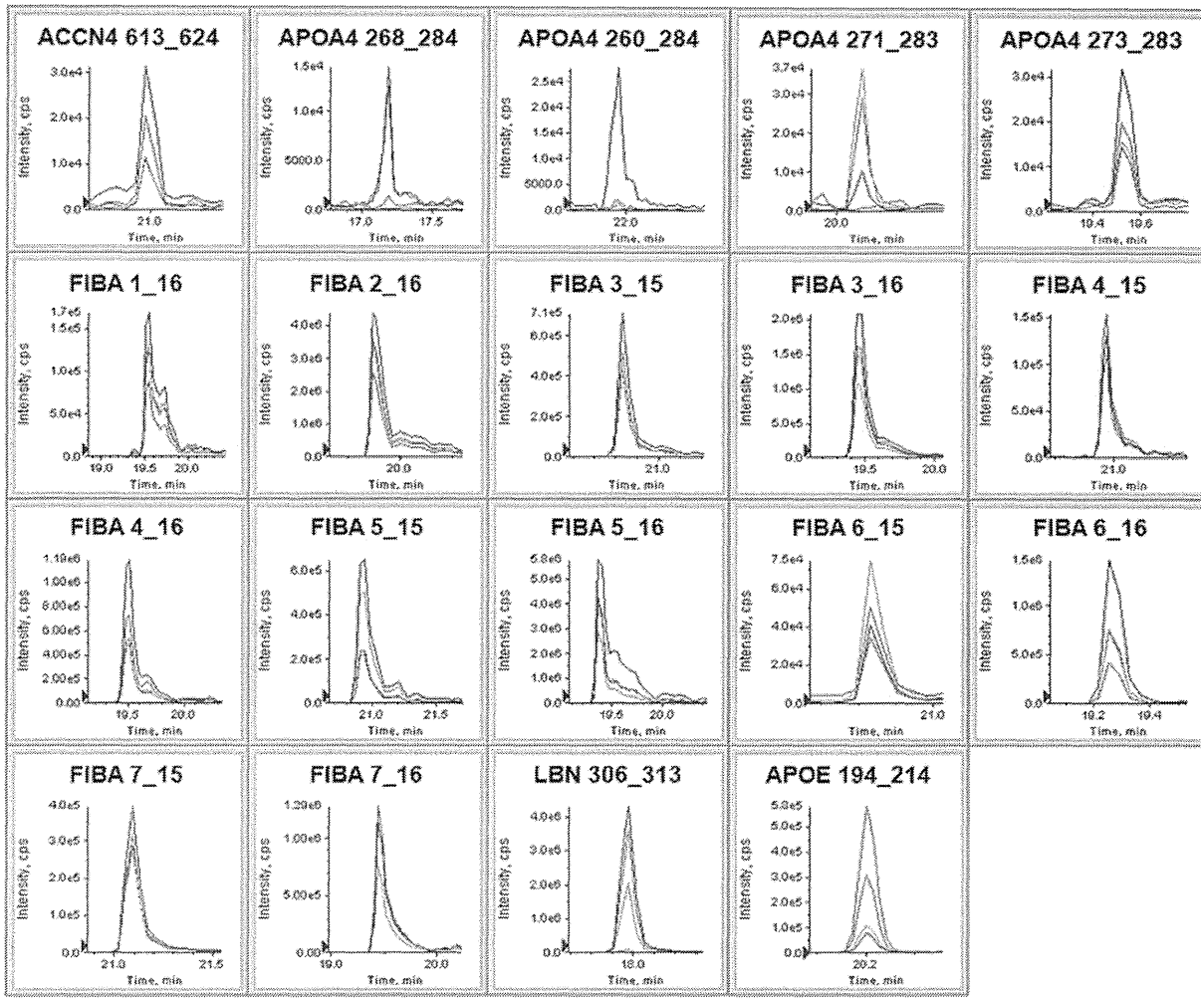
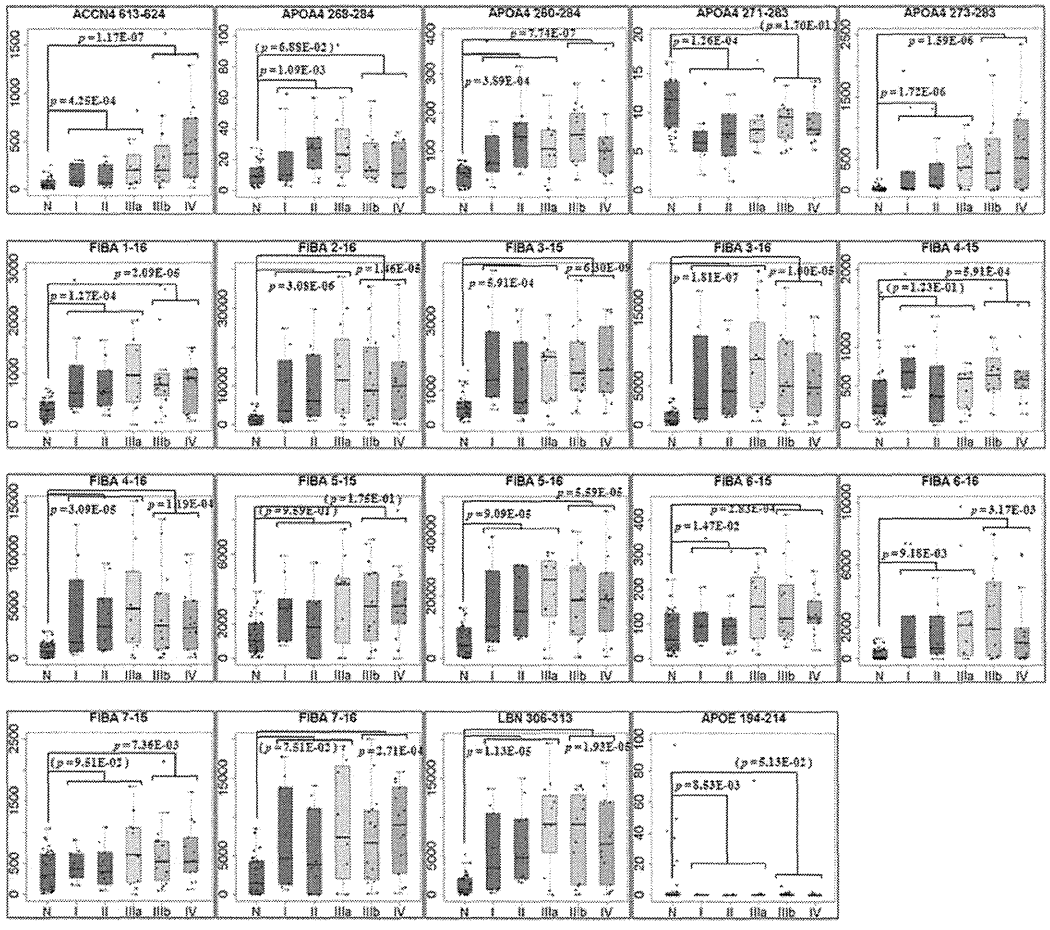


Figure 5. Selection and confirmation of the optimum MRM transitions for 19 candidates. Four pairs of precursor m/z and fragment m/z (Q1/Q3 channels) were set as MRM transitions for each peptide. The blue, red, green, or gray MRM chromatogram monitored the fragment ion which showed the 1st, 2nd, 3rd, or 4th most intense peaks in QSTAR-Elite LC/MS/MS analysis, respectively.
doi:10.1371/journal.pone.0018567.g005

analysis of ClinProt magnetic beads-purified samples, covered only limited spectra of serum peptidome. Most of studies utilizing ion-exchange selection or reversed phase extraction of peptidome on ProteinChip arrays [25,26,27] or magnetic beads [28,29] allowed at most 200 peak detections within the mass range 1,000 to 20,000. Meanwhile our peptidome profiling technology consisting of gel-filtration chromatography, custom-made high resolution C18 tip-column, QSTAR-Elite mass spectrometer, and Expressionist proteome server platform analysis enabled us to detect 12,396 non-redundant molecules with charge state of +1 to +10. The number of detected peaks here denoted the enormous advantage of our methodology for the analytical comprehensiveness compared to other existing methods. Although we focused on serum peptides involved in 3,537 clusters with charge stage of +2 to +6 in this study, 12,396 clusters might include non-peptide serum components such as metabolites, which should be also valuable for biomarker screening. Additionally, regarding the capacity of sample numbers to be analyzed simultaneously, the Expressionist server platform has a potential to handle a larger number of clinical samples. Because we in fact needed only less than an hour to process 92 LC/MS/MS data in the Refiner MS module (Fig. 1*A*), a comprehensive analysis of up to 1,000 cases would be feasible in a day. Hence our peptidome profiling technology provides the outstanding features of data comprehensiveness and quantitative performance, which absolutely fit the in-depth screening of novel biomarkers from clinical samples such as serum and plasma compared to previous technologies described above, whereas estimating actual dynamic range of detected peptide concentrations would be needed by,for instance, MRM-based absolute quantification analysis in the future. It could be tailored to many diagnostic and pharmaco-dynamic purposes as comprehensive interpretations of catalytic and metabolic activities in body fluids or tissues.

By using this technology, we finally identified 19 serum peptides as candidate lung cancer biomarkers (Table 1). The subsequent MRM-based validation experiments and t-test resulted in the confirmation of 12 candidates as reliable lung cancer biomarkers (Fig. 6*A*). Eight of them were fragments derived from fibrinopep-

A


ACCN4 613-624
APOA4 268-284
APOA4 260-284
APOA4 271-283
APOA4 273-283
FIBA 1-16
FIBA 2-16
FIBA 3-15
FIBA 3-16
FIBA 4-15
FIBA 4-16
FIBA 5-15
FIBA 5-16
FIBA 6-15
FIBA 6-16
FIBA 7-15
FIBA 7-16
LBN 306-313
APOE 194-214
N I II IIIa IIIb IV

B

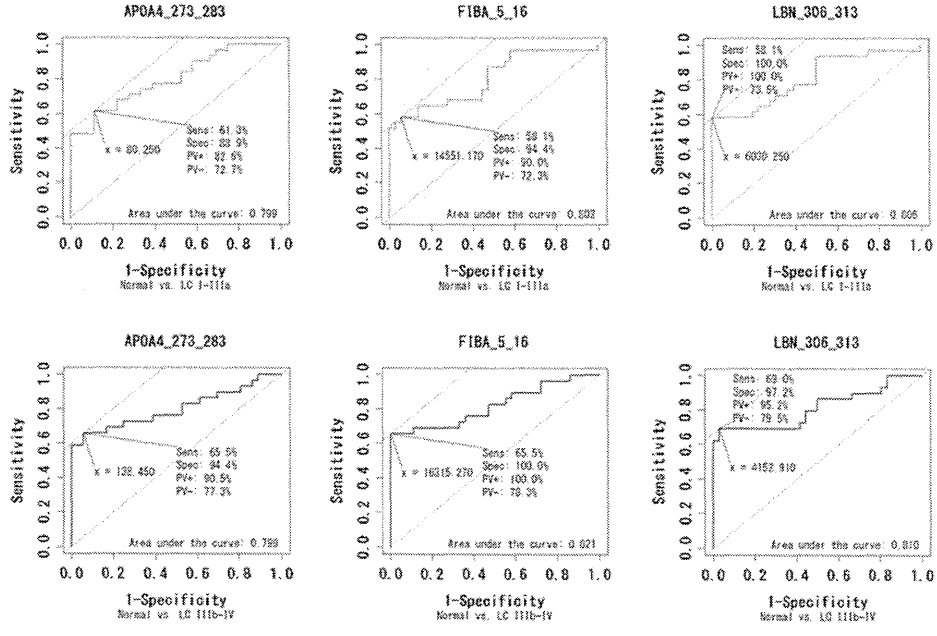

APOA4_273_283
FIBA_5_16
LBN_306_313
Sensitivity
1-Specificity
Area under the curve: 0.799
Area under the curve: 0.832
Area under the curve: 0.806
Area under the curve: 0.799
Area under the curve: 0.821
Area under the curve: 0.810
Normal vs. LC I-IIIa
Normal vs. LC IIIb-IV

Figure 6. Statistical assessment of MRM-based validation experiments. (*A*) Box plots representing the stage-dependent distributions of serum levels of the 19 candidate biomarkers. The p-values from t-test between "normal group (n = 36) and lung cancer stage-I, II, and IIIa (n = 30)" or "normal group (n = 36) and lung cancer stage-IIIb and IV (n = 30)" are shown. The p-values that did not show significant differences were provided in parentheses. N: normal group, I, II, IIIa, IIIb, and IV: lung cancer stage-I, II, IIIa, IIIb, and IV group, respectively. (*B*) ROC curves for APOA4 206–284, FIBA 2–16, and LBN 306–313 were depicted by R. The green or blue graph shows comparison of "normal group (n = 36) and lung cancer stage-I, II, and IIIa (n = 30)" or "normal group (n = 36) and lung cancer stage-IIIb and IV (n = 30)", respectively. The cut-off value was set at the point whose distance from the (sensitivity, specificity) = (1, 1) reached the minimum. The sensitivity (Sens), specificity (Spec), positive predictive value (PV+), negative predictive value (PV-), and area under the curve (AUC) were shown on each graph.
doi:10.1371/journal.pone.0018567.g006

tide A (FPA) which is N terminally cleaved product from fibrinogen α (FIBA). In fact, both our screening and validation results suggested that all of these eight FPA fragments were potential lung cancer-associated biomarkers showing the significant increase of concentrations in lung cancer patients' sera. However, since anomalous turnover of FPA was previously reported in several other diseases including gastric cancer [30], diabetic nephropathy [31], coronary heart disease [32], and others, these 8 FPA fragments could not be defined as lung cancer-specific biomarkers. The other two candidates were generated from apolipoprotein A-IV (APOA4). APOA4 protein itself was already identified as an up-regulated biomarker for ovarian cancer [33], whereas this was also known to be regulated by nutritional and metabolic stress [34]. But both quantitative information and physiological functions of endogenously-processed APOA4 peptides in human serum were still unknown. Interestingly, the APOA4 273–283 fragment demonstrated pathological stage-dependent up-regulation in lung cancer patients' sera, while the two-residue longer fragment APOA4 271–283 was significantly decreased in lung cancer samples (Fig. 6*A*). This indicates the existence of lung cancer-associated endo- or exopeptidases responsible for the cleavage at the C-terminus of APOA4 a.a. 272. Additional two candidate biomarkers, LBN 306–313 and ACCN4 613–624, derived from limbin (LBN) and amiloride-sensitive cation channel 4 (ACCN4) proteins, were reported as cellular membrane proteins. LBN is also known as Ellis-van Creveld syndrome 2 (EVC2) that is expressed in the heart, placenta, lung, liver, skeletal muscle, kidney and pancreas. Defects in LBN (EVC2) are a cause of acrofacial dysostosis Weyers type (WAD, also known as Curry-Hall syndrome) [35]. ACCN4 is a newly identified member of the acid-sensing ion channel family expressed in pituitary gland and weakly in brain [36]. Neither of them was detected in serum previously. Since our study provided the first evidence of LBN 306–313 and ACCN4 613–624 detection in human serum, further analysis of physiological functions and measurement in other diseases should be required for the proper use in clinical lung cancer diagnosis. Hence, the three candidate biomarkers illustrated in Figure 6*B* (APOA4 273–283, FIBA 5–16, and LBN 306–313) were individually considered as clinically useful biomarkers for both early detection and tumor staging of lung cancer, however, integrative measurement of biomarkers such as Figure 4*B* would provide more accurate diagnosis, that could be achievable by MRM-based diagnostic approaches in the future. Consequently the sensitivity of these biomarkers was higher than the currently-used screening biomarker CEA especially at even stage-I or II [8], indicating that new biomarkers addressed in this study had great potential to realize the early detection system for lung cancer. However further validation experiments using high risk groups of lung cancer as the controls (such as heavy smokers or COPD patients) will be necessary to prove the specificity and clinical usefulness of our biomarkers because more practical target population of the early diagnosis of lung cancer should be them rather than healthy individuals.

Finally we grasped the birds-eye view of human peptidome as a snapshot of the specific disease state. We are recently willing to use our peptidome profiling technology to establish an in-house quantitative serum/plasma peptidome database and contribute to the worldwide efforts such as Peptide Atlas (http://www.peptideatlas.org/). This framework would represent a new insight of protease/peptidase activities reflecting a clinical status at a specific time-point of disease and provide essential resources for next-generation extracorporeal diagnostic systems based on mass spectrometry. We therefore hope that researchers at global sites would utilize the peptidome profiling method addressed here and share data to construct mutually beneficial networks and databases which could contribute to the development of future diagnostic technologies worldwide.

Supporting Information

Figure S1 The bar charts illustrating the quantitative screening results for 19 candidates. The normalized peak intensities of 118 candidate biomarker peptides were calculated from 92 serum samples and displayed with bar charts.
(TIF)

Figure S2 MS/MS spectra used for the construction of MRM transitions and peptide identification. All MS/MS spectra were acquired with QSTAR-Elite mass spectrometer in the screening phase (the upper panels). The 1st, 2nd, 3rd, or 4th most intense peaks in each MS/MS spectrum were used for the optimization of MRM transitions (Fig. 5) The middle and the lower panels show the identified fragment ions in MASCOT database search. The ion scores and Expectation values were also indicated in the lower panels.
(TIF)

Figure S3 ROC curves for 19 lung cancer biomarker candidates were depicted by R. The green or blue graph shows comparison of "normal group (n = 36) and lung cancer stage-I, II, and IIIa (n = 30)" or "normal group (n = 36) and lung cancer stage-IIIb and IV (n = 30)", respectively. The cut-off value was set at the point whose distance from the (sensitivity, specificity) = (1, 1) reached the minimum. The sensitivity (Sens), specificity (Spec), positive predictive value (PV+), negative predictive value (PV-), and area under the curve (AUC) were shown on each graph.
(TIF)

Table S1
(DOC)

Table S2
(DOC)

Table S3
(DOC)

Table S4
(DOC)

Acknowledgments

We thank Ms. Mari Kikuchi at AB Sciex for technical assistances.

Author Contributions

Conceived and designed the experiments: KU. Performed the experiments: KU NS ST DK AT. Analyzed the data: KU NS ST AT HN. Contributed reagents/materials/analysis tools: AT YD NI NK KT TS MN T-AS YN HN. Wrote the paper: KU YN HN.

References

1. Parkin DM (2001) Global cancer statistics in the year 2000. Lancet Oncol 2: 533–543.
2. Samet JM, Avila-Tang E, Boffetta P, Hannan LM, Olivo-Marston S, et al. (2009) Lung cancer in never smokers: clinical epidemiology and environmental risk factors. Clin Cancer Res 15: 5626–5645.
3. Goldstraw P, Crowley J, Chansky K, Giroux DJ, Groome PA, et al. (2007) The IASLC Lung Cancer Staging Project: proposals for the revision of the TNM stage groupings in the forthcoming (seventh) edition of the TNM Classification of malignant tumours. J Thorac Oncol 2: 706–714.
4. Henschke CI, Yankelevitz DF, Libby DM, Pasmantier MW, Smith JP, et al. (2006) Survival of patients with stage I lung cancer detected on CT screening. N Engl J Med 355: 1763–1771.
5. Gail MH, Muenz L, McIntire KR, Radovich B, Braunstein G, et al. (1988) Multiple markers for lung cancer diagnosis: validation of models for localized lung cancer. J Natl Cancer Inst 80: 97–101.
6. McCarthy NJ, Swain SM (2001) Tumor markers: should we or shouldn't we? Cancer J 7: 175–177.
7. Brundage MD, Davies D, Mackillop WJ (2002) Prognostic factors in non-small cell lung cancer: a decade of progress. Chest 122: 1037–1057.
8. Sung HJ, Cho JY (2008) Biomarkers for the lung cancer diagnosis and their advances in proteomics. BMB Rep 41: 615–625.
9. Maurya P, Meleady P, Dowling P, Clynes M (2007) Proteomic approaches for serum biomarker discovery in cancer. Anticancer Res 27: 1247–1255.
10. Gamez-Pozo A, Sanchez-Navarro I, Nistal M, Calvo E, Madero R, et al. (2009) MALDI profiling of human lung cancer subtypes. PLoS One 4: e7731.
11. Wollscheid B, Bausch-Fluck D, Henderson C, O'Brien R, Bibel M, et al. (2009) Mass-spectrometric identification and relative quantification of N-linked cell surface glycoproteins. Nat Biotechnol 27: 378–386.
12. Schiess R, Wollscheid B, Aebersold R (2009) Targeted proteomic strategy for clinical biomarker discovery. Mol Oncol 3: 33–44.
13. Ueda K, Takami S, Saichi N, Daigo Y, Ishikawa N, et al. (2010) Development of serum glycoproteomic profiling technique; simultaneous identification of glycosylation sites and site-specific quantification of glycan structure changes. Mol Cell Proteomics 9: 1819–1828.
14. Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, et al. (2010) Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. PLoS One 5: e15003.
15. Villanueva J, Nazarian A, Lawlor K, Tempst P (2009) Monitoring peptidase activities in complex proteomes by MALDI-TOF mass spectrometry. Nat Protoc 4: 1167–1183.
16. Shen Y, Tolic N, Liu T, Zhao R, Petritis BO, et al. (2010) Blood peptidome-degradome profile of breast cancer. PLoS One 5: e13133.
17. Lopez-Otin C, Bond JS (2008) Proteases: multifunctional enzymes in life and disease. J Biol Chem 283: 30433–30437.
18. Overall CM, Blobel CP (2007) In search of partners: linking extracellular proteases to substrates. Nat Rev Mol Cell Biol 8: 245–257.
19. Palermo C, Joyce JA (2008) Cysteine cathepsin proteases as pharmacological targets in cancer. Trends Pharmacol Sci 29: 22–28.
20. Lopez-Otin C, Matrisian LM (2007) Emerging roles of proteases in tumour suppression. Nat Rev Cancer 7: 800–808.
21. Egeblad M, Werb Z (2002) New functions for the matrix metalloproteinases in cancer progression. Nat Rev Cancer 2: 161–174.
22. Albrethsen J, Bogebo R, Gammeltoft S, Olsen J, Winther B, et al. (2005) Upregulated expression of human neutrophil peptides 1, 2 and 3 (HNP 1-3) in colon cancer serum and tumours: a biomarker study. BMC Cancer 5: 8.
23. Sasaki K, Takahashi N, Satoh M, Yamasaki M, Minamino N (2010) A peptidomics strategy for discovering endogenous bioactive peptides. J Proteome Res 9: 5047–5052.
24. Anderson NL, Anderson NG, Pearson TW, Borchers CH, Paulovich AG, et al. (2009) A human proteome detection and quantitation project. Mol Cell Proteomics 8: 883–886.
25. Liu M, Li CF, Chen HS, Lin LQ, Zhang CP, et al. (2010) Differential expression of proteomics models of colorectal cancer, colorectal benign disease and healthy controls. Proteome Sci 8: 16.
26. Qiu FM, Yu JK, Chen YD, Jin QF, Sui MH, et al. (2009) Mining novel biomarkers for prognosis of gastric cancer with serum proteomics. J Exp Clin Cancer Res 28: 126.
27. Wang Q, Shen J, Li ZF, Jie JZ, Wang WY, et al. (2009) Limitations in SELDI-TOF MS whole serum proteomic profiling with IMAC surface to specifically detect colorectal cancer. BMC Cancer 9: 287.
28. Wong MY, Yu KO, Poon TC, Ang IL, Law MK, et al. (2010) A magnetic bead-based serum proteomic fingerprinting method for parallel analytical analysis and micropreparative purification. Electrophoresis 31: 1721–1730.
29. Huang Z, Shi Y, Cai B, Wang L, Wu Y, et al. (2009) MALDI-TOF MS combined with magnetic beads for detecting serum protein biomarkers and establishment of boosting decision tree model for diagnosis of systemic lupus erythematosus. Rheumatology (Oxford) 48: 626–631.
30. Ebert MP, Niemeyer D, Deininger SO, Wex T, Knippig C, et al. (2006) Identification and confirmation of increased fibrinopeptide a serum protein levels in gastric cancer sera by magnet bead assisted MALDI-TOF mass spectrometry. J Proteome Res 5: 2152–2158.
31. Gianazza E, Mainini V, Castoldi G, Chinello C, Zerbini G, et al. (2010) Different expression of fibrinopeptide A and related fragments in serum of type 1 diabetic patients with nephropathy. J Proteomics 73: 593–601.
32. Zito F, Drummond F, Bujac SR, Esnouf MP, Morrissey JH, et al. (2000) Epidemiological and genetic associations of activated factor XII concentration with factor VII activity, fibrinopeptide A concentration, and risk of coronary heart disease in men. Circulation 102: 2058–2062.
33. Dieplinger H, Ankerst DP, Burges A, Lenhard M, Lingenhel A, et al. (2009) Afamin and apolipoprotein A-IV: novel protein markers for ovarian cancer. Cancer Epidemiol Biomarkers Prev 18: 1127–1133.
34. Hanniman EA, Lambert G, Inoue Y, Gonzalez FJ, Sinal CJ (2006) Apolipoprotein A-IV is regulated by nutritional and metabolic stress: involvement of glucocorticoids, HNF-4 alpha, and PGC-1 alpha. J Lipid Res 47: 2503–2514.
35. Galdzicka M, Patnala S, Hirshman MG, Cai JF, Nitowsky H, et al. (2002) A new gene, EVC2, is mutated in Ellis-van Creveld syndrome. Mol Genet Metab 77: 291–295.
36. Grunder S, Geissler HS, Bassler EL, Ruppersberg JP (2000) A new member of acid-sensing ion channels from pituitary gland. Neuroreport 11: 1607–1611.

Toyama et al. Proteome Science 2011, 9:18
http://www.proteomesci.com/content/9/1/18

PROTEOME SCIENCE

RESEARCH Open Access

Deglycosylation and label-free quantitative LC-MALDI MS applied to efficient serum biomarker discovery of lung cancer

Atsuhiko Toyama[1,2,5], Hidewaki Nakagawa[2], Koichi Matsuda[1,3], Nobuhisa Ishikawa[4], Nobuoki Kohno[4], Yataro Daigo[1,3], Taka-Aki Sato[5], Yusuke Nakamura[3] and Koji Ueda[2*]

Abstract

Background: Serum is an ideal source of biomarker discovery and proteomic profiling studies are continuously pursued on serum samples. However, serum is featured by high level of protein glycosylations that often cause ionization suppression and confound accurate quantification analysis by mass spectrometry. Here we investigated the effect of N-glycan and sialic acid removal from serum proteins on the performance of label-free quantification results.

Results: Serum tryptic digests with or without deglycosylation treatment were analyzed by LC-MALDI MS and quantitatively compared on the Expressionist Refiner MS module. As a result, 345 out of 2,984 peaks (11.6%) showed the specific detection or the significantly improved intensities in deglycosylated serum samples ($P < 0.01$). We then applied this deglycosylation-based sample preparation to the identification of lung cancer biomarkers. In comparison between 10 healthy controls and 20 lung cancer patients, 40 peptides were identified to be differentially presented ($P < 0.01$). Their quantitative accuracies were further verified by multiple reaction monitoring. The result showed that deglycosylation was needed for the identification of some unique candidates, including previously unreported O-linked glycopeptide of complement component C9.

Conclusions: We demonstrated here that sample deglycosylation improves the quantitative performance of shotgun proteomics, which can be effectively applied to any samples with high glycoprotein contents.

Background

Since analyses of the serum proteome hold great promise for non-invasive detection of cancers and other diseases, various techniques for quantitative proteomic profiling have been developed to identify novel protein biomarkers [1,2]. These include labeling methods using stable isotopes such as ICAT (Isotope-coded affinity tags) [3], ^{13}CNBS (2-nitrobenzenesulfenyl) [4], SILAC (Stable isotope labeling with amino acids in cell culture) [5] and iTRAQ (Isobaric tags for relative and absolute quantification) [6]. Control and test samples are labeled with reagents with different isotopic composition of $^{12/13}$C, $^{14/15}$N and/or $^{16/18}$O, and detected simultaneously by mass spectrometry so that the intensities of isotopically resolved peak-pairs (or peak groups) represent the quantitative ratio of control and test samples. Although the precision of quantification is very high (typically 10% relative standard deviation) [7] because of the identical separation and detection, isotopic labeling limits the number of samples to be directly compared, which makes it unsuitable for analysis of a large number of clinical samples needed for biomarker discoveries. In contrast, label-free quantification methods deal with independently-acquired mass spectrometry data from essentially unlimited number of samples. Quantification based on ion intensities (extracted ion chromatograms) is known to have at least three orders of linear dynamic range [8,9], and can potentially cover wide proteome in complex samples such as serum. It is advantageous that label-free systems do not involve sample mixing prior to detection because target proteins that are only presented in test samples are effectively diluted by mixing with

* Correspondence: k-ueda@riken.jp
[2]Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, Tsurumiku-Suehirocho1-7-22, Yokohama, Japan
Full list of author information is available at the end of the article

control samples, rendering them more difficult to detect. Therefore, label-free quantification has emerged as an alternative approach for biomarker discovery, which requires sufficient sample sizes to overcome individual variability in clinical samples and technical bias in sample preparation and analysis batch [10].

High content of glycoproteins is another feature of serum that should be considered when performing quantitative proteomic analysis. Recent advances in glycoproteomic analysis using mass spectrometry have made it possible to exhaustively identify N-linked glycopeptides and their glycosylation sites [11,12]. These techniques involve enrichment of glycopeptides followed by enzymatic cleavage of N-glycans in order for efficient mass spectrometric analysis. Deglycosylation can be coupled with the incorporation of ^{18}O stable isotope resulting in +3 Da mass shift of asparagine residues, which allows deterministic identification of glycosylation sites [13]. As these studies indicated, most of serum proteins are heavily glycosylated, however, potential effect of glycopeptides on ionization suppression of coexisting peptides had been overlooked. Glycopeptides carry large, hydrophilic carbohydrate moieties, which can cause substantial ionization suppression [14], hampering precise quantification particularly at low-concentration range.

To elucidate the extent to which the ionization of peptides is interfered by glycopeptides, the first part of this study describes the changes caused by serum deglycosylation in the MS peak profiles obtained by label-free shotgun proteomic analysis. Having shown the utility of deglycosylation, we next applied this principle to the biomarker screening of lung cancer. Because of the vast number of incidence and high mortality rate, lung cancer is considered to be one of the highest priorities for biomarker development. Using the serum samples of 10 healthy control, 10 early-stage (Stage I-II) cases and 10 advanced stage (Stages IIIb-IV) cases, we conducted a study that uniquely combined sample deglycosylation, label-free MALDI and multiple reaction monitoring (MRM) mass spectrometry [15] and verified the result by western blotting. Taken together, we show herein that enzymatic removal of carbohydrate moieties results in recognizable improvement in the data quality of shotgun proteomics in terms of sensitivity and reproducibility, which should facilitate quantitative analysis of glycoprotein-rich samples.

Results

Quantitative MS analysis of deglycosylated serum

In order to examine the effect of deglycosylation on data content and quality, tryptic digest of serum proteins was prepared with or without the removal of N-glycans and sialic acids (n = 6), and analyzed on the LC-MALDI label-free quantification platform as summarized in Figure 1. Figure 2 shows the 2-dimensional MS signal intensity maps after data processing by Expressionist Refiner MS (Genedata). In total, 27,357 single peaks were detected, comprising 4,444 groups of peaks each representing unique peptide species (termed "peak clusters"). Examples of the differences in the peak profile between deglycosylated and untreated samples are illustrated in the 2D map. The broken box in Figure 2a shows emergence of prominent peaks, and the expanded views in Figure 2c and 2d (arrows) illustrate the loss of intact glycopeptide peaks. The peak clusters subjected to comparative analysis were selected by eliminating peaks that were not presented in all of the 6 replicate runs. This filtering was performed separately for deglycosylated and untreated samples, yielding 2,984 and 2,610 peak clusters, respectively. Deglycosylation thus resulted in 14.3% increase in the number of reproducibly detectable peaks. The signal intensities of these peak profiles were then directly compared by t-test as summarized in the volcano plot (Figure 3). 221 peak clusters displayed altered intensities, of which 188 were higher in deglycosylated samples and 33 peak clusters vice versa ($P < 0.01$). Present/absent search revealed that 157 peak clusters were found specifically in deglycosylated samples (Figure S-1, Additional File 1). Combined, 345 of 2,984 peak clusters (11.6%) detected in deglycosylated samples were enhanced by deglycosylation, as opposed to 33 of 2,610 peak clusters (1.3%) in untreated serum.

The composition of the peak profiles was analyzed by exhaustive MS/MS analysis, resulting in 1,735 peptide identifications, including 153 originally-glycosylated peptides that were identified with ^{18}O-incorporated N-glycosylation sites (Table S-2, Additional File 1). By matching the peptide IDs to the statistical analysis, we found that 97 of 345 peak clusters overrepresented in deglycosylated serum were originally-glycosylated peptides. These peak clusters emerged as the product of deglycosylation and accounted for most of the major fold-changes in the volcano plot as indicated with squares in Figure 3. The remaining 248 peak clusters, as well as the 33 peak clusters that diminished by deglycosylation, were non-glycopeptides. There were no apparent features common to these peptides. Rather, signal intensities were affected by the local peptide composition, such that alleviation of suppressive analytes led to signal enhancement and emergence of competing analytes led to diminished signal. The result shown here, that the number of enhanced peaks was by far greater than those that diminished, suggest that the ionization suppression effect exerted by intact glycopeptides was more extensive than the deglycosylated counterpart.

To investigate whether deglycosylation had any effect on the quality of quantitative data, reproducibility of six

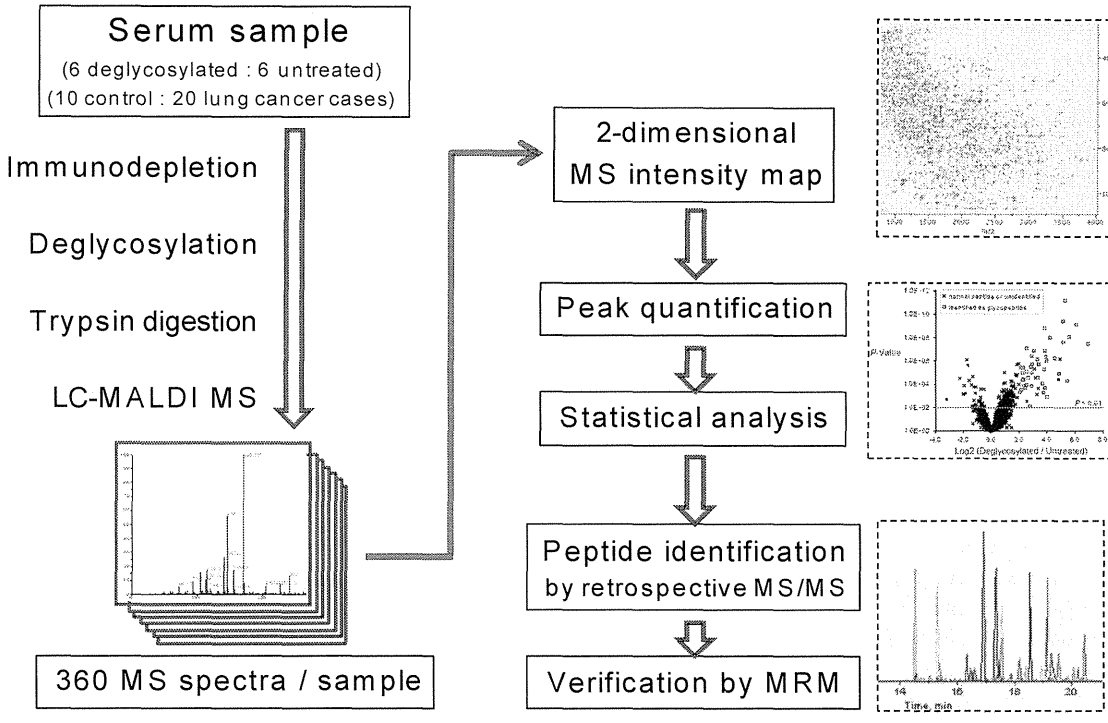


Figure 1 Schematic of the LC-MALDI label-free quantification with deglycosylation.

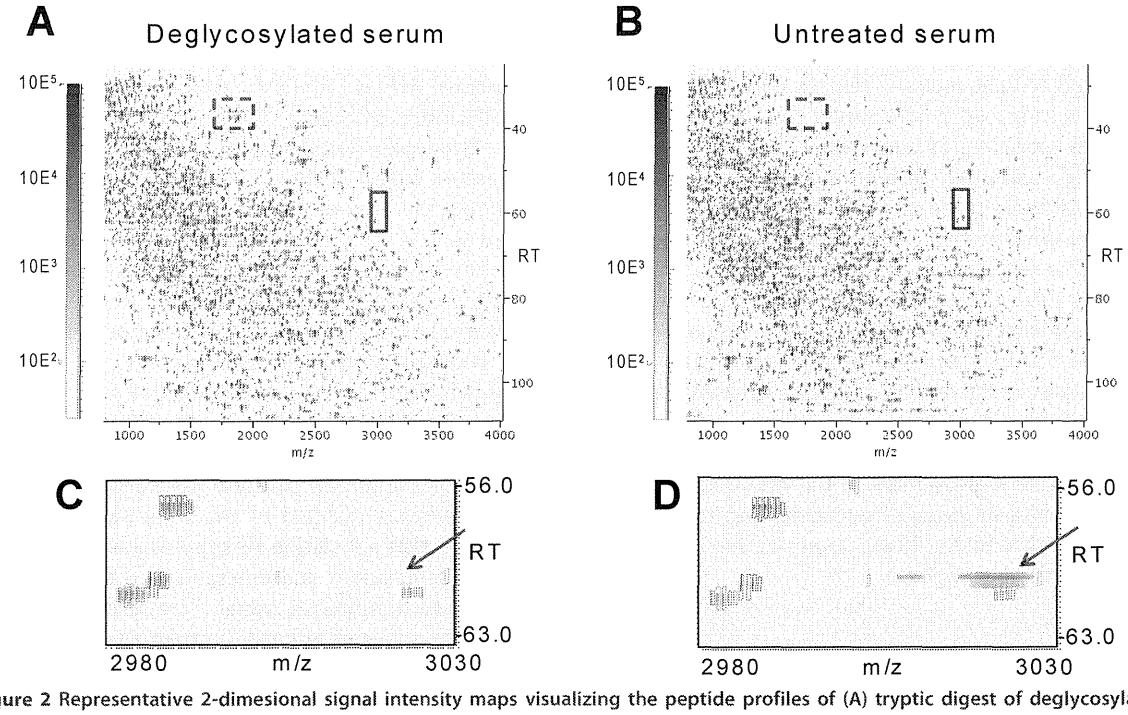


Figure 2 Representative 2-dimesional signal intensity maps visualizing the peptide profiles of (A) tryptic digest of deglycosylated serum, and (B) untreated serum, within the range of m/z 800-4000, and 25-110 minutes of retention time. An example of obvious peak appearance in deglycosylated serum is indicated by broken boxes. (C) and (D) display the expanded views of the closed box in deglycosylated sample and untreated sample, respectively. The rectangles outlining the MS signal represent peak detection and their color coordination represents peak clustering.

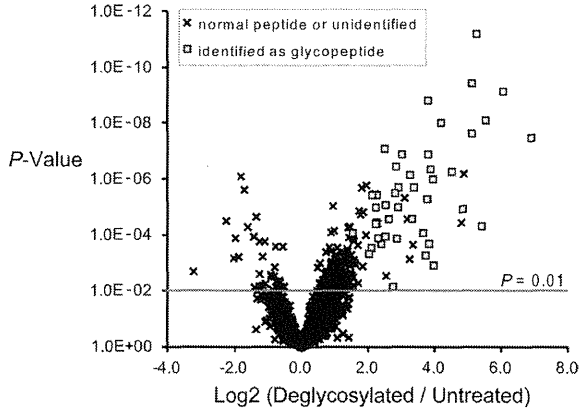


Figure 3 Volcano plot showing the difference in the peptide profiles of deglycosylated and untreated serum sample. Crosses represent peak clusters identified as normal peptides or not identified in the analysis, and squares represent deglycosylated peptides identified with +3 Da modification.

replicating runs was evaluated by calculating the correlation coefficients of the intensities of 2,356 clusters that were detected in all 12 runs. The correlation coefficients were calculated for all 15 possible combinations of 6 individual replicates within the experimental group. They ranged from 0.895 to 0.916 with mean value 0.899 (n = 15) in the untreated group and from 0.898 to 0.944 with mean value 0.914 (n = 15) in the deglycosylated group, suggesting that deglycosylation generally results in more reproducible data than the untreated (P < 0.002, t-test). Furthermore, coefficients of variation (CV) of the signal intensities for the same 2,356 peak clusters were calculated and compared between the two experimental groups (Figure 4). The frequency distribution of CV showed that deglycosylationtreatement resulted in slightly lowering the median CV from 34% to 31%. It is

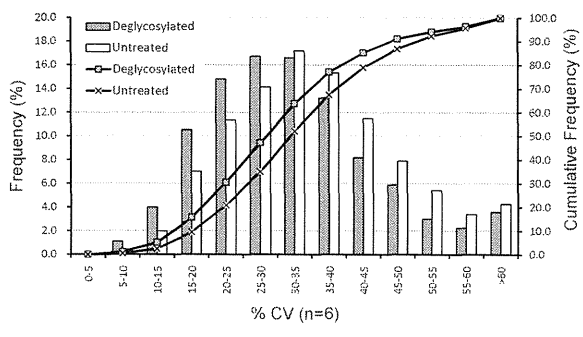


Figure 4 A frequency distribution plot of the coefficients of variations (CV) calculated for each of the 2,356 common peak clusters. Bars represent the frequency percentage and lines represent the cumulative frequency. Filled bars and boxes, replicates with deglycosylation; blank bars and crosses, untreated replicates (n = 6).

clear from these data that the effect of deglycosylation well extends to non-glycosylated peptides and enhances sensitivity and reproducibility.

Lung cancer biomarker screening

Deglycosylation-LC-MALDI platform was applied to the profiling of serum proteins for lung cancer biomarker screening. Control and lung cancer sera were immunodepleted, deglycosylated, purified on SDS-PAGE, in-gel digested and analyzed by LC-MALDI in the same way as described above. 30 mass chromatograms were processed simultaneously on Expressionist Refiner MS, and 23,453 peaks were detected, which were grouped into 6,186 peak clusters. Then we compared the two experimental groups, control (n = 10) and lung cancer (n = 20) by t-test and identified 63 peak clusters showing P < 0.01. In addition, 13 peak clusters were selected that have valid value of less than or equal to 2 in one experimental group and greater than 50% valid value proportion in the other. The total of 76 clusters was manually inspected and retrospective MS/MS was acquired by selecting the highest-expressing sample and the fraction spot of maximum elution for optimum MS/MS acquisition efficiency. As a result, 25 candidate proteins, comprising 40 candidate peptides, were identified (Table 1).

Next, these candidates were verified at peptide level by the MRM-based relative quantification analysis using the same preparation batch of serum tryptic digest as used for MALDI MS analysis as it is previously reported that LC-MALDI measurement is the most significant source of technical variability above sample preparation [16] (see Table S-3, Additional File 1 for the list of MRM transitions). This strategy was aimed at eliminating false-positive results from the long list of candidates and facilitating the selection of appropriate target for validation study. Of the 23 peptides that we found working MRM transitions, 11 peptides showed significant correlation (P < 0.05 by Pearson's correlation coefficient) in results obtained by MRM and MALDI MS, and 10 peptides (6 proteins) fulfilled P < 0.05 in both analyses. These peptide candidates were derived from ceruloplasmin, complement C3, complement component C9, inter-alpha-trypsin inhibitor heavy chain H3, inter-alpha-trypsin inhibitor heavy chain H4 and kininogen-1. The quantification results were summarized as dot-plots in Figure 5 to illustrate the potential performance of the 10 peptides as biomarker candidates. Raw data was normalized to the total detection and the samples were grouped as normal control, early stage (Stage I/II) and advanced stage (stage IIIb/IV) lung cancer cases. As the P-values indicate, most of the candidates showed significant response to cancer state even at early stages.

Further verification was performed by western blotting of complement C3 protein. Interestingly, expression of

Table 1 List of lung cancer biomarker candidates screened by label-free MALDI MS and their verification result on MRM.

			MALDI MS		MRM		
Uniprot ID	Protein Name	Peptide Sequence	t-test †	Log2 (LC/Control)	t-test ‡	Log2 (LC/Control)	Correlation Coefficient
P04217	Alpha-1B-glycoprotein	CEGPIPDVTFELLREGETKAVK	-	0.67	-	-	-
		FALVREDR	3.8E-03	-0.38	0.13	0.24	0.315
P01011	Alpha-1-antichymotrypsin	EQLSLLDRFTEDAK	1.4E-03	0.80	0.077	0.33	0.106
		FTEDAKRLYGSEAFATDFQDSAAAK	4.6E-03	0.58	-	-	-
		PQDTHQSR	5.6E-03	0.84	-	-	-
P43652	Afamin	DGLKYHYLIR	1.1E-03	-0.93	-	-	-
P02746	Complement C1q subcomponent subunit B	GNLCVNLMR	6.5E-03	0.87	-	-	-
P00736	Complement C1r subcomponent	CLPVCGKPVNPVEQR	-	0.42	-	-	-
		DYFIATCK	6.9E-03	0.49	0.15	0.14	-0.238
P00450	Ceruloplasmin	YTVNQCR	0.017	1.1	3.1E-04	0.29	0.411*
P10909	Clusterin	YVNKEIQNAVNGVK	0.01	0.33	0.19	-0.87	-0.015
P06681	Complement C2	TAVDHIREILNINQK	2.7E-03	1.4	0.051	0.23	N/A
P01024	Complement C3	AGDFLEANYMNLQR	5.9E-05	-1.4	3.8E-04	-1.3	0.690**
		ILLQGTPVAQMTEDAVDAER	2.1E-04	-1.3	0.019	-0.87	0.448*
		KGYTQQLAFR	2.8E-03	-1.1	3.9E-05	-1.3	0.800**
		QPSSAFAAFVKR	6.7E-03	-1.3	2.0E-03	0.24	-0.455
		WLNEQR	2.5E-03	-1.0	7.8E-06	-1.5	0.521**
P01031	Complement C5	FWKDNLQHKDSSVPNTGTAR	0.012	0.67	-	-	-
		TLRVVPEGVKR	0.066	0.80	-	-	-
P13671	Complement component C6	IEEADCKNKFR	0.011	1.2	-	-	-
P10643	Complement component C7	VFSGDGKDFYR	9.5E-03	0.81	-	-	-
P02748	Complement Component C9	FTPTETNKAEQCCEETASSISLHGK	2.4E-04	1.4	6.1E-03	0.49	0.472*
		QYTgTSYDPELTESSGSASHIDCR	1.9E-03	1.1	3.8E-03	0.57	0.769**
P22792	Carboxypeptidase N subunit 2	SQCTYSNPEGTWLACDQAQCR	1.4E-03	0.56	0.062	0.16	0.372
P00748	Coagulation factor XII	CTHKGRPGPQPWCATTPNFDQDQR	4.3E-03	1.2	-	-	-
Q9UGM5	Fetuin-B	MSPPQLALNPSALLSR	3.0E-03	0.62	0.03	0.58	0.189
P02751	Fibronectin	AQITGYR	1.1E-03	-0.47	0.077	-0.18	0.774**
		GFNCESKPEAEETCFDKYTGNTYR	1.9E-03	-0.69	0.42	-0.11	0.146
		IGFKLGVRPSQGGEAPR	5.7E-03	-0.90	-	-	-
P26927	Hepatocyte growth factor-like protein	RVDRLDQR	6.1E-03	-0.57	0.013	0.38	-0.128
P18065	Insulin-like growth factor-binding protein 2	LAACGPPPVAPPAAVAAVAGGAR	2.5E-03	0.67	-	-	-
Q06033	Inter-alpha-trypsin inhibitor heavy chain H3	EHLVQATPENLQEAR	4.6E-03	0.97	3.6E-03	0.65	0.578**
Q14624	Inter-alpha-trypsin inhibitor heavy chain H4	EKNGIDIYSLTVDSR	4.7E-04	0.65	0.062	0.40	0.368
		ETLFSVMPGLK	-	0.09	0.01	0.42	0.331
		MNFRPGVLSSR	9.8E-03	0.40	1.0E-03	0.63	0.413*
		SPEQQETVLDGNLIIRYDVDR	0.013	1.2	-	-	-
P03952	Plasma kallikrein	CQFFTYSLLPEDCKEEKCK	-	-0.32	-	-	-
P01042	Kininogen-1	RPPGFSPF	2.0E-04	-2.0	5.1E-04	-1.74	0.876**
P27918	Properdin	TCNHPVPQHGGPFCAGDATR	7.2E-04	-0.52	-	-	-
Q13103	Secreted phosphoprotein 24	DSGEDPATCAFQR	3.7E-03	-0.53	-	-	-

†: "-" indicates candidate screened by present/absent search,

‡: "-" indicates no MRM data obtained, *: $P < 0.05$, **: $P < 0.01$,

g: indicates site of O-linked glycosylation.

N/A: valid value too small for calculation of correlation threshold.

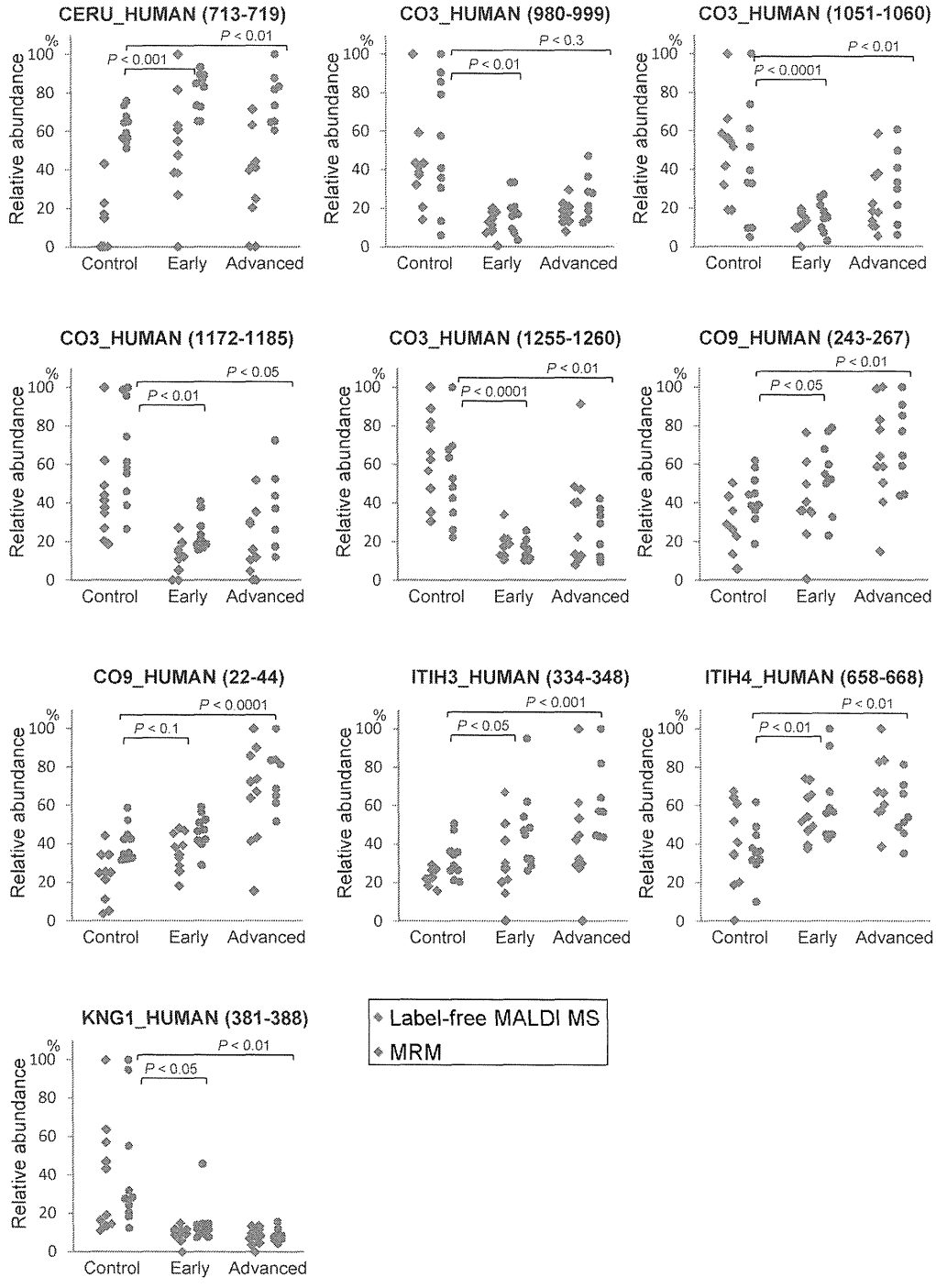


Figure 5 Dot plots showing the relative quantifications in 10 controls, 10 early stage and 10 advanced stage lung cancer cases of the 10 candidate peptides, subtitled with SwissProt IDs and amino acid numbers in parenthesis. Diamond dots aligned to the left represent quantification by MALDI MS, and circular dots aligned to the right by MRM, normalized and plotted on the same scale. The *P*-values presented were calculated using the results of MRM.

39 kDa subunit of C3 protein was strongly suppressed in early-stage patients (Figure 6A). Since this subunit was also detected by the monoclonal antibody raised against C3d fragment (Figure S-2 Additional File 1), the 39 kDa subunit was assigned to be C3dg fragment [17]. This fragment encompasses all of the four C3 peptide candidates identified by LC-MALDI screening, and semi-quantitative analysis of the immunoblot (Figure 6B) almost exactly reproduced the screening result. Moreover, C3dg fragment was shown to escape immunodepletion by MARS-Hu14 column (Figure S-2). Therefore, the apparent difference in complement C3 abundance observed in the screening was reflecting the degree of proteolytic degradation associated with lung cancer.

Finally, the benefit of deglycosylation in this biomarker screening was assessed by mining the candidate peptides from the control experiment (comparing deglycosylated and untreated serum) and verifying whether or not deglycosylation facilitated biomarker identification. Figure 7 shows the levels of complement component C9 peptides in the control experiment, which were clearly overrepresented by deglycosylation. Notably, the signal intensity of non-glycopeptide (243-267) was doubled, reiterating the observation that non-glycopeptide was also subject of signal enhancement by deglycosylation. Moreover, the signal intensity of peptide (22-44) was increased by 10-fold. Since this peptide was identified as O-linked glycopeptide (attachment of N-acetylgalactosamine and galactose as predicted from the m/z shift), the addition of sialidase probably contributed to reduction of glycan complexity and increased ionization efficiency. The sialylated counterpart was not detected in the untreated control.

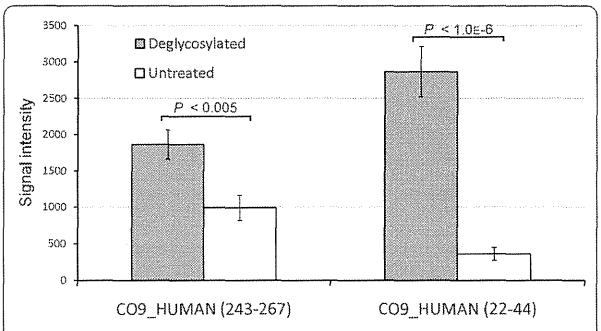


Figure 7 Bar chart representing the signal intensities of complement component C9 peptides detected from the tryptic digest of deglycosylated serum (filled bars) and untreated serum (open bars), n = 6. The numbers in parenthesis are amino acid numbers: 243-267 corresponds to $_{243}$FTPTETNKAEQCCEETASSISLHG$_{267}$K and 22-44 corresponds to $_{22}$ QYTgTSYDPELTESSGSASHIDC$_{44}$R, where "g" represents the site of O-linked glycan attachment. Error bars are one standard deviation.

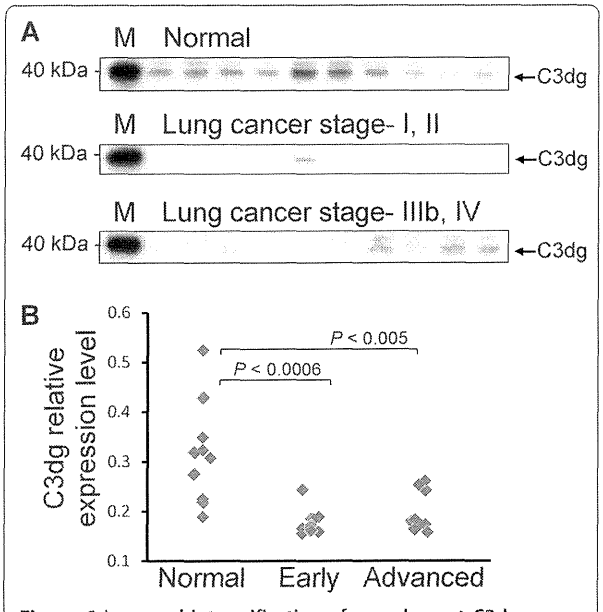


Figure 6 Immunoblot verification of complement C3dg serum levels in the study subjects. (A) Detection of 39 kDa C3dg fragment from 0.02 μL of crude serum using anti-C3 antibody. Lanes represent 10 subjects each of normal control, early stage lung cancer (stage-I/II) and advanced stage lung cancer (stage-IIIb/IV). (B) Dot plot of the data shown in (A), quantified as the ratio of C3dg detection level to the normalization lane M, a 40 kDa size marker that has cross-reactivity with the secondary antibody.

Discussion

The aim of this study was to introduce deglycosylation as a facile and universal sample preparation step in shotgun proteomic analysis because we expected that the undermining effect exerted by the large proportion of glycopeptides was more extensive than previously considered. Therefore, for the first time, we have performed a direct comparison between deglycosylated and untreated serum samples, and showed that deglycosylation actually results in improvement of the shotgun peak profile. This was observed in terms of both acquisition of unique peaks and enhancement of existing peaks.

The acquisition of unique peaks by deglycosylation was expected, as it is widely recognized that deglycosylated peptides have much higher ionization efficiency than the corresponding glycopeptides [18,19]. It is well established that deglycosylation results in the detection of "new" peaks and increases the depth of information acquired by shotgun analysis [20,21]. However, this study revisited the same phenomenon from different perspective. Our novel finding was that signal enhancement by deglycosylation extended up to 8% of existing non-glycopeptide peaks (Figure 3). This result strongly suggested that there was notable ionization suppression effect exerted by glycopeptides on co-existing analytes,

and that signal intensities were enhanced through alleviation of suppression in deglycosylated sample. Incorporation of ^{18}O into the site of deglycosylation by $H_2{}^{18}O$ was utilized to facilitate the identification of glycopeptides. The potential pitfalls of this approach as pointed out by Angel et al. [22] was circumvented by performing deglycosylation before tryptic digestion, and our data showed that 73% of identified glycopeptides fulfilled the biological consensus NxT/S. This strategy helped to confirm that only a small proportion of peaks that were enhanced by deglycosylation were actually glycosylated. We further demonstrated the evidence that deglycosylation improves the reproducibility of replicate measurements to some extent (Figure 4). The level of technical variability, median CV of 31%, was comparable to previously reported label-free quantification methods based on LC-ESI-MS [8] or LC-MALDI MS [16].

Unfortunately, previous studies on ionization suppression effects had mainly focused on selective detection of glycopeptides in mixtures of peptides [23], and cannot explain the phenomenon addressed here. Separate investigation needs to be conducted in order to elucidate the mechanism and the extent to which glycopeptides interfere with the ionization of co-existing peptides.

Serum deglycosylation coupled with label-free quantification was applied to biomarker screening for lung cancer and led to the identification of unique biomarker candidates including the fragmentation state of complement C3, complement component C9 peptide with novel O-linked carbohydrate and Kininogen-1 peptide with C-terminal Phe [24]. The benefit of deglycosylation in the biomarker screening was demonstrated by enhanced detectability of the complement component C9 peptides, particularly for O-linked glycopeptide whose intact form with sialic acid was hardly detected. Since many O-linked glycans contain sialic acid, the data presented here demonstrates a high potential of sialidase usage for comprehensive analysis of O-linked glycans.

As with other label-free quantitative proteomics, more proteins were quantified by single peptide than those quantified by multiple peptides [25] due to the high technical variability associated with label-free shotgun analysis. Therefore, we employed MRM to complement the screening by MALDI MS with the aim of eliminating false-positive results. This approach successfully ruled out many candidates while retaining confident candidates, as verified by western blotting experiment of complement C3.

Complement C3 is the major component of the classical complement pathway. Upon antigenic stimulation, C3 convertase cleaves C3 into C3a and C3b, which subsequently triggers reaction cascade leading to the formation of membrane attack complex, by either the classical or alternative pathway [17]. C3a contains a multiply interacting motif known as anaphylatoxin [26]. Recently, a number of proteomic studies have identified C3a as biomarker candidates for colon cancer [27], chronic hepatitis C and related hepatocellular carcinoma [28], insulin resistance/type-2 diabetes [29] and chronic lymphoid malignancies [30]. While upregulation of C3a is widely reported and interpreted as an indicator of primary inflammatory response, there is limited association reported between C3dg and cancer. C3dg is known as the ligand of complement receptor 2 [31] and may be critically involved in cancer recognition. The mechanism by which the production of C3dg is suppressed in response to the onset of lung cancer requires further investigation.

In the Expressionist label-free quantification platform employed here, peptide peak clusters are defined by retention time-m/z coordinate on the 2D-map, enabling quantitative analysis without MS/MS information that were essential in other platforms [32-35]. The feature that provides the ground for this concept is perfect alignment of mass chromatograms in the retention time dimension because slight drift is unavoidable even in well-optimized separation system. In this respect, Expressionist demonstrated spectacular computational strength. The range of retention time drift in the 30 LC-MALDI analyses performed in this study was from -5 minutes to +5 minutes, a maximum of 10 minutes deviation, but the software was still capable of good alignment without any obvious retention time mismatch (data not shown). Therefore, variation in experimental conditions, such as changing the analytical column lot, should easily be tolerated. This feature enables integration of several, even retrospective, analyses, which is needed for the continuous pursuit for biomarker identification and validation.

Moreover, being a server-based module, Expressionist has greater data processing capability than other standalone software. This is another advantageous feature because, in general, attempts to increase proteome coverage involve vast increase in data amount, whether it be a multi-dimensional fractionation strategy [36,37] or an extremely long gradient separation [38]. Importantly, considering the fact that low-abundance analytes are more prone to ionization suppression [39], we speculate that the benefits of sample deglycosylation we addressed here would take greater effect with increasing dynamic range of detection. Such in-depth label-free analysis is currently not available, however, we demonstrated herein that current technology is already capable of large-scale label-free analysis, and we addressed its potentiality as a biomarker discovery platform. Taken together, we believe that sample deglycosylation will prove to be a valuable sample preparation protocol in

shotgun proteomic analysis in near future for analyzing glycoprotein-rich samples.

Conclusions

The studies described herein demonstrated that serum deglycosylation has positive effect on both data content and reproducibility through production of deglycosylated peptides and possibly through alleviation of ionization suppression by intact glycopeptides. The results therefore suggested the role of deglycosylation as a simple, indispensible method to improve the general performance of label-free quantification. Its first application to serum proteomic profiling by label-free LC-MALDI MS demonstrated that this strategy could lead to the identification of unique candidates, which could be effectively applied to any samples with high glycoprotein contents, such as other clinical body fluids, membrane proteomics or secretome analysis.

Methods

Reagents

Trizma base pH 8.3, iodoacetamide, ammonium bicarbonate, ammonium citrate, formic acid were purchased from Sigma (Saint Louis, MO). PlusOne grade SDS and dithiothreitol (DTT) were purchased from GE Healthcare (Uppsala, Sweden). CHAPS was purchased from Chemical Dojin (Kumamoto, Japan). N-glycosidase F was purchased from Roche (Basel, Switzerland). α2-3,6,8,9-neuraminidase was purchased from Merck (Darmstadt, Germany). Trypsin Gold was purchased from Promega (Madison, WI). $H_2{}^{18}O$ was supplied from Cambridge Isotope Laboratories Inc. (Andover, MA). Alpha-cyano-4-hydroxycinnamic acid (CHCA) was purchased from Shimadzu-GLC (Kyoto, Japan). LC/MS grade acetonitrile and 25% trifluoroacetic acid (TFA) were purchased from Wako Pure Chemicals (Osaka, Japan).

Serum Samples

Archived human serum samples were obtained with informed consent from 20 patients with lung adenocarcinoma and at Hiroshima University Hospital. Serum samples as normal controls were also obtained with informed consent from 13 healthy volunteers who received medical checkup at Hiroshima University Hospital (Table S-1, Additional File 1). Serum was collected using standard protocol from whole blood by centrifugation at 1500 × g for 10 min and stored at -150°C. This study was approved by individual institutional ethical committees.

Immunodepletion

20 μL serum aliquots obtained from healthy volunteer were subjected to Multiple Affinity Removal System (Hu-14, 4.6 mm × 100 mm, Agilent Technologies, Santa Clara, CA) according to the manufacturer's protocol using a conventional HPLC system (Shimadzu Corp., Kyoto, Japan). The flow-through fraction was desalted with a protein separation column (mRP-C18, 4.6 mm × 50 mm, Agilent Technologies). Desalted serum proteins were dried with a SpeedVac evaporator.

Deglycosylation

All solutions in the following deglycosylation step were freshly prepared with $H_2{}^{18}O$. Protein aliquots were dissolved in 12.5 μL of 2% SDS, 20 mM DTT, 20 mM Trizma-base pH 8.3 and heated to 100°C for 5 minutes. After cooling, 25 μL of 10% CHAPS, 83.4 μL 20 mM Trizma-base pH 8.3, 1 μL N-glycosidase F and 0.6 μL α2-3,6,8,9-neuraminidase were added in the written order with thorough mixing. $H_2{}^{18}O$ was added in place of the enzymes for "untreated" samples. The reaction mixture was incubated at 37°C overnight.

Tryptic digestion

Deglycosylated proteins were reduced by 10 mM DTT and incubated at 56°C for 15 minutes, followed by alkylation by 50 mM iodoacetamide at ambient temperature for 45 minutes in dark. 20% of the total reaction mixture was purified by SDS-PAGE, applying voltage until all of the proteins had entered the separating gel. Whole lanes were cut out and subjected to in-gel tryptic digestion. Briefly, gel slices were cut into small pieces and were washed 3 times in 30% acetonitrile 50 mM ammonium bicarbonate before digesting with 200 ng of trypsin in 100 μL 50 mM ammonium bicarbonate at 37°C overnight. Peptides were extracted by 2 rounds of 50% acetonitrile and 100% acetonitrile washes. Recovered peptides were dried in SpeedVac and reconstituted in 10 μL of 2% acetonitrile 0.1% TFA for LC-MALDI analysis.

LC-MALDI Analysis

Serum tryptic digest with or without deglycosylation was separated using DiNa nano-HPLC system (KYA Technologies, Tokyo, Japan). Solvent A was 2% acetonitrile and 0.1% TFA in water and solvent B was 70% acetonitrile and 0.1% TFA in water. 2 μL sample, a final amount equivalent to 0.4 μL serum, was injected onto a trap column (L-column ODS, 5 μm, 0.3 × 5 mm, CERI, Saitama, Japan) and loaded by 8 μL/min flow of solvent A. At 5 min, valve was switched and the peptides were separated by an in-house packed analytical column (L-column ODS, 3 μm, in 0.1 mm × 200 mm capillary) at 200 nL/min flow rate using the following gradient: 5 min, 2% solvent B; 6 min, 10% solvent B; 90 min, 55% solvent B; 95 min, 100% solvent B; and 110 min, 100% solvent B. The column end was connected directly to the spotting tip of DiNA MAP target plate spotting

device (KYA Technologies). CHCA matrix solution was prepared at 1.5 mg/mL concentration in 70% acetonitrile, 0.1% TFA and 0.03 mg/mL ammonium citrate, which was pumped to the spotting tip at 2.2 μL/min flow rate and therein mixed with column elution. The mixture was deposited onto a 1536-well μFocusing plate (Hudson Surface Technologies Inc., Newark, NJ) every 15 seconds between 20.0 to 109.75 minutes for a total of 360 spot fractions. Mass spectrometric analysis was performed using 4800 Plus MALDI TOF/TOF Analyzer (AB Sciex, Foster City, CA) operated on 4000 Series Explorer software version 3.5. For each fraction spot, data was accumulated from 1000 laser shots in a randomized raster of 400 μm diameter over mass range m/z 800-4000. The laser repeat rate was 200 Hz and the laser power was fixed at 3500 units throughout the experiment. 5 calibration spots comprising 6 standard peptides were used for external calibration.

Data analysis

Individual MALDI MS raw data was exported as t2d file, ordered in chromatographic order and imported into Expressionist Refiner MS system (Genedata AG, Basel, Switzerland), where they were combined and displayed as mass chromatograms. Default processing parameters were applied unless otherwise specified. The chromatogram data was first simplified by subtracting the background noise by using the following criteria: 0.3 min RT window, 40% quantile subtraction, 0.15 point RT smoothing. After subtraction, all data points below threshold intensity of "100" were clipped to zero. A set of chromatograms were then aligned in the RT direction by nonlinear transformation, mapping the original time onto a common universal retention time, to ensure that equal RT values correspond to the elution of the same compounds. The following parameters were applied: RT transformation window, 5 min; RT search interval, 30 min; m/z window, 0.2 Da; gap penalty, 1. Peak signals were detected by summed peak detection algorithm, which computes a temporary averaged chromatogram over all input chromatograms, thereby allowing them to share the matching set of peaks with identical boundaries. Here, the summation windows of 0.2 Da in the m/z direction and 1 minute in the RT direction were selected. The detected peaks were grouped into isotopic clusters of individual compounds by summed isotope clustering activity, using the following parameters: minimum charge, 1; maximum charge, 2; maximum missing peaks, 1; first allowed gap position, 3; RT window, 1 min; peptide isotope shaping tolerance, 0.8.

Statistical analysis

The cluster information generated by Refiner MS was imported into Genedata Expressionist Analyst software for statistical analysis. The clusters were first filtered by a valid value proportion of 100% (i.e. signal was detected in all of the experimental replicates). All of these clusters were subjected to t-test for extracting differentially expressed clusters between the experimental groups, where $P < 0.01$ was considered to be significant. Present/absent search was performed to select for clusters with 0 or 1 counterpart detection, which were omitted by t-test.

Protein identification

As an exhaustive study, full MS/MS analysis was performed on three of the replicate runs. Precursor peaks were selected according to the software interpretation algorithm, while limiting the maximum number of acquisition to 10 per spot. Precursor peaks were measured in descending order of intensity. Precursor ions were isolated at 150 FWHM resolution, fragmentation was induced without the use of collision gas at 6 kV and fragment ions were further accelerated at 15 kV. Laser power of 4200 units was used, and the acquisition was summed over 2000 laser shots or until 4 fragment peaks exceeded S/N 100. Protein Pilot software version 2.0 was used to generate MS/MS peak lists for searching by MASCOT [32] version 2.2.03 (Matrix Science, London, U.K.) against 20345 human sequences of SwissProt version 57.14. Prior to search, a custom +3 Da modification on asparagine residue resulting from deglycosylation in $H_2{}^{18}O$ was defined. The search parameters were as follows: enzyme, trypsin (allow up to 2 missed cleavages); fixed modification, carbamidomethyl; variable modifications, ^{18}O-deglycosylation (Asn); peptide tolerance, 300 ppm; MSMS tolerance, 0.5 Da. Ion expectation score of 0.05 was used for the cut-off line for identification. For candidate biomarker peptides that were not identified by this method, searching was iteratively repeated in different search parameters, such as "semitrypsin" enzyme restriction, "N-terminal pyroglutamic acid" and "+365 Da modification on Thr" (corresponding to O-linked N-acetylhexosamine and hexose attached to a threonine residue) as variable modifications.

Multiple reaction monitoring

2 μL of the serum tryptic digest analyzed by LC-MALDI was diluted by adding 6 μL of solvent A and 4 μL of tBSA proteomic standard (KYA Technologies) dissolved at 50 fmol/μL. 1 μL of this mixture was injected for a single analysis. Paradigm nano-HPLC system with PAL autoinjector was used for separation. Solvent A was 2% ACN in 0.1% formic acid, solvent B was 90% ACN in 0.1% formic acid, and sample was loaded with 2% ACN in 0.1% TFA. The trap column was L-column ODS, 5 μm, 0.3 × 5 mm, and the analytical column was L-column ODS loaded in-house directly into a sprayer tip (GL Science, Tokyo, Japan).

MRM was performed using 4000QTRAP mass spectrometer (AB Sciex) during a 13 minutes gradient (2-55% solvent B) at 200 nL/min flow rate. 70 ions were monitored simultaneously for 30 minutes, each transition with 20 ms dwell time with 5 ms interval, taking a total of 1.75 s per scan. Transitions were selected from series of pilot experiment in which in-silico developed transitions for each peptide were tested for signal intensity and specificity. Transitions were considered as the derivative of target peptide only when all of them responded simultaneously and the retention time of detection matched that of LC-MALDI data. Instrument settings were as follows: declustering potential, 70; entrance potential, 10; curtain gas, 10; collision gas, 4; ion spray voltage, 2100; ion source gas, 10; interphase heater temperature, 150°C. Peak areas were integrated using MultiQuant software version 1.1.0.26. Raw data was normalized to the total signal acquired, which includes two spiked-in BSA fragments detected at highest intensity.

Immunoblot analysis

For the verification study, crude serum samples from fresh aliquots for all screening sample set (except for 3 normal controls N-3, 4, and 10, which were substituted by N-11, 12, and 13, respectively) were analyzed. 0.5 μL of crude serum was diluted 100-fold with SDS-PAGE sample buffer, boiled and 20 μL was used for immunoblot analysis. SDS-PAGE was performed using NuPAGE Bis-Tris 4-12% acrylamide gel with 2-morpholinoethanesulfonic acid buffer system, and electroblotted onto a PVDF membrane. The blots were probed with anti-C3 polyclonal antibody (Sigma, product code GW20073F) diluted 5000-fold in 5% skim milk, followed by incubation with horseradish peroxidise-conjugated secondary antibody (Sigma, product code A9046) diluted 10000-fold in 2% BSA. The reactivity was visualized on X-ray films using ECL detection kit (GE Healthcare). Immunoblot was also performed with anti-C3d monoclonal antibody (Abbiotec, San Diego, CA) and horseradish peroxidise-conjugated secondary antibody (GE Healthcare) to confirm specificity of the polyclonal antibody (Figure S-2).

Additional material

Additional file 1: Supplementary Information. This PDF file contains the following material: Figure S-1, a histogram and summary of the number of reproducible peaks. Figure S-2, an immunoblot showing the specificity of antibodies used. Table S-1, the list of serum samples. Table S-2, the list of glycopeptides identified in this study. Table S-3, the detail of MRM transitions used for verification analysis.

Author details
[1]Graduate School of Frontier Sciences, The University of Tokyo, Kashiwanoha5-1-5, Chiba, Japan. [2]Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, Tsurumiku-Suehirocho1-7-22, Yokohama, Japan. [3]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Shirokanedai4-6-1, Tokyo, Japan. [4]Department of Molecular and Internal Medicine, Hiroshima University, Minamiku-Kasumi1-2-3, Hiroshima, Japan. [5]Shimadzu Corporation, Nishinokyo-Kuwabaracho1, Kyoto, Japan.

Authors' contributions
AT performed all experiments and drafted the manuscript. HN and KM added statistical interpretation. NI and NK provided the serum samples and revised the manuscript. YD and TAS contributed to mass spectrometry analysis. YN critically revised the manuscript. KU conceived the study, helped in experimental design and critically revised the manuscript. All authors read and approved the final manuscript.

Competing interests
This study is supported in part by the grants from Toppan Printing Co., Ltd., Tokyo, Japan, and Shimadzu Corporation, Kyoto, Japan.

Received: 8 December 2010 Accepted: 8 April 2011
Published: 8 April 2011

References

1. Anderson NL, Anderson NG: **The human plasma proteome: history, character, and diagnostic prospects.** *Mol Cell Proteomics* 2002, 1:845-867.
2. Ong SE, Mann M: **Mass spectrometry-based proteomics turns quantitative.** *Nat Chem Biol* 2005, 1:252-262.
3. Gygi SP, Corthals GL, Zhang Y, Rochon Y, Aebersold R: **Evaluation of two-dimensional gel electrophoresis-based proteome analysis technology.** *Proc Natl Acad Sci USA* 2000, 97:9390-9395.
4. Kuyama H, Watanabe M, Toda C, Ando E, Tanaka K, Nishimura O: **An approach to quantitative proteome analysis by labeling tryptophan residues.** *Rapid Commun Mass Spectrom* 2003, 17:1642-1650.
5. Ong SE, Blagoev B, Kratchmarova I, Kristensen DB, Steen H, Pandey A, Mann M: **Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.** *Mol Cell Proteomics* 2002, 1:376-386.
6. Ross PL, Huang YN, Marchese JN, Williamson B, Parker K, Hattan S, Khainovski N, Pillai S, Dey S, Daniels S, *et al*: **Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.** *Mol Cell Proteomics* 2004, 3:1154-1169.
7. Schulze WX, Usadel B: **Quantitation in Mass-Spectrometry-Based Proteomics.** *Annu Rev Plant Biol* 2010.
8. Wang W, Zhou H, Lin H, Roy S, Shaler TA, Hill LR, Norton S, Kumar P, Anderle M, Becker CH: **Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards.** *Anal Chem* 2003, 75:4818-4826.
9. Bondarenko PV, Chelius D, Shaler TA: **Identification and relative quantitation of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry.** *Anal Chem* 2002, 74:4741-4749.
10. Omenn GS, States DJ, Adamski M, Blackwell TW, Menon R, Hermjakob H, Apweiler R, Haab BB, Simpson RJ, Eddes JS, *et al*: **Overview of the HUPO Plasma Proteome Project: results from the pilot phase with 35 collaborating laboratories and multiple analytical groups, generating a core dataset of 3020 proteins and a publicly-available database.** *Proteomics* 2005, 5:3226-3245.
11. Zhang H, Li XJ, Martin DB, Aebersold R: **Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry.** *Nat Biotechnol* 2003, 21:660-666.
12. Liu T, Qian WJ, Gritsenko MA, Camp DG, Monroe ME, Moore RJ, Smith RD: **Human plasma N-glycoproteome analysis by immunoaffinity subtraction, hydrazide chemistry, and mass spectrometry.** *J Proteome Res* 2005, 4:2070-2080.
13. Kaji H, Saito H, Yamauchi Y, Shinkawa T, Taoka M, Hirabayashi J, Kasai K, Takahashi N, Isobe T: **Lectin affinity capture, isotope-coded tagging and mass spectrometry to identify N-linked glycoproteins.** *Nat Biotechnol* 2003, 21:667-672.
14. Constantopoulos TL, Jackson GS, Enke CG: **Effects of salt concentration on analyte response using electrospray ionization mass spectrometry.** *J Am Soc Mass Spectrom* 1999, 10:625-634.

15. Makawita S, Diamandis EP: **The bottleneck in the cancer biomarker pipeline and protein quantification through mass spectrometry-based approaches: current strategies for candidate verification.** *Clin Chem* 2010, **56**:212-222.
16. Hattan SJ, Parker KC: **Methodology utilizing MS signal intensity and LC retention time for quantitative analysis and precursor ion selection in proteomic LC-MALDI analyses.** *Anal Chem* 2006, **78**:7986-7996.
17. Sahu A, Lambris JD: **Structure and biology of complement protein C3, a connecting link between innate and acquired immunity.** *Immunol Rev* 2001, **180**:35-48.
18. Lochnit G, Geyer R: **An optimized protocol for nano-LC-MALDI-TOF-MS coupling for the analysis of proteolytic digests of glycoproteins.** *Biomed Chromatogr* 2004, **18**:841-848.
19. Wuhrer M, Catalina MI, Deelder AM, Hokke CH: **Glycoproteomics based on tandem mass spectrometry of glycopeptides.** *J Chromatogr B Analyt Technol Biomed Life Sci* 2007, **849**:115-128.
20. Liu T, Qian WJ, Gritsenko MA, Xiao W, Moldawer LL, Kaushal A, Monroe ME, Varnum SM, Moore RJ, Purvine SO, *et al*: **High dynamic range characterization of the trauma patient plasma proteome.** *Mol Cell Proteomics* 2006, **5**:1899-1913.
21. Wang Y, Wu SL, Hancock WS: **Approaches to the study of N-linked glycoproteins in human plasma using lectin affinity chromatography and nano-HPLC coupled to electrospray linear ion trap–Fourier transform mass spectrometry.** *Glycobiology* 2006, **16**:514-523.
22. Angel PM, Lim JM, Wells L, Bergmann C, Orlando R: **A potential pitfall in 18O-based N-linked glycosylation site mapping.** *Rapid Commun Mass Spectrom* 2007, **21**:674-682.
23. Fukuyama Y, Nakaya S, Yamazaki Y, Tanaka K: **Ionic liquid matrixes optimized for MALDI-MS of sulfated/sialylated/neutral oligosaccharides and glycopeptides.** *Anal Chem* 2008, **80**:2171-2179.
24. Seraglia R, Molin L, Tonidandel L, Pucciarelli S, Agostini M, Urso ED, Bedin C, Quaia M, Nitti D, Traldi P: **An investigation on the nature of the peptide at m/z 904, overexpressed in plasma of patients with colorectal cancer and familial adenomatous polyposis.** *J Mass Spectrom* 2007, **42**:1606-1612.
25. Xue H, Lu B, Zhang J, Wu M, Huang Q, Wu Q, Sheng H, Wu D, Hu J, Lai M: **Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach.** *J Proteome Res* 2010, **9**:545-555.
26. Huber R, Scholze H, Paques EP, Deisenhofer J: **Crystal structure analysis and molecular model of human C3a anaphylatoxin.** *Hoppe Seylers Z Physiol Chem* 1980, **361**:1389-1399.
27. Ward DG, Suggett N, Cheng Y, Wei W, Johnson H, Billingham LJ, Ismail T, Wakelam MJ, Johnson PJ, Martin A: **Identification of serum biomarkers for colon cancer by proteomic analysis.** *Br J Cancer* 2006, **94**:1898-1905.
28. Lee IN, Chen CH, Sheu JC, Lee HS, Huang GT, Chen DS, Yu CY, Wen CL, Lu FJ, Chow LP: **Identification of complement C3a as a candidate biomarker in human chronic hepatitis C and HCV-related hepatocellular carcinoma using a proteomics approach.** *Proteomics* 2006, **6**:2865-2873.
29. Zhang R, Barker L, Pinchev D, Marshall J, Rasamoelisolo M, Smith C, Kupchak P, Kireeva I, Ingratta L, Jackowski G: **Mining biomarkers in human sera using proteomic tools.** *Proteomics* 2004, **4**:244-256.
30. Miguet L, Bogumil R, Decloquement P, Herbrecht R, Potier N, Mauvieux L, Van Dorsselaer A: **Discovery and identification of potential biomarkers in a prospective study of chronic lymphoid malignancies using SELDI-TOF-MS.** *J Proteome Res* 2006, **5**:2258-2269.
31. Diefenbach RJ, Isenman DE: **Mutation of residues in the C3dg region of human complement component C3 corresponding to a proposed binding site for complement receptor type 2 (CR2, CD21) does not abolish binding of iC3b or C3dg to CR2.** *J Immunol* 1995, **154**:2303-2320.
32. Lu P, Vogel C, Wang R, Yao X, Marcotte EM: **Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation.** *Nat Biotechnol* 2007, **25**:117-124.
33. Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M: **Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein.** *Mol Cell Proteomics* 2005, **4**:1265-1272.
34. Rower C, Vissers JP, Koy C, Kipping M, Hecker M, Reimer T, Gerber B, Thiesen HJ, Glocker MO: **Towards a proteome signature for invasive ductal breast carcinoma derived from label-free nanoscale LC-MS protein expression profiling of tumorous and glandular tissue.** *Anal Bioanal Chem* 2009, **395**:2443-2456.
35. Monroe ME, Tolic N, Jaitly N, Shaw JL, Adkins JN, Smith RD: **VIPER: an advanced software package to support high-throughput LC-MS peptide identification.** *Bioinformatics* 2007, **23**:2021-2023.
36. Washburn MP, Wolters D, Yates JR: **Large-scale analysis of the yeast proteome by multidimensional protein identification technology.** *Nat Biotechnol* 2001, **19**:242-247.
37. Wu CC, MacCoss MJ, Howell KE, Yates JR: **A method for the comprehensive proteomic analysis of membrane proteins.** *Nat Biotechnol* 2003, **21**:532-538.
38. Iwasaki M, Miwa S, Ikegami T, Tomita M, Tanaka N, Ishihama Y: **One-dimensional capillary liquid chromatographic separation coupled with tandem mass spectrometry unveils the Escherichia coli proteome on a microarray scale.** *Anal Chem* 2010, **82**:2616-2620.
39. Knochenmuss R, Zenobi R: **MALDI ionization: the role of in-plume processes.** *Chem Rev* 2003, **103**:441-452.

doi:10.1186/1477-5956-9-18
Cite this article as: Toyama *et al.*: Deglycosylation and label-free quantitative LC-MALDI MS applied to efficient serum biomarker discovery of lung cancer. *Proteome Science* 2011 **9**:18.

C12orf48, Termed PARP-1 Binding Protein, Enhances Poly(ADP-Ribose) Polymerase-1 (PARP-1) Activity and Protects Pancreatic Cancer Cells from DNA Damage

Lianhua Piao,[1] Hidewaki Nakagawa,[1,2] Koji Ueda,[2] Suyoun Chung,[1] Kotoe Kashiwaya,[1] Hidetoshi Eguchi,[3] Hiroaki Ohigashi,[3] Osamu Ishikawa,[3] Yataro Daigo,[1] Koichi Matsuda,[1] and Yusuke Nakamura[1]*

[1]Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan
[2]Laboratory for Biomarker Development, Center of Genomic Medicine, RIKEN, Yokohama, Japan
[3]Department of Surgery, Osaka Medical Center for Cancer and Cardiovascular Diseases, Osaka, Japan

To identify novel therapeutic targets for aggressive and therapy-resistant pancreatic cancer, we had previously performed expression profile analysis of pancreatic cancers using microarrays and found dozens of genes trans-activated in pancreatic ductal adenocarcinoma (PDAC) cells. Among them, this study focused on the characterization of a novel gene *C12orf48* whose overexpression in PDAC cells was validated by Northern blot and immunohistochemical analysis. Its overexpression was observed in other aggressive and therapy-resistant malignancies as well. Knockdown of C12orf48 by siRNA in PDAC cells significantly suppressed their growth. Importantly, we demonstrated that C12orf48 protein could directly interact with Poly(ADP-ribose) Polymerase-1 (PARP-1), one of the essential proteins in the repair of DNA damage, and positively regulate the poly(ADP-ribosyl)ation activity of PARP-1. Depletion of C12orf48 sensitized PDAC cells to agents causing DNA damage and also enhanced DNA damage-induced G2/M arrest through reduction of PARP-1 enzymatic activities. Hence, our findings implicate C12orf48, termed PARP-1 binding protein (PARPBP), or its interaction with PARP-1 to be a potential molecular target for development of selective therapy for pancreatic cancer.

INTRODUCTION

Pancreatic cancer is the fourth leading cause of cancer death in the western world and shows the worst mortality among common malignancies with a 5-year survival rate of lower than 5% (DiMagno et al., 1999; Wray et al., 2005). It is estimated that a total of 42,470 new cases were diagnosed to have pancreatic cancer and 35,240 deaths were caused by it in the United States in 2009 (Jemal et al., 2009). At present, only surgical resection can offer a chance for cure or long-term survival to the patients suffering from pancreatic cancer. However, only 10–20% of patients with pancreatic cancer are able to have radical surgery because most of the patients are already at an advanced stage at the time of diagnosis (DiMagno et al., 1999; Wray et al., 2005). Gemcitabine or 5-fluorouracil chemotherapy coupled with radiotherapy could improve the quality of life of the patients (DiMagno et al., 1999; Wray et al., 2005), but its survival benefit is very limited. Hence, there has been no substantial improvement in relative 5-year survival rate for pancreatic cancer in the past 3 decades (Jemal et al., 2009). To overcome this dismal situation, development of novel molecular therapies targeting a molecule specifically functioning in pancreatic cancer is eagerly awaited. We had previously performed extensive genome-wide expression profile analysis of pancreatic cancer cells in combination with microdissection to enrich cancer cell population (Nakamura et al., 2004), and demonstrated some genes trans-activated in PDAC cells to be possible molecular targets for development of new therapeutic modalities to treat pancreatic cancers (Taniuchi et al., 2005a,b; Iiizumi et al., 2006; Takehara et al., 2006, 2007; Hosokawa et al., 2007, 2008; Kashiwaya et al., 2009).

Poly(ADP-ribose) polymerase-1 (PARP-1), a nuclear enzyme, catalyzes the transfer of the ADP-ribose unit from its substrate, NAD^+, to some protein acceptors such as histones, p53, and PARP-1 itself. The addition of negatively charged polymers profoundly alters the properties

Additional Supporting Information may be found in the online version of this article.

Supported by: Japan Society for the Promotion of Science. Grant number: 00L01402.

*Correspondence to: Yusuke Nakamura, M.D., Ph.D., Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. E-mail: yusuke@ims.u-tokyo.ac.jp

Received 28 July 2010; Accepted 2 September 2010

DOI 10.1002/gcc.20828

Published online 7 October 2010 in Wiley Online Library (wileyonlinelibrary.com).

and functions of the target proteins. Through its physical association with partner proteins or by the poly(ADP-ribosyl)ation of them, PARP-1 is involved in multiple cellular processes including DNA repair, transcriptional regulation, chromatin modification, cell cycle progression, or genomic stability (Ogata et al., 1981; Kameshita et al., 1984). PARP-1 is a molecular nick-sensor of DNA breaks and has a critical role in the spatial and temporal organization of the DNA repairs (de Murcia et al., 1994). The activation of PARP-1 after DNA damage provides rapid signals to halt transcription and recruits enzymes required for DNA repair to the site of DNA damage, including XRCC1, DNA ligase III, and DNA polymerase β. PARP-1 is essential in the repair of both DNA single-strand breaks (SSB) as well as double-strand breaks (DSB) (Durkacz et al., 1980; D'Silva et al., 1999; Dantzer et al., 2000; Audebert et al., 2004, 2006; Wang et al., 2006). The involvement of PARP-1 in the DNA repair system prompted us to investigate the effect of PARP-1 inhibition on DNA-damaging anticancer therapies (Daniel et al., 2009; Horton et al., 2009). Inhibition of PARP-1 enhanced the cytotoxicity of DNA-damaging agents and seemed to overcome one of the causes of resistance in cancer cells to anticancer treatment (Hoeijmakers et al., 2001; Longley and Johnston, 2005). Currently, several PARP-1 inhibitors have already been taken into the clinical trials as chemopotentiating or radio-potentiating agents, and have shown promising results (Miknyoczki et al., 2007; Plummer et al., 2008; Rottenberg et al., 2008; Horton et al., 2009; Jones et al., 2009; O'Shaughnessy et al., 2009).

We here focus on the characterization of a novel gene *C12orf48* (*Chromosome 12 open reading frame 48*). We demonstrate that C12orf48 protein can interact with PARP-1 directly and be involved in the repair of DNA breaks through enhancing PARP-1 activity. Thus, we termed this molecule PARP-1 binding protein (PARPBP). These findings indicate that C12orf48, or its interaction with PARP-1 could be a promising molecular target for the development of novel treatment for pancreatic cancer.

MATERIALS AND METHODS

Cell Lines

PDAC cell lines, KLM-1, SUIT-2, KP-1N, PK-1, PK-45P, and PK-59, were provided from Cell Resource Center for Biomedical Research, Tohoku University (Sendai, Japan). MIAPaCa-2, Panc-1 and COS7 cell lines were purchased from the American Type Culture Collection (ATCC, Rockville, Maryland). KLM-1, SUIT-2, PK-1, PK-45P, PK-59 and Panc-1, were grown in RPMI 1640 (Sigma-Aldrich, St. Louis, Missouri), and COS7, MIAPaCa-2 in DMEM (Sigma-Aldrich), with 10% fetal bovine serum and 1% antibiotic/antimycotic solution (Sigma-Aldrich).

Semi-Quantitative Reverse Transcription-PCR

Microdissection of PDAC cells and normal pancreatic ductal cells were described previously (Nakamura et al., 2004). RNAs from these cells were subjected to two rounds of RNA amplification using T7-based in vitro transcription (Epicenter Technologies, Madison, Wisconsin). Total RNAs from human PDAC cell lines were extracted using Trizol reagent (Invitrogen) according to the manufacturer's instructions. Extracted RNAs were treated with DNase I (Roche, Mannheim, Germany) and reversely transcribed to single-stranded cDNAs using oligo $(dT)_{12-18}$ primer with Superscript II reverse transcriptase (Invitrogen). The primer sequences were 5′-TTGGCTTGACTCAGGATTTA-3′ and reverse 5′-ATGCTATCACCTCCCCTGTG-3′ for β-*actin* (*ACTB*), and 5′-CTCAGCTGGG AAAGCTACAGAT-3′ and 5′-CATGCCAGGT AGTTCTTCCATC-3′ for *C12orf48* (GenBank Accession no. NM_017915). Each PCR regime involved initial denaturation at 94°C for 2 min followed by 23 cycles (for *ACTB*), 28 cycles (for *C12orf48*) at 94°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min.

Northern Blot Analysis

One μg each of polyA RNA extracted from eight PDAC cell lines (KLM-1, PK-59, PK-45P, MIAPaCa-2, KP-1N, Panc-1, PK-1, and SUIT-2) and seven adult normal tissues (heart, lung, liver, kidney, brain, testis, and pancreas, from BD Bioscience, Palo Alto, CA) was blotted onto a nylon membrane. The 305-bp probe specific to *C12orf48* was prepared by PCR using the primer set described above. The cancer membrane and human Multiple Tissue Northern blot membrane (Clontech, Mountain View, CA) were hybridized with the cDNA probe labeled with α^{32}P-dCTP using Mega Label kit (GE Healthcare, Piscataway, New Jersey). Prehybridization, hybridization, and

washing were performed according to the manufacturer's instruction. The blots were autoradiographed at −80°C for 10 days.

Generation of Antibodies to C12orf48 and Immunocytochemistry/Immunohistochemistry

Plasmids expressing two fragments of C12orf48 (codons 1–150 and 328–498) in pET21a(+) vector (Novagen, Madison, Wisconsin) were constructed to produce recombinant proteins in *E. coli*. The recombinant C12orf48 proteins were purified using Ni-NTA resin agarose (Qiagen, Valencia, CA) and used to immunize rabbits. The sera from the immunized rabbits were purified by antigen-Affi-Gel 10 (Bio-Rad Laboratories, Hercules, CA) affinity column chromatography. For immunocytochemical analysis, KLM-1 cells fixed by 4% paraformaldehyde were incubated with rabbit anti-C12orf48 polyclonal antibody for 1 hr. After washing with PBS, the cells were stained by Alexa 488-conjugated anti-rabbit IgG secondary antibodies (Molecular Probes, Eugene, Oregon) for 1 hr. Stained preparations were mounted with VECTASHIELD mounting medium with DAPI (Vector Laboratories, Burlingame, CA). For immunohistochemistry, tissue sections of PDACs were obtained from the Osaka Medical Center for Cancer and Cardiovascular Diseases under the written informed consent. Human PDAC tissue microarrays were purchased from ISU-ABXIS (Accurate Chemical Corp., Westbury, New York). The sections were deparaffinized and autoclaved at 108°C in Dako Cytomation Target Retrieval Solution High pH (Dako Cytomation, Carpinteria, CA) for 15 min. After blocking, the sections were incubated with rabbit anti-C12orf48 antibody (dilution 1:2,500) at room temperature for 1 hr, washed three times in PBS, and incubated with peroxidase labeled anti-rabbit immunoglobulin (Envision kit; Dako Cytomation). Finally, the reactants were developed with 3,3′-diaminobenzidine. Counterstaining was performed using hematoxylin.

Short-Hairpin RNA-Expressing Constructs

The psiU6BX3.0 vector for expression of short-hairpin RNA (shRNA) was constructed to knock down the expression of the target genes, as described previously (Taniuchi et al., 2005b). The target sequences for *C12orf48* were 5′-CACAGTATCTCCTAGTCAA-3′ (si1), 5′-GTT GCTCAGGATTTGGATT-3′ (si2), 5′-GCAGC TAATGCTCCTACCA-3′ (si3), and 5′-GAAG CAGCACGACTTCTTC-3′ (siEGFP) as a negative control. PDAC cell lines, KLM-1 and SUIT-2, were transfected with each of these shRNA-expression vectors using FuGENE6 (Roche), and selected with Geneticin (GIBCO, 0.5 mg/mL for KLM-1 cells and 0.9 mg/mL for SUIT-2 cells, respectively). Cell viability was measured using cell-counting kit-8 (DOJINDO, Kumamoto, Japan) 6 days after the transfection. Absorbance was measured at 490 nm, and at 630 nm as reference, with a Microplate Reader 550 (Bio-Rad). After 2 weeks of the selection, cancer cells were fixed with 100% methanol and stained with 0.1% of crystal violet-H_2O.

Immunoprecipitation and Mass-Spectrometric Analysis

The pCAGGS Flag-C12orf48-HA vector was constructed by PCR cloning, and was transfected to HEK293 cells. The transfected cells were lysed in lysis buffer (50 mmol/L Tris-HCl [pH 8.0], 0.4% NP-40, 150 mmol/L NaCl, Protease Inhibitor Cocktail Set III [Calbiochem, San Diego, CA]). Cell extracts were precleared by incubation with CL-4B sepharose (Sigma-Aldrich) at 4°C for 1 hr, and incubated with anti-FLAG M_2-agarose (Sigma-Aldrich) for 1 hr. The proteins were separated in 5–20% gradient SDS-PAGE gels (Bio-Rad) and stained with a silver-staining kit (Invitrogen). Protein bands that specifically observed in the cell extracts transfected with pCAGGS Flag-C12orf48-HA were excised and analyzed by liquid chromatography-mass spectrometry (LC-MS/MS). The excised proteins were reduced in 10-mM tris(2-carboxyethyl)phosphine (Sigma-Aldrich) with 50-mM ammonium bicarbonate (Sigma-Aldrich) for 30 min at 37°C and alkylated in 50-mM iodoacetamide (Sigma-Aldrich) with 50-mM ammonium bicarbonate for 45 min in the dark at 25°C. Porcine trypsin (Promega, San Luis Obispo, CA) was added for a final enzyme to protein ratio of 1:20. The digestion was conducted at 37°C for 16 hr. The resulting peptide mixture was separated on a 100 μm × 150 mm HiQ-Sil C18W-3 column (KYA Technologies, Tokyo, Japan) using 30 min linear gradient from 5.4 to 29.2% acetonitrile in 0.1% trifluoroacetic acid (TFA) with total flow of 300 nL/min. The eluting peptides were automatically mixed with matrix solution (4 mg/mL α-cyano-4-hydroxy-cinnamic acid (Sigma-Aldrich), 0.08 mg/mL ammonium citrate in 70% acetonitrile, 0.1% TFA) and spotted

onto MALDI target plates (KYA Technologies). Mass spectrometric analysis was performed on 4800 Plus MALDI/TOF/TOF Analyzer (AB SCIEX Foster City, CA). MS/MS peak list was generated by the Protein Pilot version 2.0.1 software (AB SCIEX) and exported to a local MASCOT search engine version 2.2.03 (Matrix Science) for protein database search.

Flow Cytometry and Synchronization

KLM-1 cells were transfected with *C12orf48*-specific siRNA duplex (5′-CUAGUCAACUA CUGGAUUU-3′), *PARP-1*-specific siRNA duplex (5′-GAUAGAGCGUGAAGGCGAA-3′), and siEG FP duplex (5′-GAAGCAGCACGACUUCUUC-3′) as a negative control, respectively, by using Lipofectamine RNAiMAX (invitrogen) according to the manufacturer's recommendations. 96 hr after the transfection, the cells were fixed with 70% ethanol in PBS at 4°C, and incubated with 500 μL of PBS containing 0.5 mg of boiled RNase at 37°C for 30 min. Finally, 2×10^4 cells stained with 50 μg/mL propidium iodide were analyzed by means of Cell Lab Quanta™ SC MPL Flow Cytometer (Beckman Coulter, USA). A complete block at G1/S-phase was achieved by treatment with 2 μg/mL aphidicolin for 24 hr. Then, cells were released from the cell-cycle arrest, harvested, and prepared for flow cytometry analysis (FACS).

In Vitro PARP-1 Auto-Poly(ADP-Ribosyl)ation Assays

In vitro PARP-1 automodification assays were performed as described previously (Di Palma et al., 2008). Briefly, 200 ng of the purified C12orf48 recombinant protein and 25 ng of the recombinant human PARP-1 (Alexis, San Diego, CA) were incubated in binding buffer (10 mM Tris-HCl, pH 7.5, 1 mM $MgCl_2$, 1 mM DTT) plus 10 μg/mL of sonicated DNA at 37°C for 10 min. The reactions were started by adding ^{32}P-labeled NAD^+, and incubated at 37°C for 10 min. After terminating the reactions with SDS sample buffer, the proteins were fractionated by 8% SDS-PAGE gel. Incorporation of ^{32}P-labeled NAD^+ to poly(ADP-ribosyl)ated proteins was visualized by autoradiography.

PARP-1 Activity in Cell Extracts

KLM-1 and SUIT-2 cells were transfected with *C12orf48*-siRNA, *PARP-1*-siRNA, or siEGFP (as a control), and collected 72 hr after the transfection. The knockdown effects were confirmed with anti-C12orf48 antibody and anti-PARP-1 antibody (Santa Cruz Biotechnology), respectively. PARP-1 activities in cell extracts were assayed using the universal colorimetric PARP assay kit (Trevigen, Gaithersburg, Maryland) based on the incorporation of biotinylated ADP-ribose onto histone H1 proteins. Briefly, cell extracts were loaded into a 96-well plate coated with histone H1, and incubated with biotinylated poly(ADP-ribose) and nicked DNA (Trevigen), size of which are 200–500 base pairs that are considered to be optimal for the PARP activation, for 1 hr. After wash with PBS containing 0.1% (v/v) Triton X-100, streptavidin-HRP (horseradish peroxidase) was added and incubated additionally for 20 min. TACS-Sapphire™ was added subsequently to develop colors and the reaction was stopped by addition of 5% phosphoric acid. Finally, the absorbance was measured at 450 nm in a spectrometrophotometer. PARP-1 enzymatic activities were also evaluated by the use of mouse anti-poly(ADP-ribose) (PAR) monoclonal antibody (Trevigen). 25 μg of the cell extracts obtained from the KLM-1 cells that were transfected with *C12orf48*-siRNA, *PARP-1*-siRNA, or siEGFP (as a control), or 5 ng of recombinant human PARP-1 (Trevigen) were incubated for 20 min at 37°C in binding buffer (10 mM Tris-HCl, pH 7.5, 1 mM $MgCl_2$, 1 mM DTT) plus 10 μg/mL of sonicated DNA, and 200 μM NAD^+ (Sigma-Aldrich).

Sensitivity to DNA Damage

KLM-1 cells were transfected with oligo *C12orf48*-siRNA or siEGFP (as a control), and incubated for 48 hr to knockdown C12orf48 expression as described above. These transfected KLM-1 cells were trypsinized, and the number of living cells was counted. 5×10^5 cells were re-seeded into the 6-well plates, and incubated with indicated concentrations of Adriamycin for 24 hr, H_2O_2 for 6 hr, or exposed to indicated intensity of UV radiation, and then incubated for 24 hr. Cell viability was measured using Cell-counting kit-8 as described above.

RESULTS

Overexpression of C12orf48 in PDAC Cells

Among the transactivated genes that were identified through our genome-wide microarray analysis of pancreatic cancer cells (Nakamura