SUPPORTING INFORMATION

Identification of Blood Protein Biomarkers of Acute Liver Injury by Targeted Quantitative Proteomics in Acetaminophen and Carbon tetrachloride treated Mouse Models and Acetaminophen Overdose Patients

Shizhen Qin[1], Yong Zhou[1], Li Gray[1], Ulrike Kusebauch[1], Laurence McEvoy[3], Daniel J Antoine[3], Lucy Hampson[3], Kevin B Park[3], David Campbell[1], Juan Caballero[1], Gustavo Glusman[1], Xiaowei Yan[1], Taek-Kyun Kim[1], Yue Yuan[1], Kai Wang[1], Lee Rowen[1], Robert L Moritz[1], Gilbert S Omenn[1,2], Munir Pirmohamed [3]* and Leroy Hood[1]*

1. Institute for Systems Biology, USA
2. Departments of Computational Medicine & Bioinformatics, Internal Medicine, an Human Genetics and School of Public Health, University of Michigan, USA
3. Institute of Translational Medicine at University of Liverpool, England

Table of contents

SUPPORTING INFORMATION

Figure S1: Characterizing individual SRM assays for the 5 informative proteins and ALT, AST in APAP overdose patients; Figure

Figure S2: Detection of endogenous peptide VNEAACDIAR after APAP or CCl4 treatment in mouse plasma; Figure

Figure S3: Collision energy optimization of 185 peptides

Supplementary Tables S1-S7 - Qin et al.xlsx.

Table S1: Number of mice in each treatment group
Table S2: Information for 14 APAP overdose patients

Table S3: Numbers of targeted proteins and their proteotypic peptides (Protein/Peptide) filtered for SRM in each step of target protein selection

Table S4.1: SRM methods to monitor mouse liver proteins in APAP and CCL4 liver-toxicity mouse models

Table S4.2: SRM methods to monitor human liver proteins in APAP overdose serum samples

Table S5: No relationship between plasma APAP concentration in blood and liver injury after APAP injection

Table S6: Total Glutathione concentration and GSH/GSSG ratio in 5 NOD mice at drug administration

Table S7: Fold change of 23 level-up proteins in individual APAP overdose patients measured by SRM as compared to healthy volunteer controls.

SUPPORTING INFORMATION

Methods

Liver injury assessment after drug administration in mice

The degree of liver injury in mice was assessed by the ALT and AST enzyme activities in mouse plasma following the manufacturer's instructions (TECO Diagnostics, Anaheim, CA). Reaction products were read in a Beckman DU800 spectrophotometer at 340nm wavelength (250nm background). Specimens were analyzed in duplicate on the day of collection. Samples collected at night were kept at 4°C and analyzed the next morning. This colorimetric method has an acceptable accuracy (1, 2). Liver injury was also studied by histology H&E staining. Sections of mouse livers were preserved in 10% buffered formalin for H&E staining and image interpretation in the Department of Pathology at the University of Washington. Briefly, following fixation, tissues were processed into paraffin, sectioned and placed on slides. Tissue sections were then deparaffinized in xylene, hydrated through graded ethanol and stained with Mayer's hematoxylin followed by eosin Y. The slides were then dehydrated with graded ethanol, cleared in xylene and coverslipped using Permount.

Construction of mouse and human liver-enriched protein list

Liver-enriched proteins for mouse and human were generated respectively with multiple microarray and RNA-Seq datasets. Gene Atlas microarray datasets include 3 human and 3 mouse multi-tissue transcriptome series: NCBI/GEO: GSE1133, for both human and mouse (3), GSE3526, human (4), GSE2361, human (5), GSE9954, mouse (6) and GSE10246, mouse (7). For the mouse and human liver-enriched gene list, microarray raw data were processed with R:Bioconductor (8) and normalized with the Robust Multi-Array Average expression measure method (9). For each series, the liver-enrichment of each transcript was computed as the log10-ratio of the liver sample against the maximal value of the other tissues combined. The

distribution of log10 ratios is used to compute the distribution and respective *p*-value. The 165 mouse transcripts with >10-fold liver enrichment over the maximal value of the rest of any tissues (not-liver) and *p*-value $<10^{-6}$ were filtered as liver-enriched.

The human liver-enriched protein list, in addition to microarray data analysis, data mining results from Illumina Body Map and an in-house ISB human-tissue atlas were incorporated. Transcriptomes of normal human tissues were obtained from Illumina 75 bp single-end raw RNAseq data of 16 normal adult human tissues (NCBI/GEO GSE30611, Illumina Body Map). We also used an in-house ISB TranscriptomeAtlas of 75 bp single-end transcriptomes with 50 normal embryonic and adult human tissues totaling 2,014,386,871 sequence reads (> 2 TB data) [*Caballero, manuscript in preparation*]. Datasets were preprocessed, filtered, mapped with BLAT (10) and assembled with Cufflinks (11) through an in-house RNAseq pipeline. We quantified the expression patterns of 32,313 genes (Ensembl 62 annotation) and 136,825 isoforms thereof, as well as 12,803 putative novel genes and their 23,677 isoforms. For each series, we computed the liver-enrichment as the log10-ratio of the liver sample and the maximal RPKM value of the other tissues combined. The distribution of log10-ratios is used to compute the distribution and respective p-value. We filtered the results by selecting genes with >10-fold liver enrichment over the maximal value of the rest of any tissues (not-liver) and p-value$<10^{-6}$.

SRM method generation and optimization

As described previously (12), target peptides and transitions were selected to ensure they are proteotypic to the protein of interest. ISB proteomics databases, e.g., PeptideAtlas for mouse (13) and SRMAtlas for human (14), were used to facilitate this process.

In the mouse study, three steps were involved in target peptide filtering and transition selection. First, an initial test was performed to assess the detectability of peptides selected from liver-enriched proteins in pooled depleted plasma under our SRM conditions. Up to 5 peptides for each protein were selected and tested. For each peptide, both charge 2 and charge 3 precursors and as many as possible suitable transitions were monitored by SRM without heavy-labeled internal standards. The initial list contained a total of 547 peptides derived from 131 liver-enriched proteins. Peptides eluting within a window of ±3 min of the expected RT and with all monitored Q1/Q3 transitions observed were counted as "possibly detected".

Second, crude unpurified peptide standards corresponding to the possibly detected light peptide analogues were synthesized with heavy isotopic lysine ($^{13}C_6{}^{15}N_2$) or arginine ($^{13}C_6{}^{15}N_4$) at the C-termini (heavy PEPscreen® peptides, Sigma-Aldrich, St. Louis, MO, USA). SRM method conditions were optimized and titration curves were generated for each peptide.

Third, successfully detected heavy peptides were spiked into the pooled mouse samples to verify the correct identification of the light peptides detected in the first step. An example of the selection and verification for peptide *VNEAACDIAR* selected from a liver enriched-protein, BHMT1, is shown in Supplementary Figure 2. Similar steps were employed in the human study. Since we noticed a great agreement of transition rank and peak intensities between our results and SRMAtlas, only the top 2 peptides from each protein (a few proteins only have one good representative peptide) and top 4 transitions from each peptide were chosen, resulting in significantly less peptides and transitions tested in step one in the human study. Validated

peptides and transitions were used for the SRM final tests in mouse or human samples. The numbers of proteins and peptides in each step are summarized in Supplementary Table 3.

Detailed information including gene symbol, protein name, Uniprot ID, peptide sequence, precursor m/z, protein cellular location, function, process and pathway, and secretion signal prediction of the 52 mouse and 66 human proteins is summarized in Supplementary Table S4_Mouse and S4_Human.

Collision energy optimization and titration of heavy standards

Collision energies (CE) calculated with the instrument vendor-provided default formula (*CE=0.036*m/z − 4.8*) were optimized for the 134 mouse peptides as charge 2 and charge 3 precursors with 4 additional CE conditions (±5V, ±10V). In general, charge 2 precursors performed better with the calculated default or a higher CE (default plus 5V). This is in agreement with the previous observation that the default formula worked well for charge 2 precursors (15). As for charge 3 precursors, a 5V lower CE (CE values between default and default minus 5V were not tested) contributed to stronger signal intensity for the majority of the 132 peptides investigated (Supplementary Figure 3). The best precursors and each of their best 4 transitions under optimized conditions were selected for the SRM analysis on mouse samples. Human heavy peptide standards selected based on SRMAtlas were tested with the calculated CE without further optimization. Heavy peptides were titrated in a mouse plasma or human serum background to generate titration curves for the determination of the linear relationship between peak area under the curve (AUC) and the desired spike-in peptide amount of each peptide standard.

Figure legends

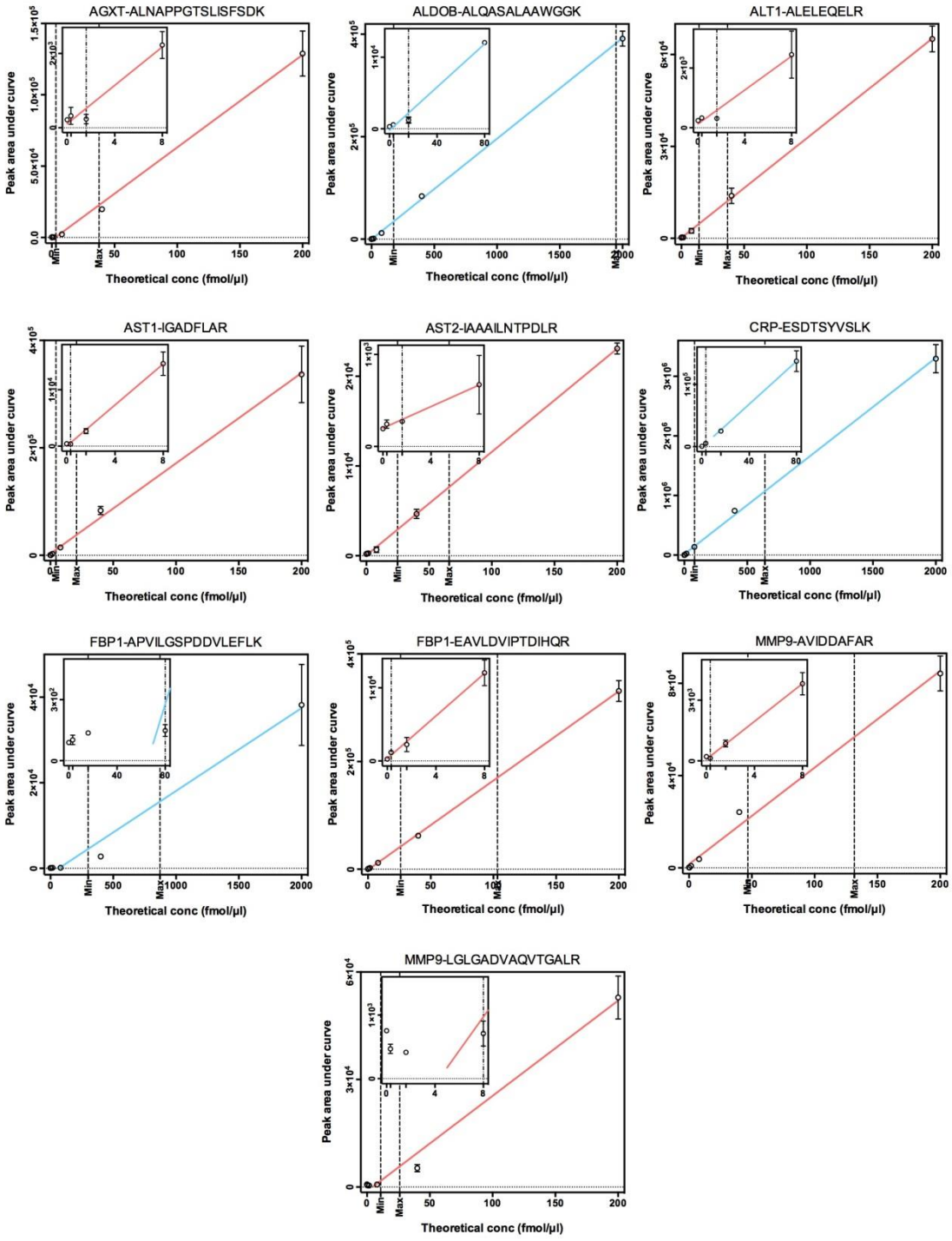
Supplemental Figure 1. Characterizing individual SRM assays for the 5 informative proteins and ALT, AST in APAP overdose patients. The calibration curves were generated in a reference depleted serum matrix. Error bars at each point are one standard deviation from the mean. The two vertical grid lines in each figure represent the detected range (Min and Max) in APAP overdose patients. The insets show a zoomed portion of the curve fitting around the limit of detection. The vertical dot lines in inserts are estimated limit of quantification (LOQ), which is lower than the targeted range in APAP overdose patients. With a relatively low LOD, all peptides spanned a linear region in 2-3 orders of magnitude (n = 2 for each transition at all concentration points).

Supplementary Figure 2. BHMT1 (peptide VNEAACDIAR) transitions were detected from NOD mouse plasma samples treated with APAP or CCl4 and later confirmed with corresponding synthetic heavy peptide standards. **A1**: Light (endogenous) peptide was not detected in plasma from drug vehicle-treated mice, only non-specific peaks are shown; **A2**: light peptide was detected in plasma from APAP-treated mouse plasma; **A3**: heavy peptide standard peptide shown similar retention time (RT) and transition peaks as light peptides in **A2**. **B1-B3**: same as **A1-A3** except that mice were treated with CCl4.

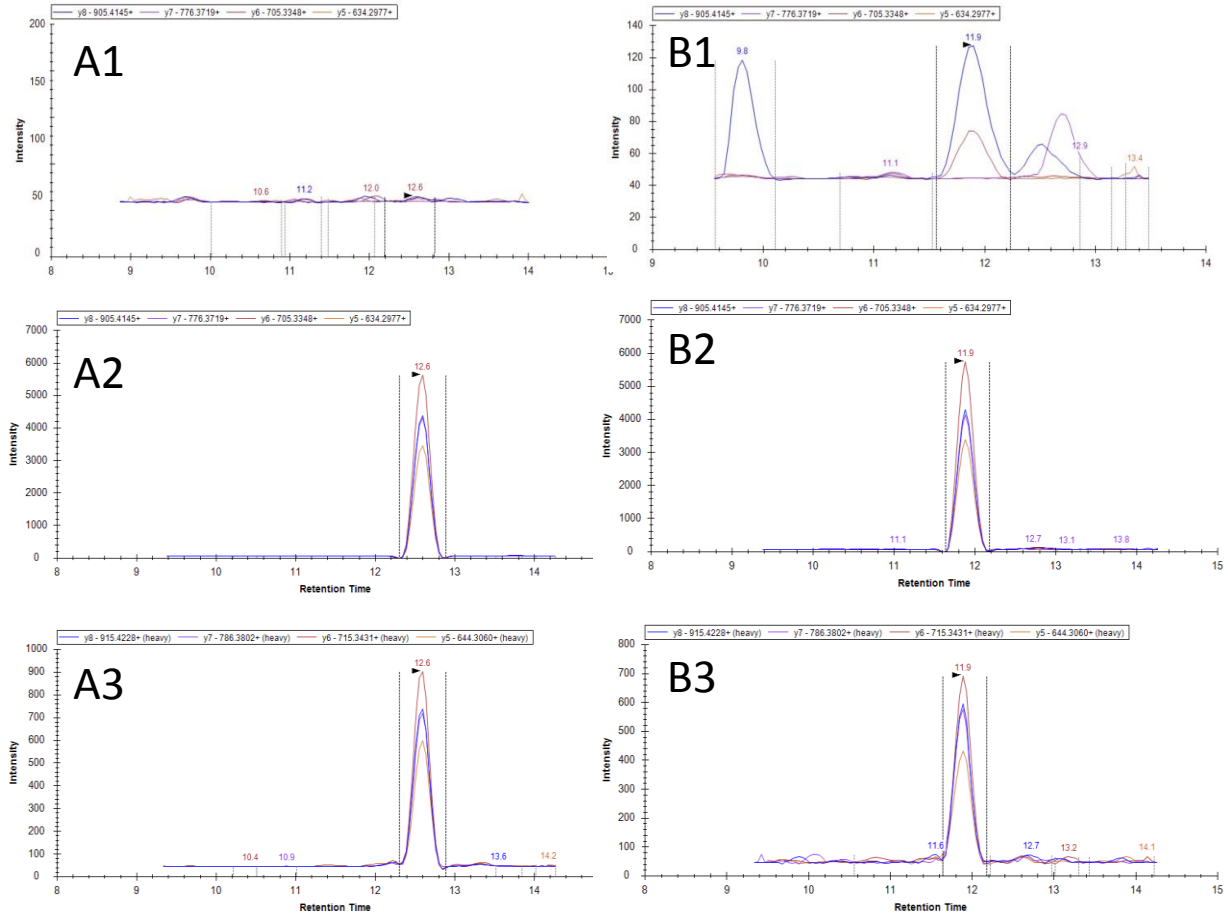
Supplementary Figure 3. Both charge 2 and charge 3 precursors of 185 peptides were optimized as described in methods. Peptides were tested under default CE (calculated) and default -10V, default – 5V, default + 5V and default +10V CEs. Some peptides were detected only as charge 2 or charge 3 precursors. **3A**. Most of the 168 charge 2 precursor ions tend to

perform better with default CE or higher CE (default plus 5 or 10V). **3B**. Most of the 166 charge 3 precursors performed better by lowering the CE 5V from default (default minus 5), a few performed better at default minus 10V. Each bar represents the normalized average value of duplicate runs of total AUC of the peptide tested at indicated CE value. Red line is the average AUC of all peptides tested at each CE value with one standard deviation error bars. X-axis, collision energy, Y-axis, normalized total area under curve for each charge 2 or charge 3 peptide measured with indicated CE.

Supplementary Figure 1

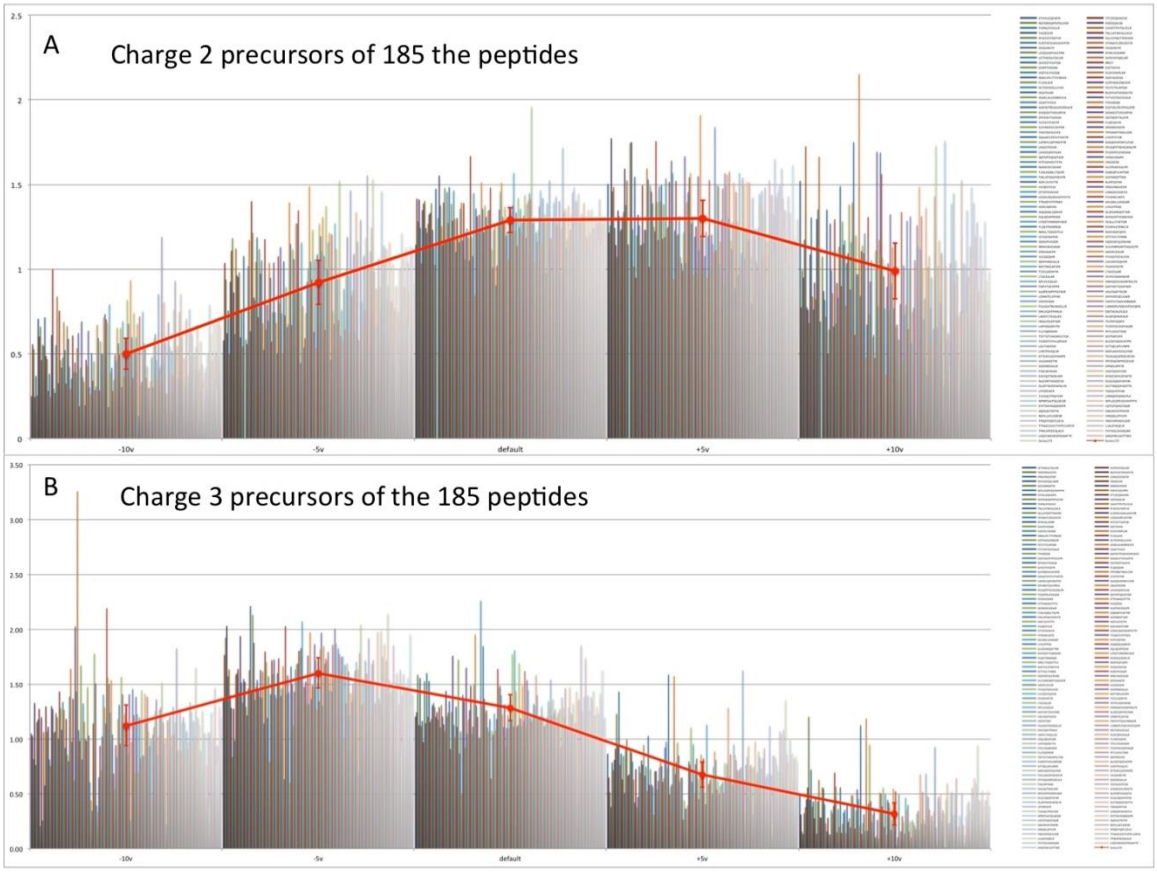


Supplemental Figure 2



SUPPORTING INFORMATION

Supplemental Figure 3



SUPPORTING INFORMATION

Reference:

1. Doumas, B.; Biggs, H. G., A colorimetric method for assaying serum aspartate aminotransferase activities. *Clin Chim Acta* **1969,** 23, (1), 75-82.
2. Henry, R. J.; Chiamori, N.; Golub, O. J.; Berkman, S., Revised spectrophotometric methods for the determination of glutamic-oxalacetic transaminase, glutamic-pyruvic transaminase, and lactic acid dehydrogenase. *Am J Clin Pathol* **1960,** 34, 381-98.
3. Su, A. I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K. A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G.; Cooke, M. P.; Walker, J. R.; Hogenesch, J. B., A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc Natl Acad Sci U S A* **2004,** 101, (16), 6062-7.
4. Roth, R. B.; Hevezi, P.; Lee, J.; Willhite, D.; Lechner, S. M.; Foster, A. C.; Zlotnik, A., Gene expression analyses reveal molecular relationships among 20 regions of the human CNS. *Neurogenetics* **2006,** 7, (2), 67-80.
5. Ge, X.; Yamamoto, S.; Tsutsumi, S.; Midorikawa, Y.; Ihara, S.; Wang, S. M.; Aburatani, H., Interpreting expression profiles of cancers by genome-wide survey of breadth of expression in normal tissues. *Genomics* **2005,** 86, (2), 127-41.
6. Thorrez, L.; Van Deun, K.; Tranchevent, L. C.; Van Lommel, L.; Engelen, K.; Marchal, K.; Moreau, Y.; Van Mechelen, I.; Schuit, F., Using ribosomal protein genes as reference: a tale of caution. *PLoS One* **2008,** 3, (3), e1854.
7. Lattin, J. E.; Schroder, K.; Su, A. I.; Walker, J. R.; Zhang, J.; Wiltshire, T.; Saijo, K.; Glass, C. K.; Hume, D. A.; Kellie, S.; Sweet, M. J., Expression analysis of G Protein-Coupled Receptors in mouse macrophages. *Immunome Res* **2008,** 4, 5.
8. Gentleman, R. C.; Carey, V. J.; Bates, D. M.; Bolstad, B.; Dettling, M.; Dudoit, S.; Ellis, B.; Gautier, L.; Ge, Y.; Gentry, J.; Hornik, K.; Hothorn, T.; Huber, W.; Iacus, S.; Irizarry, R.; Leisch, F.; Li, C.; Maechler, M.; Rossini, A. J.; Sawitzki, G.; Smith, C.; Smyth, G.; Tierney, L.; Yang, J. Y.; Zhang, J., Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* **2004,** 5, (10), R80.
9. Irizarry, R. A.; Hobbs, B.; Collin, F.; Beazer-Barclay, Y. D.; Antonellis, K. J.; Scherf, U.; Speed, T. P., Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **2003,** 4, (2), 249-64.
10. Kent, W. J., BLAT--the BLAST-like alignment tool. *Genome Res* **2002,** 12, (4), 656-64.
11. Trapnell, C.; Williams, B. A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M. J.; Salzberg, S. L.; Wold, B. J.; Pachter, L., Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* **2010,** 28, (5), 511-5.
12. Qin, S.; Zhou, Y.; Lok, A. S.; Tsodikov, A.; Yan, X.; Gray, L.; Yuan, M.; Moritz, R. L.; Galas, D.; Omenn, G. S.; Hood, L., SRM targeted proteomics in search for biomarkers of HCV-induced progression of fibrosis to cirrhosis in HALT-C patients. *Proteomics* **2012,** 12, (8), 1244-52.
13. Zhang, Q.; Menon, R.; Deutsch, E. W.; Pitteri, S. J.; Faca, V. M.; Wang, H.; Newcomb, L. F.; Depinho, R. A.; Bardeesy, N.; Dinulescu, D.; Hung, K. E.; Kucherlapati, R.; Jacks, T.; Politi, K.; Aebersold, R.; Omenn, G. S.; States, D. J.; Hanash, S. M., A mouse plasma peptide atlas as a resource for disease proteomics. *Genome Biol* **2008,** 9, (6), R93.
14. Kusebauch, U.; Deutsch, E. W.; Campbell, D. S.; Sun, Z.; Farrah, T.; Moritz, R. L., Using PeptideAtlas, SRMAtlas, and PASSEL: Comprehensive Resources for Discovery and Targeted Proteomics. *Curr Protoc Bioinformatics* **2014,** 46, 13 25 1-13 25 28.

15. Maclean, B.; Tomazela, D. M.; Abbatiello, S. E.; Zhang, S.; Whiteaker, J. R.; Paulovich, A. G.; Carr, S. A.; Maccoss, M. J., Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal Chem* **2010,** 82, (24), 10116-24.