Supplementary notes and figures for "A strategy for degradomic-peptidomic analysis of the human blood plasma" by Yufeng Shen[1], Tao Liu[1], Nikola Tolić[2], Brianne O. Petritis[1], Rui Zhao[2], Ronald J. Moore[1], Samuel O. Purvine[2], David G. Camp II[1], and Richard D. Smith[1,2]
[1]Biological Science Division, and [2]Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA 99352

Supplementary notes:

For the database searching-UStag method, datasets were initially searched against the IPI human sequence database (69,731 protein sequence entries, ipi.HUMAN.v3.39 downloaded from ftp://ftp.ebi.ac.uk/pub/databases/IPI/). The top 10 candidates for each FT MS/MS spectrum were output from the database search engine [SEQUEST, TurboSEQUEST - PVM Master v.27 (rev. 12), (c) 1998-2005] without processing the raw data prior to each search. For each candidate the theoretical isotopic envelopes for the molecule and fragments were generated using ICR2LS (http://ncrr.pnl.gov/software/ICR2LS.stm using parameters given in Supplementary Notes). For the candidate fragments, we matched the candidate theoretical *b* and *y* ion fragments against the de-isotoped monoisotopic masses (i.e., deconvolved isotopic envelopes) of the spectrum. The *b* and *y* fragments were accepted as assigned, provided the match was achieved with a mass error tolerance of 10 ppm. Consecutive *b* or *y* ion fragments were used to construct the amino acid sequences, which were then searched against the IPI human sequence database to examine sequence uniqueness. Only unique sequences (referred to as UStags) were used to identify peptides, and then the theoretical isotopic envelope and mass were compared to that of the precursor in the MS/MS spectrum for peptides identified. If the mass agreed to within 10 ppm, then the peptide was considered as unmodified and added directly to the identification list. If the mass error was > 10 ppm, the method described below was utilized.

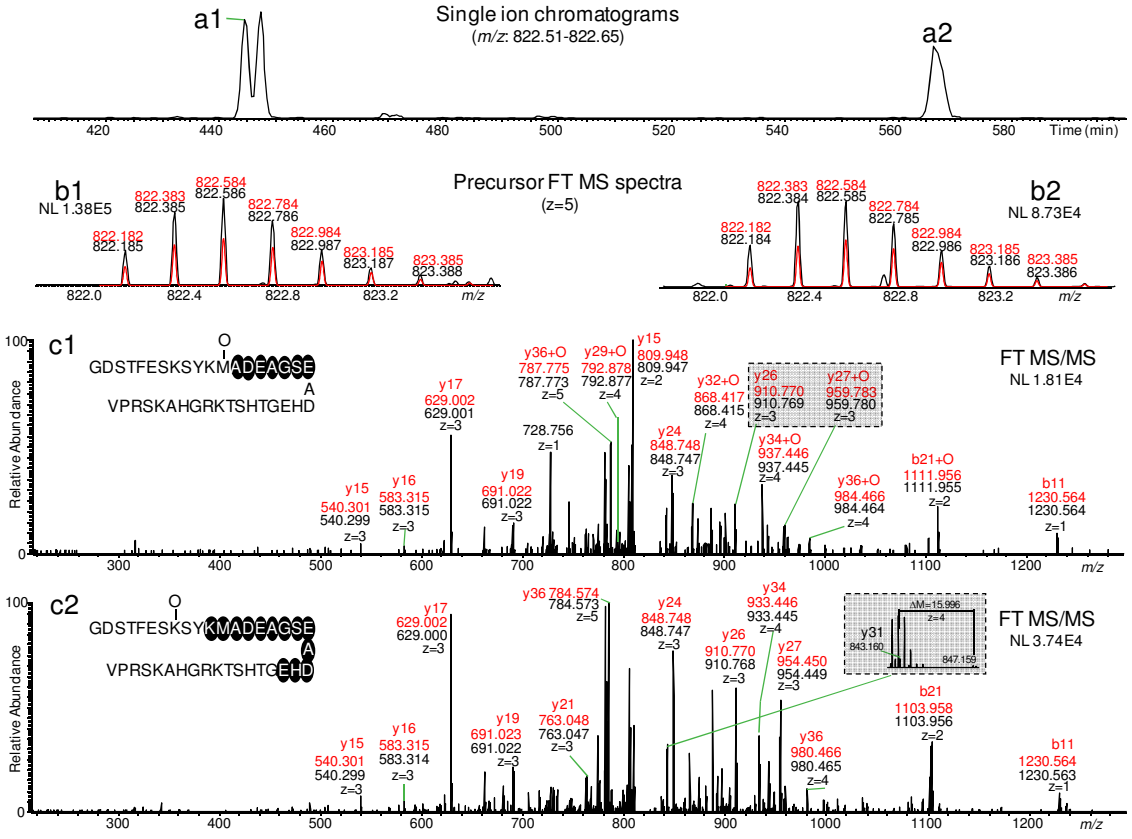
De novo sequencing-UStags were obtained based on the de-isotoped monoisotopic masses from MS/MS spectra. The *de novo* function was recursive in nature with the objective of extending the constructed sequence with an additional residue per each call (candidate peptide sequence match). A mass error tolerance of 0.005 u was used for sequencing, and no gaps of >1 amino acid residues were allowed. The amino acid sequences reported by the software were matched to the IPI human sequence database to inspect for uniqueness (i.e., the criterion for establishing a UStag), and simultaneously, to determine the location of a peptide sequence identified as a UStag. If the molecular and fragment masses of a UStags-aligned peptide sequence were in agreement with FT MS and MS/MS measurements (i.e., within a mass tolerance of 10 ppm), then the peptide was considered as unmodified and was also added to the identification list. If the masses were not in agreement, then the peptide was considered as modified, and the modification was determined as described below.

When peptides identified using UStags (obtained from either the database searching or *de novo* sequencing approaches described above) had molecular masses with >10 ppm mass errors, then prefix and suffix masses were determined. Theoretical mass values were calculated according to the location of the UStags among database sequences, while experimental values were derived from masses of the smallest fragments in the UStags (for prefixes) and the mass differences between the precursor and largest fragment (for suffixes). The differences between the experimental and theoretical values were searched against the UNIMOD modification list (http://www.unimod.org/) with a mass tolerance of 0.005 u, and modification candidates were output. Assigning specific sites of modification(s) to the residue(s) of the peptide prefix and/or suffix was accomplished by inspecting the FT MS/MS spectra manually in this work. False identification rates were examined using a decoy random database that was constructed by reversing each protein sequence in the IPI human sequence database. Finally, protein annotations and information from EBI (http://www.ebi.ac.uk/) and Swiss-Prot (http://ca.expasy.org) were used for identified proteins.

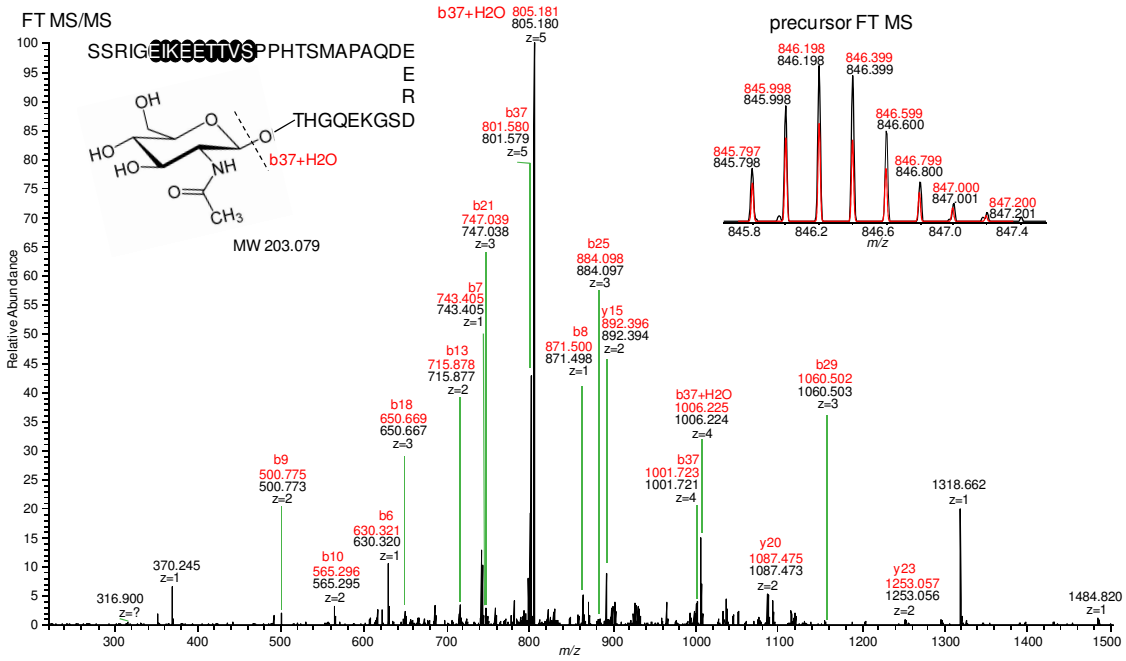
The description and parameter settings of ICR2LS used for this study are detailed as followings:
The ICR2LS is the software developed in-house for analysis of mass spectrometry data for proteomics study. The ICR2LS (http://ncrr.pnl.gov/software/ICR2LS.stm) COM interface paired with VBA in MS Excel was used to generate theoretical isotopic distributions, peak picking, and de-isotoping of experimental spectra. Implemented in ICR2LS, the theoretical isotopic distributions were generated using the Mercury function (Rockwood, A.L. et al. J. Am. Soc. Mass Spectrom. **2004**, *15*, 12-21); de-isotoping was based on THRASH algorithm (Horn, D.M. et al. J. Am. Soc. Mass Spectrom. **2000**, *11*,320-332); monoisotopic masses were derived using the Averagine algorithm (Senko, M.W. et al. J. Am. Soc. Mass Spectrom. **1995**, *6*, 229–233). The attached text file (which can be loaded into the ICR2LS from Tools menu Mass Transform Parameters command) specifies the settings.

chkRequireMultiple=1	require multiple peaks for isotopic envelope
chkAutoCorrelationCS=1	use auto correlation (Patterson) algorithm to calculate charge state
optISOleastsqarea=True	fit function is least square for area under the isotopic peaks.
txtMinPeakSN=2	local (peak) signal to noise
txtFit=.4	acceptable fit of observed and theoretical isotopic envelope. Perfect fit is 0; for the high throughputbottom-up LC-MS experiments acceptable fit is 0.25; tolerance is here wider since larger masses are targeted.
optGaussian=True	peak shape model

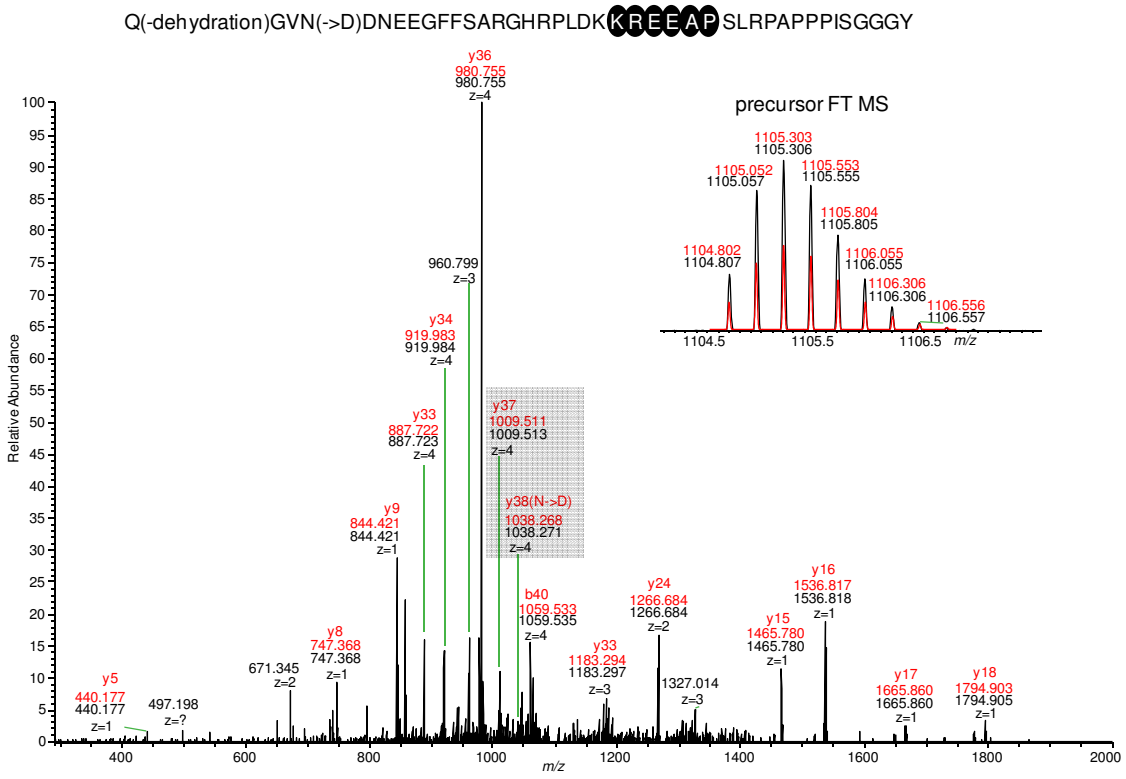
Supplementary Figures



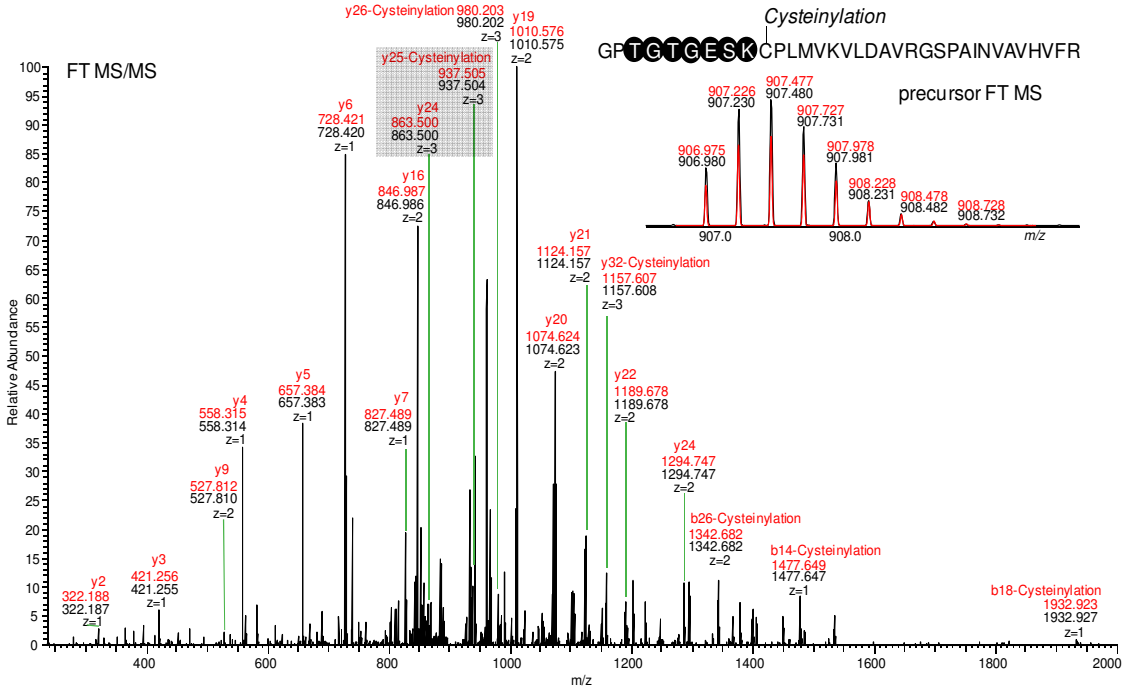
Shen *et al*. Supplementary Figure 1. Identification of oxidation on different sites of the same blood plasma peptidome peptide. The black color-circled amino acid sequences represent the UStags. Black peaks and numbers represent measurements and red, predictions of the assigned-peptide sequence. The fragment ions used for determination of the oxidized amino acids are highlighted in grey frames. The identification method is described in details in the manuscript text.



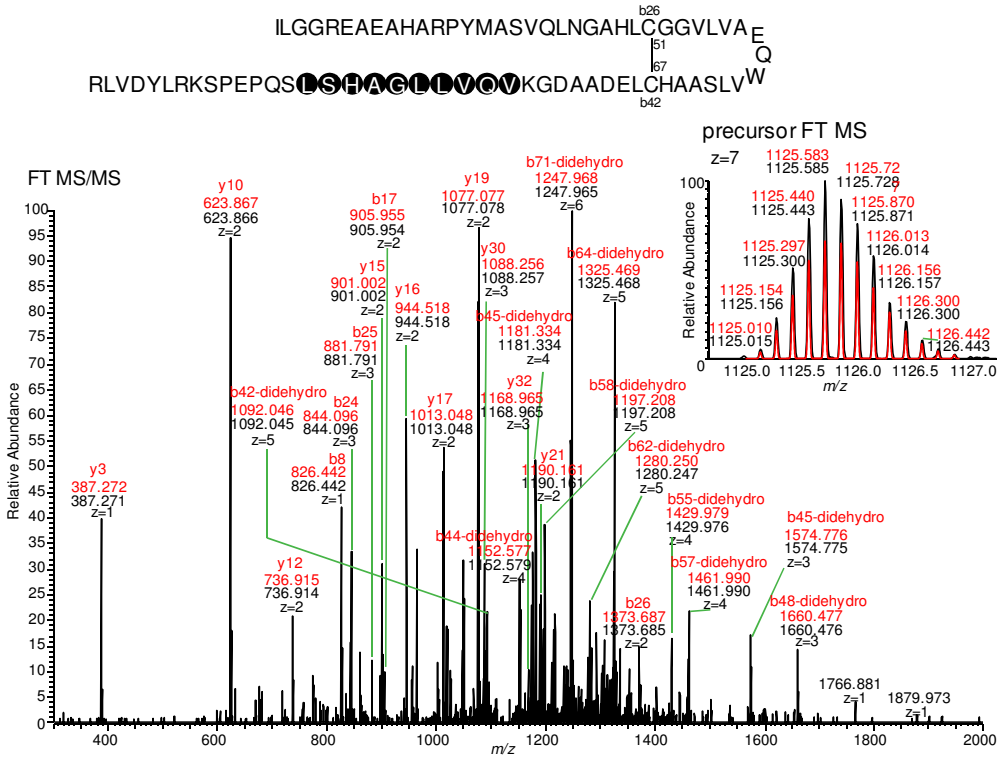
Shen *et al*. Supplementary Figure 2. Identification of acetylhexosamine on threoninesame of the blood plasma peptidome peptide. The fragment ion used for determination of acetylhexosamine (*b*37) is labeled. Others are the same as for Figure 1.



Shen *et al*. Supplementary Figure 3. Identification of N->D (or deamidation) and dehydration on Q in the blood plasma peptidome peptide. The fragment ion used for determination of the modification is highlighted in grey frame. Others are the same as for Figure 1.



Shen *et al*. Supplementary Figure 4. Identification of Cys-cysteinylation in the blood peptidome peptide. The fragment ion used for determination of the cysteinylated Cys is highlighted in grey frame. Others are the same as for Figure 1.



Shen *et al*. Supplementary Figure 5. Evidences showing existence of Cys-Cys disulfide in the blood peptidome peptide. The fragment ions used for determination are labeled. Others are the same as for Supplementary Figure 1.