Published in final edited form as:

J Genet Genomics. 2015 September 20; 42(9): 507–510. doi:10.1016/j.jgg.2015.06.001.

Serological targeted analysis of an ITIH4 peptide isoform: A preterm birth biomarker and its associated SNP implications

Zhou Tan[a,b,1], **Zhongkai Hu**[b,1], **Emily Y. Cai**[b,1], **Cantas Alev**[e], **Ting Yang**[b], **Zhen Li**[b], **Joyce Sung**[c], **Yasser Yehia El-Sayed**[c], **Gary M. Shaw**[d], **David K. Stevenson**[d], **Atul J. Butte**[d], **Guojun Sheng**[e], **Karl G. Sylvester**[b], **Harvey J. Cohen**[d], **Xuefeng B. Ling**[b,*]

[a]Institute of Developmental and Regenerative Biology, Hangzhou Normal University, Hangzhou 310029, China

[b]Department of Surgery, Stanford University, Stanford, CA 94305, US

[c]Department of Obstetrics and Gynecology, Stanford University, Stanford, CA 94305, US

[d]Department of Pediatrics, Stanford University, Stanford, CA 94305, US

[e]Lab for Early Embryogenesis, RIKEN Center for Developmental Biology, Chuo-Ku, Kobe, Hyogo 650-0047, Japan

Over 11% of all pregnancies in the US result in preterm birth, greatly contributing to perinatal morbidity and mortality (Goldenberg and Rouse, 1998). Preterm birth etiologies remain largely unknown, and effective prevention methods have yet to be developed. The use of biofluid (e.g., serum or urine) for the analysis of the naturally occurring peptidome (MW < 4000) as a source of biomarkers has been reported for different diseases (Villanueva et al., 2006; Ling et al., 2010a, b, c, 2011). Mass spectrometry-based profiling of naturally occurring peptides can provide an extensive inventory of serum peptides derived from either high-abundant endogenous circulating proteins or cell and tissue proteins (Liotta and Petricoin, 2006). These peptides are usually soluble and unaffected by endogenous proteases or peptidases, and can be directly analyzed by liquid chromatography-mass spectrometry (LCMS) without additional manipulation (e.g., tryptic digestion). Esplin et al. (2011) used serum peptidomic patterns to identify patients with preterm birth. A peptide (QLGLPGPPDVPDHAAYHPF) derived from inter-alpha-trypsin inhibitor heavy chain 4 (ITIH4) was identified as a preterm birth biomarker from a predominantly (~75%) African-American cohort. The serum level of ITIH4 peptide was shown to decrease in pregnancies that resulted in preterm births, and had a sensitivity of 65.0% and specificity of 82.5% in discriminating pregnancies delivering preterm from term, although the biological activity of the parent protein or the fragment identified is unknown.

Examination of the NCBI SNP database of common gene variations and population genetics data from 1000 genomes project (Genomes Project et al., 2010) revealed that there is a

*Corresponding author. fax: +1 650 723 1154, bxling@stanford.edu (X. B. Ling).

[1]These authors contributed equally to this work.

SUPPLEMENTARY DATA

Supplementary data associated with this article can be found at http://dx.doi.org/…

single nucleotide polymorphism (SNP, variant rs2276814) in ITIH4 (position 669) where a single coding nucleotide differs from A of amino acid codon cAa to T of cTa, resulting in an amino acid change from glutamine (Q) to leucine (L). As shown in Fig. S1 and Table S1, African American or Sub-Saharan African subjects have comparable probabilities of “A” or “T” allele, and therefore, similar chances of glutamine (Q) or leucine (L) at ITIH4 position 669. In contrast, European, Asian and Hispanic American subjects are predominantly homozygous for the “T” allele, and therefore, carry leucine at position 669 of ITIH4 (“L” isoform).

Within the European, Asian and Hispanic American populations, we postulated that the ITIH4 derived serum peptide should be in its “L” isoform (protein ITIH4 669, nucleotide level A→T, protein level Q→L). This expected “L” isoform peptide sequence should be LLGLPGPPDVPDHAAYHPF, which should share an almost identical sequence as the preterm birth serum peptide biomarker with “Q” isoform (QLGLPGPPDVPDHAAYHPF) but with the first amino acid changed from L to Q.

LCMS based quantitative proteomic analysis is a powerful method for selective quantification of specific proteins/peptides in very complex mixtures. LCMS method coupled with stable isotope dilution (SID) provides both absolute structural specificity for the analyte and relative or absolute measurement of analyte concentration (Addona et al., 2009). Unlike the label-free quantitative method, which is more error-prone due to systematic variations among individual runs and stochastic nature of the indices used for calculation (Kito and Ito, 2008), SID based peptide assay is the gold standard for absolute protein quantitation *via* mass spectrometry (Fig. S2). We postulated that targeted mass spectrometry (also known as data-dependent acquisition), if focusing on “Q” isoform’s specific mass/charge ratio (m/z) and specific liquid chromatography time as reported previously (Esplin et al., 2011), would fail to read out the ITIH4 “L” isoform peptide.

We hypothesized that the ITIH4 “L” isoform peptide could be detected in the sera of Asian, European, or Hispanic backgrounds, and further that this “L” isoform serum peptide would be also a biomarker associated with preterm birth, similar to the reported “Q” isoform (Esplin et al., 2011).

To test these hypotheses, we firstly applied quantitative proteomic methods to identification of the ITIH4 “L” isoform peptide in normal pregnancy sera. Fig. 1A shows a representative LCMS ion chromatogram from the term delivery pregnancy samples (ProMedDX, USA, http://www.promeddx.com). The ProMedDX sera were from women of uncomplicated pregnancies (age, 26.1±6.86 years; 11 were of Hispanic origin and 3 African Americans) at various gestational ages. Each of the peaks in these chromatograms was formed by the elution of serum peptides. The “L” ITIH4 peptide, eluting at average 15.9 min (standard deviation, 0.23 min) from the 60-min HPLC column, was identified as doubly charged with m/z of 1006.26 and the sequence MS/MS spectral profile was found to match to the peptide sequence LLGLPGPPDVPDHAAYHPF (Fig. 1B).

Next, we conducted quantitative proteomic analysis to determine the association between the ITIH4 “L” isoform serum peptide and the preterm birth. The LCMS ion chromatogram

and the matching MS/MS spectrum of ITIH4 “L” isoform in Stanford samples are consistent with those in the ProMedDX samples. The information of Stanford samples, cases (n = 11, preterm birth) and controls (n = 14, term delivery) was summarized in Table S2. Serum samples were collected at Stanford University Medical Center under IRB approved protocols. Our study cohort consisted of predominantly Asians, Europeans, or their mixture (case: 9.1% African-American, 27.3% Asian, 18.2% Caucasian, and 45.5% Hispanic; control: 0% African American, 14.3% Asian, 7.1%

Caucasian, 71.4% Hispanic, and 7.1% Pacific islanders). The ProMedDX samples were used for “L” isoform identification, while the Stanford samples were used for the following quantitative study. As summarized in Fig. 1C, the peaks in the chromatogram were formed by the elution of “L” isoform peptide from C18 column (see Materials and Methods for details) at the HPLC 16th time point. The normalized “L” isoform peptide serum concentration in women with preterm birth was found to be 3-fold less when compared with women who had term deliveries. Scatter plot analysis (Fig. 1D) revealed that the ITIH4 “L” isoform decreased significantly (both Student’s t-test and Mann-Whitney U test, $P < 0.001$) in preterm birth subjects relative to the controls. As summarized in Fig. 1E, we analyzed the normalized ITIH4 peptide serum abundances in the term delivery pregnancy sera (ProMedDX collection) of different gestational ages. The comparative analysis of ITIH4 “L” peptide serum abundance between early (<35 weeks) and late (≥35 weeks) gestational ages revealed no statistically significant difference between the two groups (Student’s t-test, $P = 0.747$). The level of ITIH4 “L” isoform peptide does not change with gestational ages in term delivery pregnancy sera. These supported our hypotheses that 1) ITIH4 “L” isoform serum peptide can be detected in a pregnancy cohort predominantly of Asian, Hispanic and Caucasian origin; 2) reduction of the serum level of ITIH4 “L” isoform is associated with preterm birth in our cohort. The ROC AUC (area under the curve) as a measure of the classification performance is 90.2%. The cut point analysis was summarized in Fig. 1F with test sensitivity of 0.78 and specificity of 0.80.

The proteomic observation in the current study is congruent with the SNP database allele frequency records, which shows that most subjects of European, Asian and/or Hispanic American origins have “L” at ITIH4 669 position (Fig. S1 and Table S1). In addition, our observation of the lack of effect of gestational age on level of the “L” peptide in women with full term pregnancies reinforced the notion that the reduction seen in women with preterm births is indeed associated with preterm birth pathophysiology.

Our findings of the serum ITIH4 “L” isoform peptide point out the importance of the ethnic variation when applying the ITIH4 serum peptide as a preterm birth biomarker . Population SNP allele analyses indicate that most European, Asian and Hispanic American subjects have “L” at ITIH4 669. Targeted analysis of the serum ITIH4 “Q” isoform peptide in these race/ethnic subjects would be expected to reveal minimum detection. Therefore, the “Q” peptide based algorithm (Esplin et al., 2011) would misclassify it as a high-risk marker for preterm birth regardless of their physiological conditions. This could result in the wrong inference in a future trial to establish the clinical utility of ITIH4 serum peptide for prediction of gestational timing of delivery.

We believe that the association of the ITIH4 “L” isoform with preterm birth complements the previous findings of the “Q” isoform as a preterm birth biomarker. Specifically, in patients whose genotypes are heterozygous, quantifying either “L” or “Q” peptide alone and comparing with normal ITIH4 serum peptide level may lead to misdiagnosis. Consideration of both ITIH4 peptide isoforms can help to avoid false positives arising from the analysis of serum levels of a non-detectable ITIH4 peptide isoform target in the corresponding population with a particular race/ethnic background. To comprehensively apply the ITIH4 serum peptide biomarker for preterm birth analysis in all race/ethnic backgrounds, we proposed a multiple-stage algorithm (Fig. S3). At the first stage, blood cells are processed and SNP genotyping is performed to determine whether the subject’s ITIH4 669 allele is in “Q” or “L” isoform. At the second stage, upon the genotyping results, the targeted analysis of ITIH4 “Q” (homozygous genotype), “L” (homozygous genotype), or the mixed two (heterozygous genotype) in the patient serum is conducted. Finally, the serum quantity of ITIH4 peptide isoform(s) would be used, in combination with other protein markers (Esplin et al., 2011), to gauge the patient’s risk of preterm birth.

Our examination of the ITIH4 “L” isoform as a biomarker for preterm birth is carried out in a small cohort consisting of samples with different gestational age, which is a limitation of the study. Conducting a future prospective trial of our proposed multiple-stage algorithm (Fig. S3) may lead to a clinically applicable test for preterm birth in patients of diverse race/ethnic backgrounds.

Due to the SNP (rs2276814) at position 669 of ITIH4, there are “Q” or “L” peptide isoforms (Q/LGLPGPPDVPDHAAYHPF). Serum quantitative proteomics was performed accordingly to analyze the ITIH4 “L” isoform peptide in a cohort of uncomplicated pregnancies ($n = 14$) and pregnancies complicated by preterm birth ($n = 11$) subjects. Our targeted analysis showed that the reduction of the “L” peptide’s serum quantity is associated with the clinical state of preterm birth ($P < 0.001$). We concluded that “L” peptide is a potential biomarker predictive of preterm birth in the cohort, consisting predominantly of Europeans, Asians and Hispanic Americans.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors thank Kenneth Lau for MALDI mass spectrometric analysis. The authors also thank scientists at the Pediatric Proteomics group and the March of Dimes Prematurity Research Center at Stanford University for critical discussions. This work was supported in part by March of Dimes, and the Lucile Packard Foundation for Children’s Heath. This work was also supported in part by the National Natural Science Foundation of China (NSFC) to ZT (No. 31201697)

REFERENCES

Addona TA, Abbatiello SE, Schilling B, Skates SJ, Mani DR, Bunk DM, Spiegelman CH, Zimmerman LJ, Ham AJ, Keshishian H, Hall SC, Allen S, Blackman RK, Borchers CH, Buck C, Cardasis HL, Cusack MP, Dodder NG, Gibson BW, Held JM, Hiltke T, Jackson A, Johansen EB, Kinsinger CR, Li J, Mesri M, Neubert TA, Niles RK, Pulsipher TC, Ransohoff D, Rodriguez H, Rudnick PA,

Smith D, Tabb DL, Tegeler TJ, Variyath AM, Vega-Montoto LJ, Wahlander A, Waldemarson S, Wang M, Whiteaker JR, Zhao L, Anderson NL, Fisher SJ, Liebler DC, Paulovich AG, Regnier FE, Tempst P, Carr SA, 2009. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nature biotechnology 27, 633–641. Use the journal abbreviation name

Esplin MS, Merrell K, Goldenberg R, Lai Y, Iams JD, Mercer B, Spong CY, Miodovnik M, Simhan HN, van Dorsten P, Dombrowski M, , 2011. Proteomic identification of serum peptides predicting subsequent spontaneous preterm birth. American journal of obstetrics and gynecology 204, 391 e391–398.

Genomes Project, C., Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA, 2010. A map of human genome variation from population-scale sequencing. Nature 467, 1061–1073. [PubMed: 20981092]

Goldenberg RL, Rouse DJ, 1998. Prevention of premature birth. The New England journal of medicine 339, 313–320. [PubMed: 9682045]

Kito K, Ito T, 2008. Mass spectrometry-based approaches toward absolute quantitative proteomics. Current genomics 9, 263–274. [PubMed: 19452043]

Ling XB, Lau K, Deshpande C, Park JL, Milojevic D, Macaubas C, Xiao C, Lopez-Avila V, Kanegaye J, Burns JC, Cohen H, Schilling J, Mellins ED, 2010a. Urine peptidomic and targeted plasma protein analyses in the diagnosis and monitoring of systemic juvenile idiopathic arthritis. Clinical proteomics 6, 175–193. [PubMed: 21124648]

Ling XB, Lau K, Kanegaye JT, Pan Z, Peng S, Ji J, Liu G, Sato Y, Yu TT, Whitin JC, Schilling J, Burns JC, Cohen HJ, 2011. A diagnostic algorithm combining clinical and molecular data distinguishes Kawasaki disease from other febrile illnesses. BMC medicine 9, 130. [PubMed: 22145762]

Ling XB, Mellins ED, Sylvester KG, Cohen HJ, 2010b. Urine peptidomics for clinical biomarker discovery. Advances in clinical chemistry 51, 181–213. [PubMed: 20857622]

Ling XB, Sigdel TK, Lau K, Ying L, Lau I, Schilling J, Sarwal MM, 2010c. Integrative urinary peptidomics in renal transplantation identifies biomarkers for acute rejection. Journal of the American Society of Nephrology : JASN 21, 646–653. [PubMed: 20150539]

Liotta LA, Petricoin EF, 2006. Serum peptidome for cancer detection: spinning biologic trash into diagnostic gold. The Journal of clinical investigation 116, 26–30. [PubMed: 16395400]

Villanueva J, Shaffer DR, Philip J, Chaparro CA, Erdjument-Bromage H, Olshen AB, Fleisher M, Lilja H, Brogi E, Boyd J, Sanchez-Carbayo M, Holland EC, Cordon-Cardo C, Scher HI, Tempst P, 2006. Differential exoprotease activities confer tumor-specific serum peptidome patterns. The Journal of clinical investigation 116, 271–284. [PubMed: 16395409]

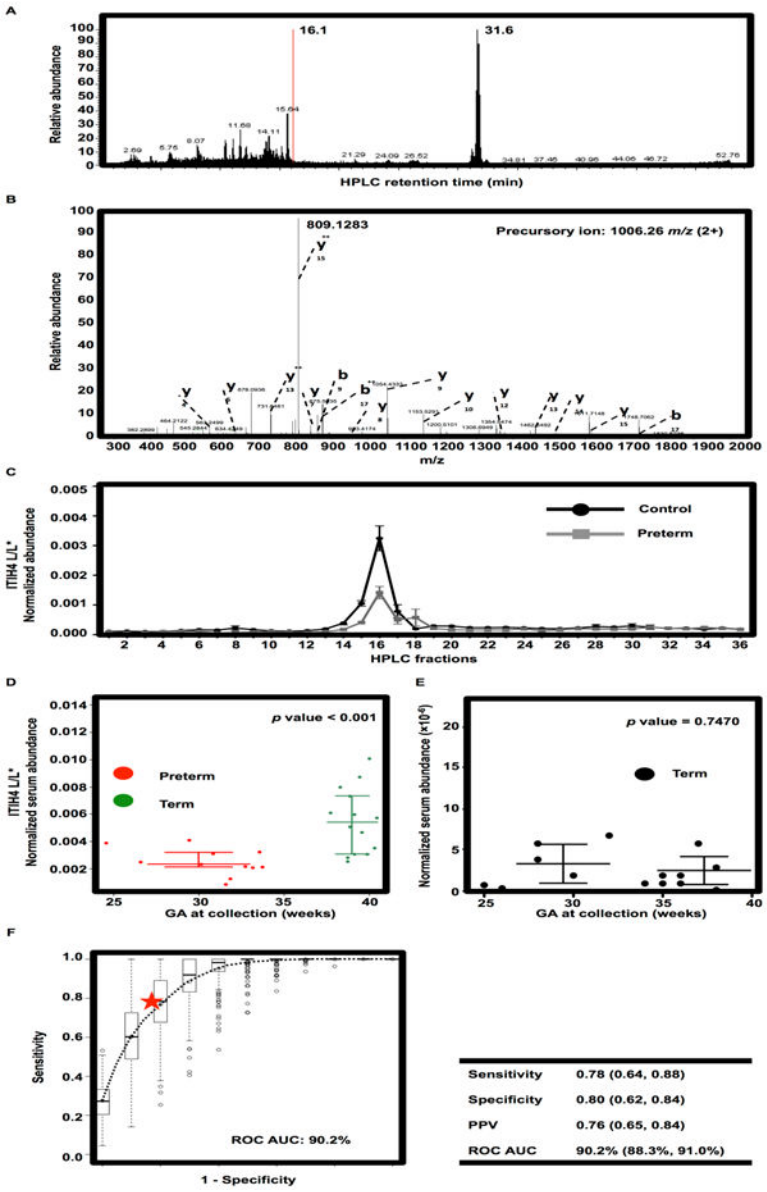

Fig. 1. The ITIH4“L” isoform serum peptide is a potential biomarker associated with preterm birth.

A: Representative LCMS ion chromatogram. The red vertical line indicates that the ITIH4 “L” peptide precursor ion with the matching MS/MS spectrum was found eluting at 16.1 min. B: An MS/MS spectrum of precursory ion m/z 1006.26 (2+) that matched to the ITIH4 “L” peptide with sequence LLGLPGPPDVPDHAAYHPF. C: ITIH4 “L” isoform peptide chromatogram derived from preterm birth cases (gray) and healthy pregnancy controls (black). The peaks in the chromatogram were formed by the elution of “L” isoform peptides at the HPLC 16th time point, and the ionic intensities of ITIH4 “L” isoform peptide were normalized with the stable isotope labeled spiked-in ITIH4 “L*” isoform peptide. The normalized “L” isoform peptide abundances, plotted with the standard deviations, were higher in healthy pregnancy controls than in preterm birth cases. D: Scatter plot analysis of each subject’s normalized serum abundance as a function of the baby gestational age at

the time of sample collection. Violet red represents preterm birth cases; sea green represents full term pregnancy controls. For either preterm birth or full term birth, bars of 25th, 50th and 75th percentiles represent and compare the overall trend of biomarker scoring as a function of the sample classification. E: ITIH4 “L” isoform peptide abundance in term delivery pregnancy sera of different gestational ages. Comparative analysis of ITIH4 “L” peptide serum abundances between early (< 35 weeks) and late (≥ 35 weeks) GA revealed no statistically significant difference between the two groups (Student’s *t* test, $P = 0.747$). F: ROC analysis of the ITIH4 “L” isoform serum peptide as a preterm birth biomarker. 500 testing data sets, generated by bootstrapping, from the normalized abundance data were used to derive estimates of standard errors and confidence intervals for our ROC analysis. The plotted ROC curve is the vertical average of the 500 bootstrapping runs, and the box and whisker plots show the vertical spread around the average. The red star denotes the cut point with the optimal sensitivity and specificity of the assay. Summary of the cut point performance of sensitivity specificity, positive predictive value, ROC AUC and their 95% confidence intervals are listed in the right table.