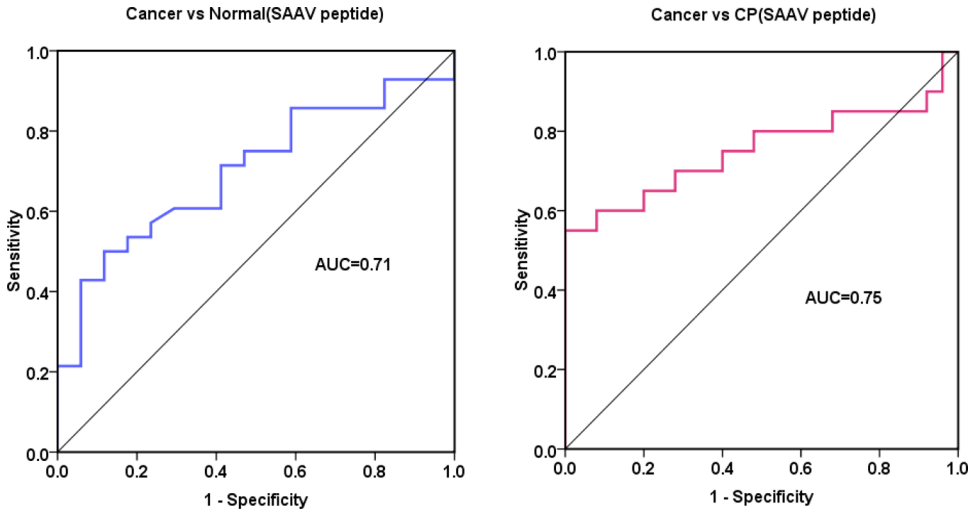


Figure 4. Performance of variant peptide based on the results of SRM assay. the ROC curve and AUC value are presented.

isoleucine (Ile) at position 448 was selected for SRM analysis. The other four significant peptides were not selected for validation because they did not meet the strict filter criterion. Using trypsin digestion, the variant site was located in the peptide of SDNCEDTPEAGYFA*V*AVVK. The optimum collision energies and transitions of the SAAV peptide were determined empirically using a stable isotopic-labeled synthetic peptide (Table 3). Three optimum transitions were selected for quantification in SRM assay.

The same coelution time and a CV of <25% in the average relative ratio of the SRM transitions are considered as being interference-free.[30,31] Figure 3A presents the same retention time and similar average relative ratio from labeled peptide and targeted peptide, which indicates no interference for these transitions. Simultaneously, a series of different concentrations of diluted synthetic peptides were analyzed in duplicate with serum matrix. The curve showed excellent linearity with a correlation of 0.99 (R2) from 15 fmol to 15 pmol (Figure 3B). The same amount of isotope labeled peptides were spiked into serum digests. Totally, 27 cancer, 20 CP and 17 normal samples were analyzed in sequence. The transition with the smallest CV was selected to represent the SAAV peptide. The scatter plot of

dx.doi.org/10.1021/pr500934u | *J. Proteome Res.* 2014, 13, 6058−6066

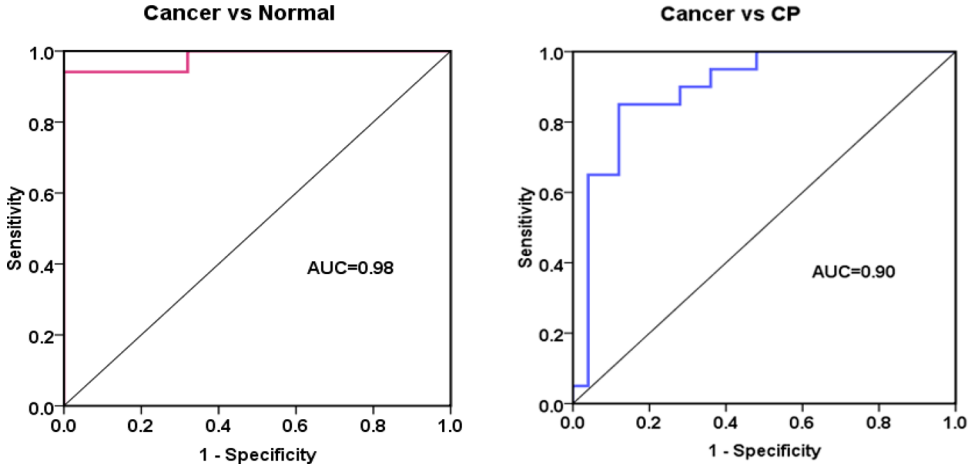


Figure 5. Performance of biomarker combination of SAAV peptide and THBS1, AACT, the ROC curve and AUC value are shown.

the quantitative results of SAAV peptides is shown in Figure 3C. Statistical analysis showed that a significant difference existed between pancreatic cancer and CP, as well as between cancer and normal samples (p-value < 0.05). The serotransferrin protein also was identified and quantified in these experiments. The results showed that there is no significant difference between cancer serum samples and controls, which means the significant difference of SAAV peptide (DTPEAGYFAVAVVkkSAS) is not derived from the difference of the parent protein (Supporting Information Figure S3). Unfortunately, the corresponding canonical peptides were not quantified in our experiments. Otherwise. we could quantify the fraction of SAAV peptides by comparing the abundance of the SAAV peptide with the corresponding canonical peptides.

Performance of Potential Biomarker Panel

To evaluate the performance of the SAAV peptide as a potential marker, AUC values were acquired by generating ROC curves. As shown in Figure 4, the AUC value between pancreatic cancer and normal is 0.71. An improved AUC value (0.75) is obtained between pancreatic cancer and CP. In previous work, the protein AACT and THBS1 were discovered as promising biomarker candidates.[19] Further ELISA analysis of the two proteins found that the SAAV peptide is complementary to these two candidate proteins. A novel biomarker panel was obtained by combining AACT, THBS1 and SAAV peptide, where its performance was investigated for diagnosing pancreatic cancer. The results revealed that the panel has a strong capability for differentiating pancreatic cancer with healthy controls (AUC = 0.98) and CP controls (AUC = 0.90) (Figure 5).

Currently, the performance of CA 19−9 marker can not satisfy the need to detect early pancreatic cancer.[32] In this study, CA 19−9 ELISA assay also was performed on the same serum samples. The CA19−9 performance between cancer and CP or normal is an AUC value of 0.79 and 0.89, respectively, which is consistent to other reports.[32,19] The biomarker panel of AACT, THBS1 and SAAV peptide has an improved performance compared with CA19−9. If we combine CA 19−9 with the biomarker panel, the performance of this combination in discriminating pancreatic cancer with CP reaches to an AUC value of ∼0.988 with 100% sensitivity at 92% specificity, and reaches to an AUC value of 0.998 in distinguishing cancer from normal controls with 100% sensitivity at 96% specificity. It should be noted that the same set of samples was used in the different assays. Thus, a risk of overfitting may exist. To reduce the effect of overfitting, two measures were performed in this data analysis. First, the markers were simply linearly combined with weights proportional to their marginal correlation coefficient with the outcome instead of using optimized regression fits in the multimarker analysis. Second, cross validation was performed in all performance metrics calculation.

CONCLUSIONS

The largest-scale quantification data set of SAAV peptides to date was discovered in pancreatic cancer serum using quantitative strategies. A total of 96 unique SAAV peptides were quantified in this quantitative analysis, of which five variant peptides showed a significant difference between pancreatic cancer and other controls. The SAAV peptide from serotransferrin protein showed its potential as a potential biomarker by further SRM validation analysis. A promising combination of AACT, THBS1 proteins and the SAAV peptide was achieved, which has an excellent performance in discriminating pancreatic cancer with other controls. These results show that large-scale analysis of SAAV peptides in serum may hold a promising future for biomarker discovery research.

ASSOCIATED CONTENT

Supporting Information

Supplementary Figure S1, MS/MS spectra of the variant peptide DEGkLQHL*V*NELTH (A) and canonical peptide DEGkLQHL*E*NELTH (B) from alpha-1-antitrypsin (A1AT). Supplementary Figure S2, MS/MS spectra of the synthetic isotope-labeled peptides SDNCEDTPEAGYFAVAVVK* (A) and SL*GNIIMVCR (B) from serotransferrin and complement factor H proteins, respectively. Supplementary Figure S3, the quantification results of serotransferrin protein in the serum samples from the patients with pancreatic cancer and other controls (p-value = 0.32). Supplementary Table S1, detailed identification and quantitative information for single amino acid variant peptides. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Address: Department of Surgery, The University of Michigan Medical Center, 1150 West Medical Center Drive, Building

MSRB1 Room A510B, Ann Arbor, MI 48109-0656. E-mail: dmlubman@umich.edu. Telephone: 734-647-8834. Fax: 734-615-2088.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful for support from the National Cancer Institute through SPORE Program Grant 1 P50CA130810 (S.N., D.M.L., M.T.R.), Grants 1 R01 CA154455 (D.M.L) and R01 GM49500 (D.M.L).

■ REFERENCES

(1) Abecasis, G. R.; Auton, A.; Brooks, L. D.; DePristo, M. A.; Durbin, R. M.; Handsaker, R. E.; Kang, H. M.; Marth, G. T.; McVean, G. A. An integrated map of genetic variation from 1,092 human genomes. *Nature* **2012**, *491* (7422), 56−65.

(2) Cargill, M.; Altshuler, D.; Ireland, J.; Sklar, P.; Ardlie, K.; Patil, N.; Shaw, N.; Lane, C. R.; Lim, E. P.; Kalyanaraman, N.; Nemesh, J.; Ziaugra, L.; Friedland, L.; Rolfe, A.; Warrington, J.; Lipshutz, R.; Daley, G. Q.; Lander, E. S. Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat. Genet.* **1999**, *22* (3), 231−8.

(3) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls . *Nature* **2007**, *447*, (7145), 661−78.

(4) Bozic, I.; Antal, T.; Ohtsuki, H.; Carter, H.; Kim, D.; Chen, S.; Karchin, R.; Kinzler, K. W.; Vogelstein, B.; Nowak, M. A. Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci. U.S A.* **2010**, *107* (43), 18545−50.

(5) Bignell, G. R.; Greenman, C. D.; Davies, H.; Butler, A. P.; Edkins, S.; Andrews, J. M.; Buck, G.; Chen, L.; Beare, D.; Latimer, C.; Widaa, S.; Hinton, J.; Fahey, C.; Fu, B.; Swamy, S.; Dalgliesh, G. L.; Teh, B. T.; Deloukas, P.; Yang, F.; Campbell, P. J.; Futreal, P. A.; Stratton, M. R. Signatures of mutation and selection in the cancer genome. *Nature* **2010**, *463* (7283), 893−8.

(6) Jemal, A.; Siegel, R.; Ward, E.; Hao, Y.; Xu, J.; Thun, M. J. Cancer statistics, 2009. *CA Cancer J. Clin.* **2009**, *59* (4), 225−49.

(7) Chu, D.; Kohlmann, W.; Adler, D. G. Identification and screening of individuals at increased risk for pancreatic cancer with emphasis on known environmental and genetic factors and hereditary syndromes. *J. Pncreas* **2010**, *11* (3), 203−12.

(8) Abraham, P.; Giannone, R. J.; Adams, R. M.; Kalluri, U.; Tuskan, G. A.; Hettich, R. L. Putting the pieces together: High-performance LC-MS/MS provides network-, pathway-, and protein-level perspectives in populus. *Mol. Cell Proteomics* **2012**, *12* (1), 106−19.

(9) Sohn, T. A.; Lillemoe, K. D.; Cameron, J. L.; Huang, J. J.; Pitt, H. A.; Yeo, C. J. Surgical palliation of unresectable periampullary adenocarcinoma in the 1990s. *J. Am. Coll. Surg.* **1999**, *188* (6), 658−66 discussion 666−9.

(10) Bunger, S.; Laubert, T.; Roblick, U. J.; Habermann, J. K. Serum biomarkers for improved diagnostic of pancreatic cancer: a current overview. *J. Cancer Res. Clin. Oncol.* **2011**, *137* (3), 375−389.

(11) Wingren, C.; Sandstrom, A.; Segersvard, R.; Carlsson, A.; Andersson, R.; Lohr, M.; Borrebaeck, C. A. Identification of serum biomarker signatures associated with pancreatic cancer. *Cancer Res.* **2012**, *72* (10), 2481−90.

(12) Petersen, G. M.; Amundadottir, L.; Fuchs, C. S.; Kraft, P.; Stolzenberg-Solomon, R. Z.; Jacobs, K. B.; Arslan, A. A.; Bueno-de-Mesquita, H. B.; Gallinger, S.; Gross, M.; Helzlsouer, K.; Holly, E. A.; Jacobs, E. J.; Klein, A. P.; LaCroix, A.; Li, D.; Mandelson, M. T.; Olson, S. H.; Risch, H. A.; Zheng, W.; Albanes, D.; Bamlet, W. R.; Berg, C. D.; Boutron-Ruault, M. C.; Buring, J. E.; Bracci, P. M.; Canzian, F.; Clipp, S.; Cotterchio, M.; de Andrade, M.; Duell, E. J.; Gaziano, J. M.; Giovannucci, E. L.; Goggins, M.; Hallmans, G.; Hankinson, S. E.; Hassan, M.; Howard, B.; Hunter, D. J.; Hutchinson, A.; Jenab, M.; Kaaks, R.; Kooperberg, C.; Krogh, V.; Kurtz, R. C.; Lynch, S. M.; McWilliams, R. R.; Mendelsohn, J. B.; Michaud, D. S.; Parikh, H.; Patel, A. V.; Peeters, P. H.; Rajkovic, A.; Riboli, E.; Rodriguez, L.; Seminara, D.; Shu, X. O.; Thomas, G.; Tjonneland, A.; Tobias, G. S.; Trichopoulos, D.; Van Den Eeden, S. K.; Virtamo, J.; Wactawski-Wende, J.; Wang, Z.; Wolpin, B. M.; Yu, H.; Yu, K.; Zeleniuch-Jacquotte, A.; Fraumeni, J. F., Jr.; Hoover, R. N.; Hartge, P.; Chanock, S. J. A genome-wide association study identifies pancreatic cancer susceptibility loci on chromosomes 13q22.1, 1q32.1 and 5p15.33. *Nat. Genet.* **2010**, *42* (3), 224−8.

(13) Schandorff, S.; Olsen, J. V.; Bunkenborg, J.; Blagoev, B.; Zhang, Y.; Andersen, J. S.; Mann, M. A mass spectrometry-friendly database for cSNP identification. *Nat. Methods* **2007**, *4* (6), 465−6.

(14) Su, Z. D.; Sun, L.; Yu, D. X.; Li, R. X.; Li, H. X.; Yu, Z. J.; Sheng, Q. H.; Lin, X.; Zeng, R.; Wu, J. R. Quantitative detection of single amino acid polymorphisms by targeted proteomics. *J. Mol. Cell Biol.* **2011**, *3* (5), 309−15.

(15) Ortensi, B.; Setti, M.; Osti, D.; Pelicci, G. Cancer stem cell contribution to glioblastoma invasiveness. *Stem Cell Res. Ther.* **2013**, *4* (1), 18.

(16) Song, C.; Wang, F.; Cheng, K.; Wei, X.; Bian, Y.; Wang, K.; Tan, Y.; Wang, H.; Ye, M.; Zou, H. Large-scale quantification of single amino-acid variations by a variation-associated database search strategy. *J. Proteome Res.* **2014**, *13* (1), 241−8.

(17) Su, Z. D.; Sheng, Q. H.; Li, Q. R.; Chi, H.; Jiang, X.; Yan, Z.; Fu, N.; He, S. M.; Khaitovich, P.; Wu, J. R.; Zeng, R. De novo identification and quantification of single amino-acid variants in human brain. *J. Mol. Cell Biol.* **2014**, *6* (5), 421−33.

(18) Wang, Q.; Chaerkady, R.; Wu, J.; Hwang, H. J.; Papadopoulos, N.; Kopelovich, L.; Maitra, A.; Matthaei, H.; Eshleman, J. R.; Hruban, R. H.; Kinzler, K. W.; Pandey, A.; Vogelstein, B. Mutant proteins as cancer-specific biomarkers. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (6), 2444−9.

(19) Nie, S.; Lo, A.; Wu, J.; Zhu, J.; Tan, Z.; Simeone, D. M.; Anderson, M. A.; Shedden, K. A.; Ruffin, M. T.; Lubman, D. M. Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis. *J. Proteome Res.* **2014**, *13* (4), 1873−84.

(20) Liu, Y.; He, J.; Li, C.; Benitez, R.; Fu, S.; Marrero, J.; Lubman, D. M. Identification and confirmation of biomarkers using an integrated platform for quantitative analysis of glycoproteins and their glycosylations. *J. Proteome Res.* **2010**, *9* (2), 798−805.

(21) Wu, J.; Xie, X.; Liu, Y.; He, J.; Benitez, R.; Buckanovich, R. J.; Lubman, D. M. Identification and confirmation of differentially expressed fucosylated glycoproteins in the serum of ovarian cancer patients using a lectin array and LC-MS/MS. *J. Proteome Res.* **2012**, *11* (9), 4541−52.

(22) Nie, S.; Lo, A.; Zhu, J.; Wu, J.; Ruffin, M. T.; Lubman, D. M. Isobaric protein-level labeling strategy for serum glycoprotein quantification analysis by liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2013**, *85* (11), 5353−7.

(23) Li, J.; Su, Z.; Ma, Z. Q.; Slebos, R. J.; Halvey, P.; Tabb, D. L.; Liebler, D. C.; Pao, W.; Zhang, B. A bioinformatics workflow for variant peptide detection in shotgun proteomics. *Mol. Cell Proteomics* **2011**, *10* (5), M110 006536.

(24) Abraham, P.; Adams, R. M.; Tuskan, G. A.; Hettich, R. L. Moving away from the reference genome: Evaluating a peptide sequencing tagging approach for single amino acid polymorphism identifications in the genus Populus. *J. Proteome Res.* **2013**, *12* (8), 3642−51.

(25) Chernobrovkin, A. L.; Mitkevich, V. A.; Popov, I. A.; Indeikina, M. I.; Ilgisonis, E. V.; Lisitsa, A. V.; Archakov, A. I. Identification of single amino acid polymorphisms in MS/MS spectra of peptides. *Dokl. Biochem. Biophys.* **2010**, *437*, 90−3.

(26) Sheynkman, G. M.; Shortreed, M. R.; Frey, B. L.; Scalf, M.; Smith, L. M. Large-scale mass spectrometric detection of variant peptides resulting from nonsynonymous nucleotide differences. *J. Proteome Res.* **2014**, *13* (1), 228−40.

(27) Song, C.; Wang, F.; Cheng, K.; Wei, X.; Bian, Y.; Wang, K.; Tan, Y.; Wang, H.; Ye, M.; Zou, H. Large-scale quantification of single

amino-Acid variations by a variation-associated database search strategy. *J. Proteome Res.* **2014**, *13* (1), 241−8.
(28) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222.
(29) Mohammed, Y.; Domanski, D.; Jackson, A. M.; Smith, D. S.; Deelder, A. M.; Palmblad, M.; Borchers, C. H. PeptidePicker: A scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J. Proteomics* **2014**, *106*, 151−61.
(30) Percy, A. J.; Chambers, A. G.; Smith, D. S.; Borchers, C. H. Standardized protocols for quality control of MRM-based plasma proteomic workflows. *J. Proteome Res.* **2012**, *12* (1), 222−33.
(31) Percy, A. J.; Chambers, A. G.; Yang, J.; Jackson, A. M.; Domanski, D.; Burkhart, J.; Sickmann, A.; Borchers, C. H. Method and platform standardization in MRM-based quantitative plasma proteomics. *J. Proteomics* **2013**, *95*, 66−76.
(32) Pleskow, D. K.; Berger, H. J.; Gyves, J.; Allen, E.; McLean, A.; Podolsky, D. K. Evaluation of a serologic marker, CA19−9, in the diagnosis of pancreatic cancer. *Ann. Int. Med.* **1989**, *110* (9), 704−9.