Development of an orally-administrative MELK-targeting inhibitor that suppresses the growth of various types of human cancer

Suyoun Chung[1], Hanae Suzuki[2], Takashi Miyamoto[2], Naofumi Takamatsu[2], Ayako Tatsuguchi[3], Koji Ueda[3], Kyoko Kijima[2], Yusuke Nakamura[1,4] and Yo Matsuo[2]

[1] Department of Medicine and Surgery, The University of Chicago, Chicago, IL, USA

[2] OncoTherapy Science, Inc., Kawasaki, Kanagawa, Japan

[3] Laboratory for Biomarker Development, RIKEN, Yokohama, Japan

[4] Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Correspondence to: *Yusuke Nakamura*, *email*: *ynakamura@bsd.uchicago.edu*

Yo Matsuo, *email*: *y-matsuo@oncotherapy.co.jp*

Keywords: oncogene, drug discovery, kinase inhibitor, cancer stem cell

Received: *December 05, 2012*, *Accepted*: *December 20, 2012*, *Published*: *December 21, 2012*

Copyright: © Chung et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT:

We previously reported MELK (maternal embryonic leucine zipper kinase) as a novel therapeutic target for breast cancer. MELK was also reported to be highly upregulated in multiple types of human cancer. It was implied to play indispensable roles in cancer cell survival and indicated its involvement in the maintenance of tumor-initiating cells. We conducted a high-throughput screening of a compound library followed by structure-activity relationship studies, and successfully obtained a highly potent MELK inhibitor OTSSP167 with IC_{50} of 0.41 nM. OTSSP167 inhibited the phosphorylation of PSMA1 (proteasome subunit alpha type 1) and DBNL (drebrin-like), which we identified as novel MELK substrates and are important for stem-cell characteristics and invasiveness. The compound suppressed mammosphere formation of breast cancer cells and exhibited significant tumor growth suppression in xenograft studies using breast, lung, prostate, and pancreas cancer cell lines in mice by both intravenous and oral administration. This MELK inhibitor should be a promising compound possibly to suppress the growth of tumor-initiating cells and be applied for treatment of a wide range of human cancer.

INTRODUCTION

Breast cancer is the most common malignancy among women worldwide[1]. More than 1.3 million patients are newly diagnosed with breast cancer each year, and over 400,000 patients died of the disease[2]. Treatments acting on molecular targets such as estrogen receptor or HER-2 for breast cancer have successfully improved the mortality rate, but a subset of the patients can still have little benefit with these therapies[3, 4]. Triple-negative breast cancer (TNBC), one of the breast cancer subtypes, develops more frequently in younger women and is known to be more aggressive with poor prognosis[5]. Since TNBC does not expresses either of HER-2, estrogen receptor, or progesterone receptor[6], no effective targeted therapy is presently available[5, 7]. Hence, the development of novel targeted drugs for such patients is urgently awaited.

We identified maternal embryonic leucine zipper kinase (MELK), that is a member of the AMPK serine/threonine kinase family and is involved in the mammalian embryonic development[8], to be a promising drug target molecule for breast cancer[9]. MELK was also overexpressed in various types of human cancer including TNBC and its expression was hardly detectable in normal tissues except the testis[9-11]. In addition to the involvement in cancer cell growth, MELK was also reported its critical roles in formation or maintenance of

cancer stem cells[12, 13], that have the ability to self-renew and differentiate. Emerging evidence indicated that the cancer stem cells are resistant to chemotherapy and radiation therapy, and are associated with the cancer relapse[14, 15]. Thus, targeting cancer stem cell is considered as a novel strategy for cancer treatment[16, 17]. The mechanisms how cancer cells acquired these abilities are not yet understood, but recent studies indicated that MELK is one of the marker molecules to characterize cancer stem cells in tumor, such as breast cancer and glioblastoma[13, 18]. Thus, targeting MELK could be an effective strategy to treat multiple types of human cancer.

In this study, we report development of a small-molecule MELK inhibitor OTSSP167 that can selectively and effectively inhibit MELK kinase activity. Our *in vitro* and *in vivo* studies also imply that OTSSP167 significantly suppresses mammosphere formation of breast cancer cells as well as the growth of human cancer-derived xenografts in mice, implying that OTSSP167 has great potential to apply as a novel therapeutics for cancer in a MELK-dependent manner. Furthermore, to verify the molecular mechanism of this MELK-specific inhibitor, we demonstrate identification of new substrates of MELK and inhibitory effect of the compound on activities of these molecules in breast cancer cells.

RESULTS

High-through put screening to identify MELK-specific inhibitor

To obtain small-molecule MELK inhibitors, we first conducted high-throughput screening of a library consisting of 108,269 compounds. Each compound was screened at a single concentration of 30 μM against MELK using the IMAP assay[19] optimized for the high-throughput low-volume 384-well format assays (see Supplementary Methods). The inhibition activity was measured by percent of inhibition of the MELK kinase activity relative to control. The average and standard deviation of the percent inhibition were 0.87% and 9.07%, respectively. A total of 597 compounds revealed the MELK kinase inhibitory activity by 37.1% or higher. After validation by dose-response analysis, a quinoline derivative (compound 1 in Fig 1A) was confirmed to inhibit the MELK activity with the half-maximum inhibitory concentration (IC_{50}) value of 4.8 μM. To develop high-affinity MELK inhibitors, we performed an intensive structure-activity relationship study on the basis of the structure of compound 1, and obtained novel compounds with various degrees of MELK inhibitory activity. Among them, the compound OTSSP167 (Fig 1B) was identified as one of the most effective MELK inhibitor with IC_{50} value of 0.41 nM (see Supplementary Methods for the compound synthesis and the kinase assay). OTSSP167 has a 1,5-naphthyridine core with methylketone at the 3-position, *trans*-4-((dimethylamino)methyl) cyclohexylamino at the 4-position, and 3,5-dichloro-4-hydroxyphenyl at the 6-position of the core.

Growth suppressive effect of OTSSP167 in various cancer types

Since MELK was reported to be overexpressed in other types of human cancer in addition to breast cancer[9, 10], we examined the growth inhibitory effect of OTSSP167 on the growth of various cancer cell lines. *In vitro* anti-proliferative assay using A549 (lung), T47D (breast), DU4475 (breast), and 22Rv1 (prostate) cancer cells, in which MELK was highly expressed, revealed IC_{50} values of 6.7, 4.3, 2.3, and 6.0 nM, respectively (Fig 2A-D). On the other hand, HT1197 (bladder) cancer cells, in which MELK expression was hardly detectable, revealed IC_{50} value of 97 nM (Fig 2E), clearly implying the MELK-dependent growth-inhibition effect of this compound.

Growth suppressive effect of OTSSP167 in xenograft mouse model

We subsequently investigated *in vivo* anti-tumor effect of OTSSP167 by a xenograft model using MDA-MB-231 cells (MELK-positive, triple-negative breast cancer cells). The compound was administered to mice bearing xenografts for 14 days after the tumor size reached about 100 mm^3. The tumor size was measured as a surrogate marker of drug response (tumor growth inhibition (TGI)). Intravenous administration of OTSSP167 at 20 mg/kg once every two days resulted in TGI of 73% (Fig 3A). Since the bioavailability of this compound was expected to be very high (data not shown), we attempted oral administration of this compound. The oral administration at 10 mg/kg once a day revealed TGI of 72% (Fig 3B). Due to the strong growth-suppressive effect on various cancer cell lines, we further investigated *in vivo* growth-suppressive effect using cancer cell lines of other types and found significant tumor growth suppression by OTSSP167 for multiple cancer types in dose-dependent manners with no or a little body-weight loss (Fig 3 and Supplementary Fig. S1). For example, mice carrying A549 (lung cancer) xenografts that were treated with 1, 5, and 10 mg/kg once a day of OTSSP167 by intravenous administration revealed TGI of 51, 91, and 108%, respectively (Fig 3C) and those by oral administration of 5 and 10 mg/kg once a day revealed TGI of 95 and 124%, respectively (Fig 3D). In addition, we examined DU145 (prostate cancer) and MIAPaCa-2 (pancreatic cancer) xenograft models by oral administration of 10 mg/kg once a day, and observed TGI of 106 and 87%, respectively (Fig

3E and F). To further validate the MELK-specific *in vivo* tumor suppressive effect, we examined PC-14 lung cancer cells in which MELK expression was hardly detectable (Fig 3G). Oral administration of 10 mg/kg OTSSP167 once a day for 14 days showed no tumor growth suppressive effect on PC-14 xenografts (Fig 3H), further supporting the MELK-dependent antitumor activity of OTSSP167.

OTSSP167 inhibits the phosphorylation of novel MELK substrates

To further characterize the molecular mechanism of MELK overexpression in mammary carcinogenesis and validate the functional consequence of small molecule inhibitor against MELK, we further investigated MELK substrates. Using the MELK recombinant protein, we performed *in vitro* kinase assay in combination with 2D-PAGE and identified multiple candidate spots which appeared in a MELK-specific manner. We analyzed these spots by mass spectrometry and confirmed drebrin-like (DBNL) and proteasome subunit alpha type 1 (PSMA1) to be MELK substrates by *in vitro* kinase assay using their recombinant proteins as shown in Fig 4A. We subsequently performed *in vitro* kinase assay with these two substrates to confirm the phosphorylation-inhibitory effect of OTSSP167. As shown in Fig 4B, addition of this compound in an *in vitro* kinase assay significantly suppressed the phosphorylation levels of DBNL and PSMA1, further supporting strong inhibitory effect of this compound on the MELK activity.

Phosphorylated DBNL by MELK enhances cellular invasiveness in cancer cell

DBNL is known to be an actin-binding adaptor protein that regulates the actin cytoskeleton and endocytosis[20-22]. To characterize the biological function of DBNL in human cancer, we first examined *in vivo* phosphorylation status of DBNL by western blot analysis using BT549 cells treated with Okadaic acid that can inhibit the phosphatase activity[23]. We introduced either or both of MELK and DBNL expression vectors into the cells, and detected the significant elevation of DBNL phosphorylation in the cells trasnfected with both expression vectors, compared with the cells transfected with either of the genes or the mock vector (Fig 5A). In addition, to identify the phosphorylation sites of DBNL by MELK, we performed mass spectrometry analysis in the presence or absence of MELK, and identified and confirmed Ser269 as a candidate phophorlylation site as shown in Fig 5B; the substitution of Ser269 of DBNL with an alanine completely diminished the phosphorylation by MELK, while that of Thr270 with an alanine showed no effect on the phosphorylation status. Subsequently, immunocytochemical analysis revealed drastic enhancement of membrane ruffling of the cells that were co-transfected with both DBNL and MELK (Fig 5C). Since membrane ruffling is related to tumor cell mobility and cancer metastasis[24], we performed Matrigel invasion assay (Fig 5D) and observed significantly higher invasiveness of the cells overexpressing both DBNL and MELK than the control cells or those overexpressing either of the two genes.

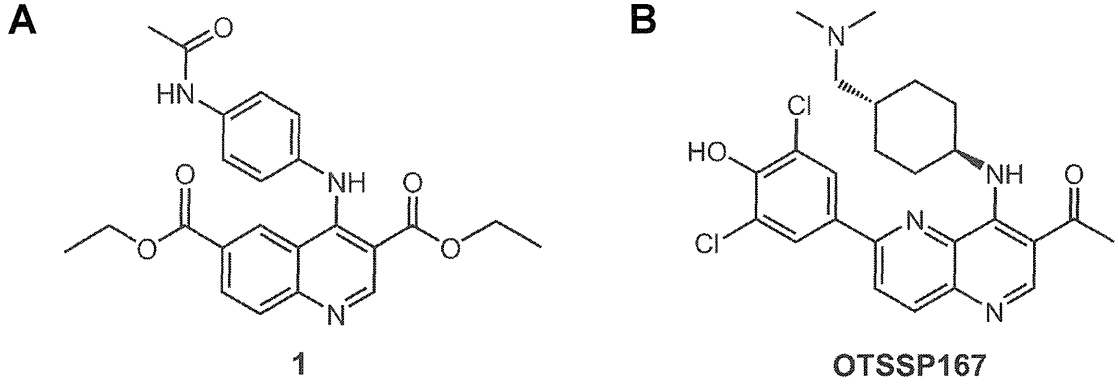

Figure 1: Novel MELK inhibitors. (A) A quinoline derivative, diethyl 4-(4-acetamidophenylamino) quinoline-3,6-dicarboxylate (compound 1; a commercially available compound), was found to have a moderate inhibitory activity against MELK (IC_{50} = 4.8 μM) through the high-throughput screening. The subsequent structure-activity relationship study led to the synthesis of a highly potent MELK inhibitor with a novel structure: (B) compound OTSSP167, 1-(6-(3,5-dichloro-4-hydroxyphenyl)-4-((*trans*-4-((dimethylamino)methyl) cyclohexyl)amino)-1,5-naphthyridin-3-yl)ethanone. The dihydrochloride salt form was used in experiments for OTSSP167.

OTSSP167 suppresses mammosphere formation through the inhibition of PSMA1 phosphorylation

PSMA1 is a subunit of the proteasome complex and was reported to be upregulated in breast cancer cells[25]. To examine the biological effect of MELK on PSMA1, we trasnfected either or both of MELK and PSMA1 into BT549 cells, and found the increase of the PSMA1 protein when PSMA1 was co-transfected with MELK, compared with the cells transfected with PSMA1 alone (Fig 6A). In concordance with this result, when we knocked-down MELK in T47D cells using siRNA, the amount of PSMA1 protein was drastically reduced, compared with the parental cells or the cells treated with control siRNA (Fig 6A), while the amount of PSMA1 transcript was unchanged (Fig 6B). These results have indicated that MELK possibly stabilizes PSMA1 protein through its phosphorylation. Since the knockdown of PSMA1 expression suppressed the proliferation of cancer cells (data not shown), the PSMA1 is also considered to be essential for survival of cancer cells. Previous studies suggested contribution of MELK in cancer stem cells due to its high level of expression in cancer stem cell populations (ex, CD133-positive glioblastoma cells) [10, 12, 13]. Our results in Fig 6C also revealed that upregulation of MELK promoted the mammosphere formation of breast cancer cells and induced the Otc3/4 expression that is well known as one of the stem cell markers while that of kinase-dead MELK (D150A) did not. Moreover, in mammosphere formation assay using MCF-7 breast cancer cells, the cells that were treated with OTSSP167 revealed stronger inhibition in its mammosphere formation than in the growth of adherent cells (Fig 6D), suggesting that OTSSP167 is likely to suppress effectively the growth of cancer stem cells. Interestingly, overexpression of PSMA1 was reported to play critical roles in hematopoietic stem progenitor cells[26]. Hence, we investigated possible involvement of PSMA1 phosphorylation by MELK in the maintenance of cancer stem cell characteristics. We performed mammosphere formation assay using MCF-7 cells which transiently over-expressed PSMA1 with either wild-type MELK or kinase-dead MELK, and found co-overexpression of PSMA1 and wild-type MELK strongly enhanced sphere formation, compared with the parental MCF-7 cells or those transfected with PSMA1 or PSMA1+ kinase-dead MELK (Fig 6E). Concordantly, the depletion of PSMA1 or MELK expression in MDA-MB-231 cells using siRNA significantly suppressed the formation of mammosphere (Fig 6F). Taken together, these results suggest that OTSSP167 suppressed mammosphere formation of cancer stem cells through the reduction of phosphorylated PSMA1 by inhibition of the MELK activity.

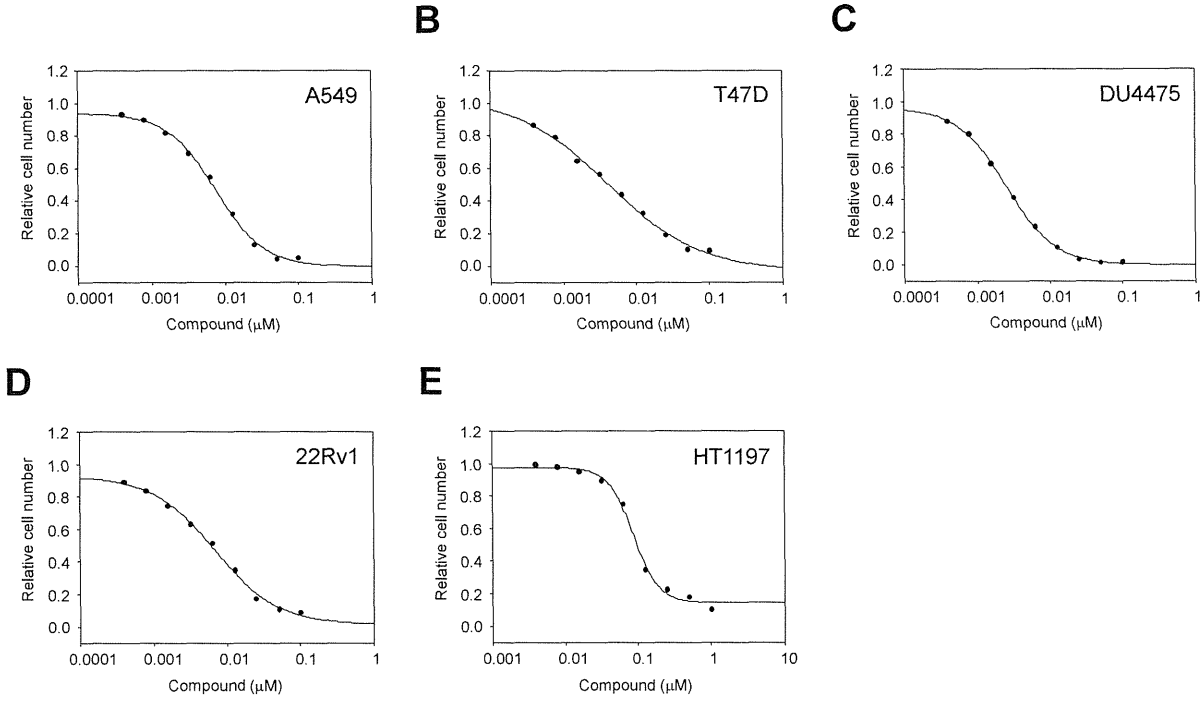

Figure 2: In vitro anti-proliferative activity of OTSSP167. The graphs indicate growth inhibition curves of OTSSP167 for various types of human cancel cell line; (A) A549 (lung cancer), (B) T47D (breast cancer), (C) DU4475 (breast cancer), and (D) 22Rv1 (prostate cancer) cells, in which MELK is highly expressed, as well as (E) HT1197 (bladder cancer) cell line , in which MELK expression is hardly detectable.

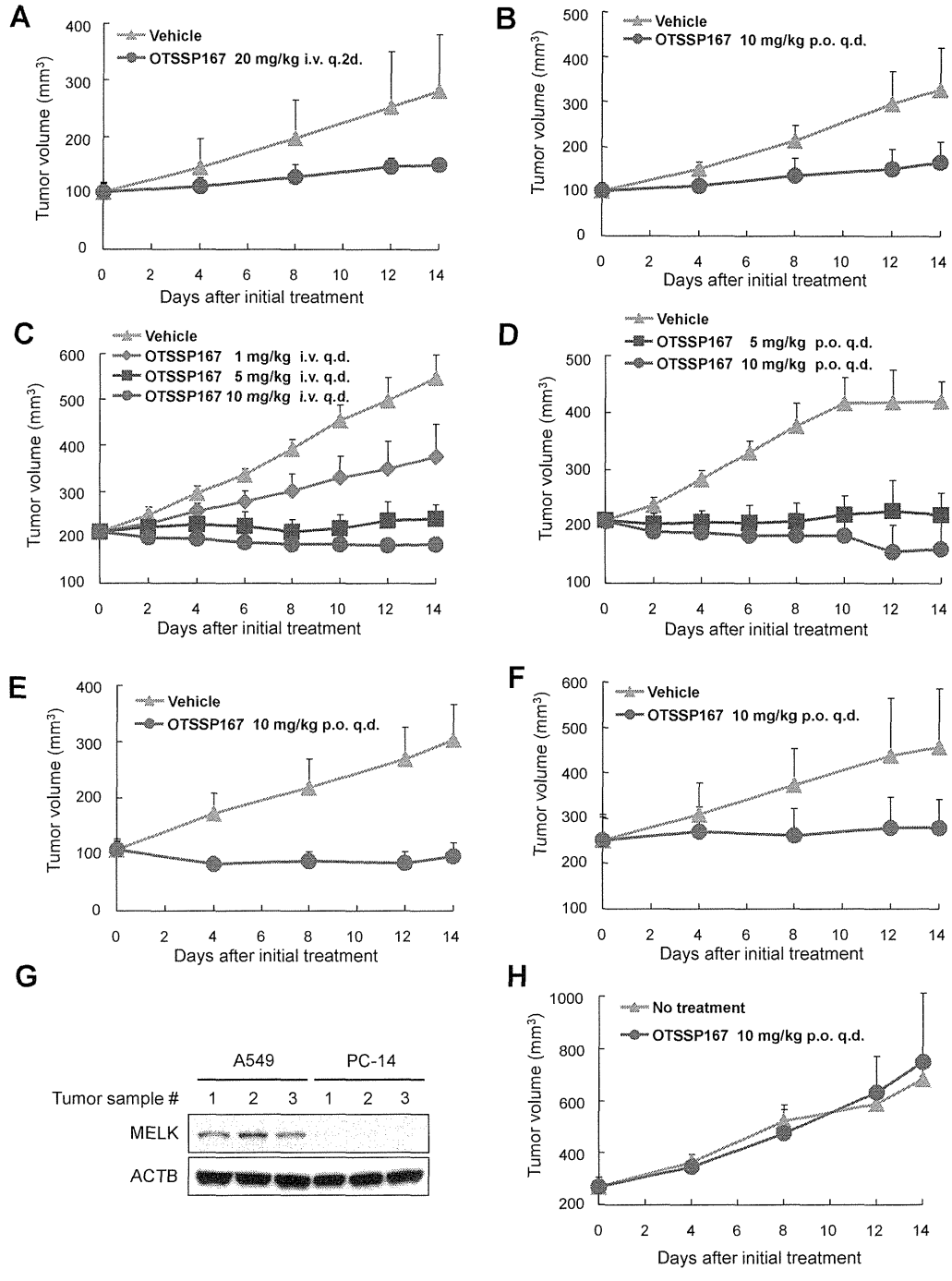

Figure 3: Mice xenograft models showing the effectiveness of OTSSP167 on the growth of various human cancer xenograft. Nude mice bearing (A,B) MDA-MB-231 (triple-negative breast cancer), (C,D) A549 (lung cancer), (E) DU145 (prostate cancer), or (F) MIAPaCa-2 (pancreatic cancer) were treated with either vehicle control or OTSSP167 of given concentrations for 14 days. The administration doses were (A) 20 mg/kg intravenously once every two days or (B) 10 mg/kg orally once a day for MDA-MB-231; (C) 1, 5, or 10 mg/kg intravenously once a day or (D) 5 or 10 mg/kg orally once a day for A549; (E) 10 mg/kg orally once a day for DU145; and (F) 10 mg/kg orally once a day for MIAPaCa-2. Mean tumor volumes ± SD ($n = 6$ for each treatment group) are shown. (G) Lysates of tumor samples taken from A549 and PC-14 xenograft mice were immunoblotted with anti-MELK and anti-ACTB antibodies. (H) OTSSP167 was administered to nude mice bearing PC-14 (MELK-negative lung cancer cells) at a dose of 10 mg/kg orally once a day. Mean tumor volumes ± SD (n = 3 per group) are shown. i.v. q.2d; intravenously once every two days, i.v. q.d.; intravenously once a day, p.o. q.d.; orally once a day.

DISCUSSION

After a great success of Imatinib in the treatment of chronic myelogenous leukemia (CML) and gastrointestinal stromal tumors (GISTs), many scientists and industries have been focusing on the development of drugs targeting on cancer-specific molecules[27]. Protein kinases are considered as attractive therapeutic targets for development of anti-cancer drugs because they play critical roles in growth-signaling pathways in cancer cells[28-31]. However, development of an inhibitor(s) which specifically suppresses target kinase activity is not so easy because most of kinase inhibitors are type 1 inhibitor which recognizes an ATP-pocket highly conserved across kinases and competes with ATP. These structural conservation leads to the unexpected cross-reactivity, in some cases yielding unexpected and unfavorable side effects[28, 32]. For discovering new kinase inhibitors with high effectiveness and minimum toxicity, the combination of identification of appropriate target molecules coupled with advanced drug-development tools including analogue synthesis, structure-informed design and fragment-based assembly is essential[28, 33].

To develop MELK-specific inhibitors in this study, we conducted the high-throughput screening for identification of hit compounds and subsequent intensive structure-activity informed study, and finally developed OTSSP167 which effectively inhibited the MELK kinase activity with IC_{50} of 0.41 nM. We then investigated the effect of OTSSP167 on the formation of mammosphere, one of the characteristics of breast cancer stem cells since MELK was reported as a key molecule for cancer stem-cell formation/maintenance[13]. Our results showed that OTSSP167 inhibited mammosphere formation in a dose-dependent manner and also revealed strong growth-suppressive effect on various types of human cancer xenograft including breast, pancreas, prostate and lung cancers without no or a little body-weight loss at the effective doses. The experiment using the MELK-negative cancer cells supported the MELK-dependent growth suppressive effect of OTSSP167 on cancer cells.

In parallel, to further characterize biological mechanisms of the MELK-signaling pathway and verify the mode of action of the MELK inhibitor OTSSP167, we screened novel MELK substrates and identified two possible candidate molecules. DBNL and PSMA1. DBNL is a member of the debrin/Abp1 family of actin-binding proteins and is a component of the immunological synapse that regulates T-cell activation[34]. Although there was no evidence of DNBL involvement in human carcinogenesis. our data have indicated that the phosphorylation of DBNL by MELK is likely to promote cancer cell invasiveness, and probably lead to tumor recurrence and poor prognosis[35]. We also found that MELK could phosphorylate Ser269 on DBNL. Since the phosphrylation of this site was reported to be critical to bind to 14-3-3 proteins[36] that has important roles in the regulation of numerous cellular signaling pathways like cell cycle regulation or apoptosis[37], we suspect that MELK might promote cell growth and mobility of cancer cells through the regulation of the DBNL-14-3-3 signaling pathway.

The other substrate, PSMA1, is one of the components of the 20S core structure of proteasome complex that is important to regulate the concentration of intracellular proteins and remove misfolded proteins through degrading them[38]. The function of PSMA1 itself was not well understood, however, its phosphorylation might affect the assembly of the proteasome complex[39]. A recent study suggested that enhancement of the proteasome assembly and activity could play crucial roles in the maintenance of human embryonic stem cells[40]. We also investigated the biological characteristics of PSMA1 in cancer cells, and found that PSMA1 was stabilized by the phosphorylation in MELK overexpressing

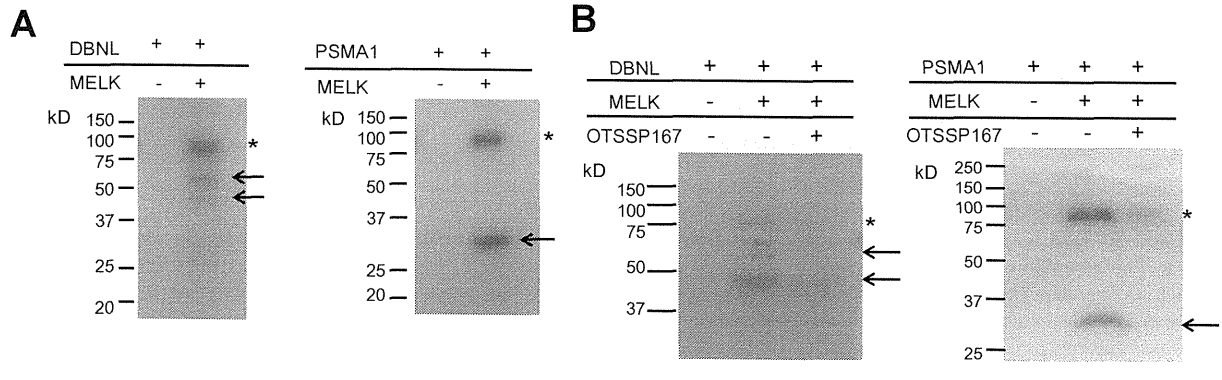

Figure 4: Identification and functional analysis of MELK substrates in breast cancer cell lines. (A) *In vitro* kinase assay using recombinant proteins confirmed DBNL and PSMA1 to be novel substrates of MELK. Arrows indicate phosphorylated substrate proteins; asterisks indicate autophosphorylated MELK. (B) *In vitro* kinase assay using OTSSP167. DBNL (55kDa) or PSMA1 (30kDa) recombinant protein was incubated with MELK with or without OTSSP167. Asterisks indicate autophosphorylated MELK; arrows indicate phosphorylated substrates. Phopshorylation of each substrate was diminished by addition of 10 nM of OTSSP167

cells and that coexistence of PSMA1 and MELK enhanced the formation of mammosphere. Interestingly, the number of mammosphere was significantly decreased in the cells in which PSMA1 expression was knocked down by siRNA for PSMA1. Our data imply that OTSSP167 possibly suppressed mammosphere formation through the reduction of PSMA1 protein.

In summary, we have demonstrated that MELK plays crucial roles in cancer progression and/or stem cell maintenance through phosphorylation of its substrate proteins. Our data have also indicated that that the selective MELK inhibitor OTSSP167 could suppress the phosphorylation of these two MELK substrates, and has the *in vitro* and *in vivo* growth suppressive effect on cancer cells, implicating a great potential of this MELK inhibitor to apply to treatment of various types of human cancer.

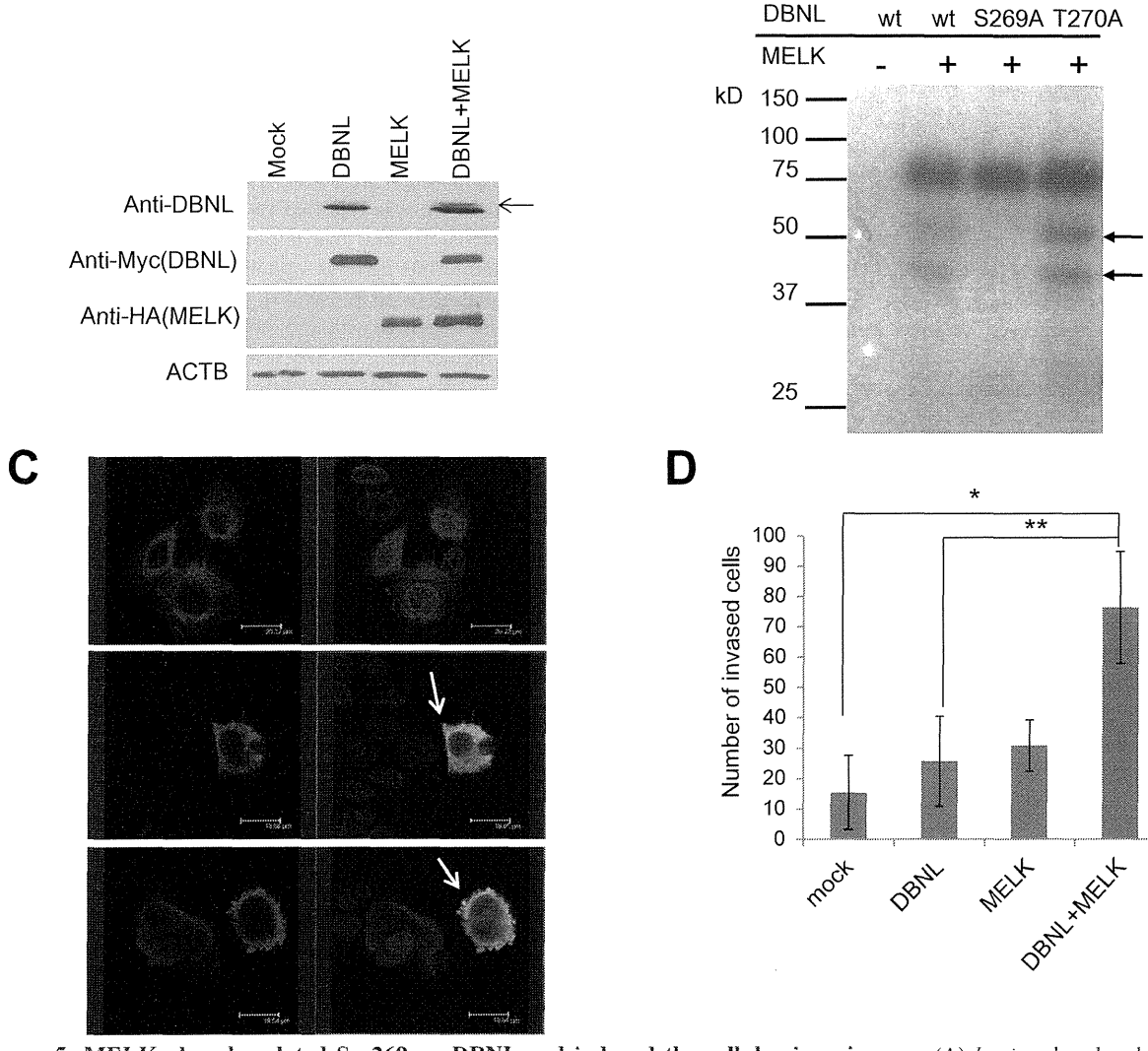

Figure 5: MELK phosphorylated Ser269 on DBNL and induced the cellular invasiveness. (A) *In vivo* phosphorylation assay. Phosphorylation of DBNL (indicated by an arrow) in BT549 cells in which DBNL and MELK were co-transfected was enhanced much stronger than that in the cells in which only DBNL was transfected. (B) Identification of phosphorylated sites by *in vitro* kinase assay. Amino acid substituted mutants of DBNL were generated and used for *in vitro* kinase assay. Phosphorylated band of DBNL in which a serine 269 was substituted with an alanine (S269A) was completely diminished, while that of DBNL in which a threonine 270 was substituted with an alanine (T270A) was unchanged. wt; wild-type. Closed arrows indicate phosphorylated DBNL. (C) Immunocytochemical analysis of cells overexpressing DBNL with/without MELK. MCF-7 cells in which both MELK and DBNL were over-expressed shows a strongly enhanced membrane-ruffling pattern (Red; DBNL, Green; MELK, Blue; DAPI) which is indicated by white arrow. (D) MCF-7 cells over-expressing DBNL revealed elevated cell invasiveness in the presence of MELK. The number of invaded cells on Y-axis indicates the average cell number of migration, that was counted by microscopic observation (*$p=0.009$, **$p=0.0209$, student's t-test). Error bars represent means ± SD of triplicates.

METHODS

High-throughput screening

A library consisting of 108,269 compounds (AMRI's Diverse AMRI Synthetic Library (DASL)) was screened using the assay protocol optimized for the high-throughput low-volume 384-well format assays. Each of the compounds (30 μM) in 342 library plates were incubated for 120 min at room temperature, with 70 μM of ATP, 100 nM of the substrate peptide, and 30 nM of MELK protein. Any plate that showed Z'<0.5 was retested (more details in Supplementary Methods).

Cell lines, plasmids, oligo siRNAs and transfection

MCF-7, MDA-MB-231, BT549, T47D, DU4475, 22Rv1, DU145, HT1197, and NIH3T3 cells were purchased from the American Type Culture Collection (ATCC) (Rockville, MD, USA). A549, PC-14, and MIAPaCa-2 cells were purchased from European Collection of Cell Cultures (ECACC) (Salisbury, UK). RIKEN BioResource Center (Tsukuba, Japan), and Japanese Collection of Research Bioresources Cell Bank (JCRB) (Suita, Japan), respectively. All cells were cultured under appropriate media recommended by suppliers with 10% FBS and 1% antibiotic-antimycotic solution (Sigma-Aldrich). All cells except MDA-MB-231 were maintained at 37 °C in humidified air with 5% CO_2. MDA-MB-231 was maintained at 37 °C in humidified air without CO_2.

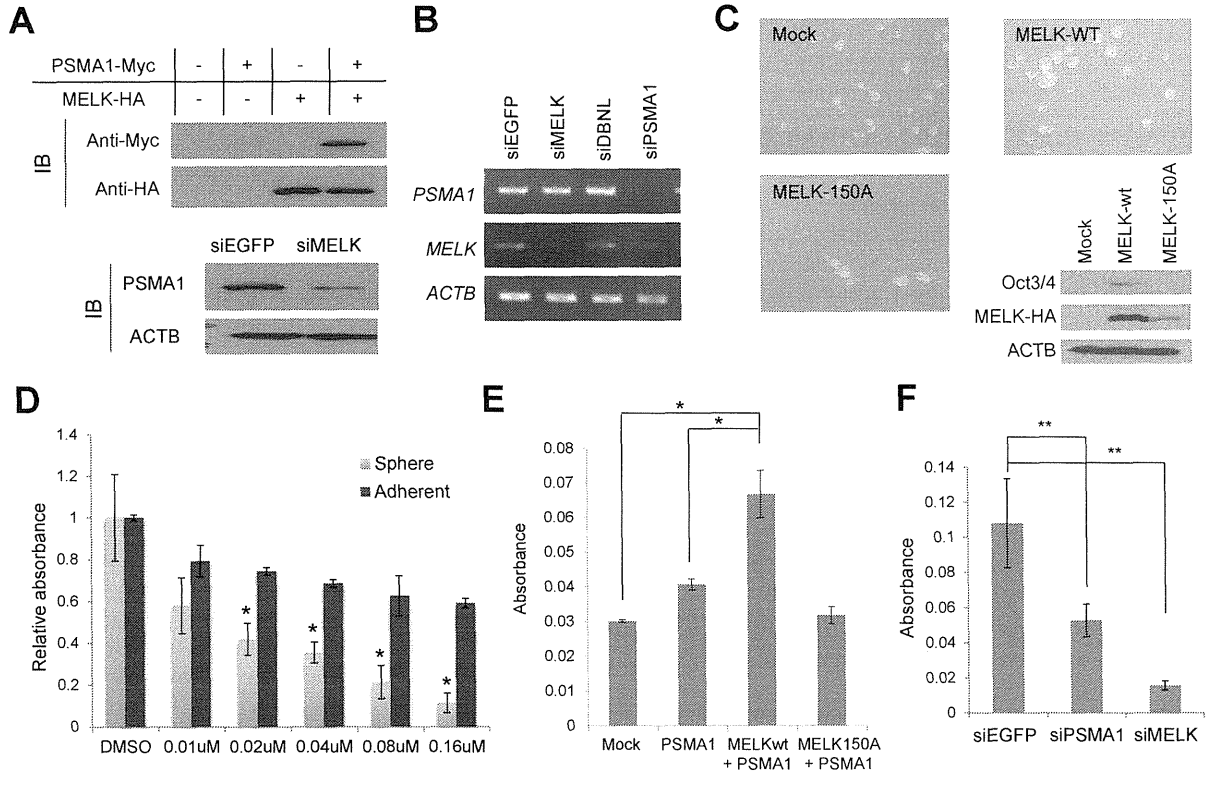

Figure 6: PSMA1 enhanced the mammosphere formation through the phosphorylation by MELK. (A, B) PSMA1 protein was stabilized through the phosphorylation by MELK in breast cancer cells (A) although transcriptional level of PSMA1 was unchanged in the cells in which MELK expression was knocked down (B). (C) Wild-type MELK (MELK-wt) or kinase-dead mutant MELK (MELK-150A) expression vector was transfected into MCF-7 cells which were seeded onto an ultra-low attachment culture plate. The formation of mammosphere was enhanced in cells in which MELK-wt was transiently introduced than those transfected with mock vector or MELK-150A. The expression levels of one of cancer stem cell markers, Oct3/4, are shown. The cells which transiently over-expressed MELK-wt induced Oct3/4 expression while those transfected with mock vector or MELK-150A revealed no Oct3/4 expression. (D) OTSSP167 suppressed more significantly the formation of mammosphere than the growth of attached MCF-7 cells. The cells were plated onto ultra-low attachment culture plate or normal culture plate without or with OTSSP167 of given concentrations (*$p<0.05$, student's t-test). (E) The MCF-7 cells in which both PSMA1 and wild-type MELK (MELK wt) were co-overexpressed revealed higher number of mammosphere formation than the parental MCF-7 cells or those transfected with PSMA1 alone or PSMA1 + kinase-dead MELK (MELK D150A) (*$p<0.0001$, student's t-test). (F) The mammosphere formation of MDA-MB-231 cells, in which PSMA1 was knocked down, was suppressed (**$p<0.05$, student's t-test). Absorbance measured at 490 nm is indicated using that at 630 nm as a reference with a microplate reader. Error bars represent means ± SD of triplicates for experiments D-F.

MELK wild-type and kinase-dead mutant (D150A) plasmids were constructed previously[9]. To construct vectors designed to express DBNL (NM_001014436.2) or PSMA1 (NM_002786.3), the entire coding sequences were amplified by RT-PCR and cloned into the pcDNA3.1-myc-his or pCAGGSnHc expression vector. We carried out site-directed mutagenesis PCR to generate DBNL substituted mutants (S269A and T270A) with a QuickChange site-directed Mutagenesis kit (Stratagene). Plasmids were transfected using Fugene6 (for NIH3T3) or FugeneHD (for human breast cancer cell lines) (Roche) according to the supplier's recommendations. For knockdown experiments, cells were transfected with oligo siRNA using Lipofectamine RNAiMAX (Invitrogen) according to manufacturer's instructions. The target sequences of oligo siRNAs were as follows: 5'-GACAUCCUAUCUAGCUGCA-3' for MELK; 5'-CAGAUACCAACACAACGAU-3' for PSMA1; 5'-GGTGCTGGCTCTGAGCACA-3' for DBNL; 5'-TTGAAGCAGCACGACUUCUUC-3' and 5'-TTGAAGAAGUCGUGCUGCUUC-3' for siEGFP.

Recombinant proteins and in vitro kinase assay for substrate screening

MELK recombinant protein was generated as described previously[9]. The full coding sequence of each of MELK substrate candidates was amplified by RT-PCR and cloned into the pGEX6p-1 vector (GE Healthcare). The GST-tagged recombinant proteins were expressed in BL21 codon-plus RIL competent cells (Stratagene) and purified using Glutathione Sepharose 4B beads (GE Healthcare) according to the supplier's instructions. The GST-tag was removed by PreScission protease (GE Healthcare) according to the supplier's instructions. For *in vitro* kinase assay, MELK recombinant protein (0.4 μg) was mixed with 5 μg of each substrate in 20 μl of kinase buffer containing 30 mM Tris-HCl (pH), 10 mM DTT, 40 mM NaF, 10 mM $MgCl_2$, 0.1 mM EGTA with 50 μM cold-ATP and 10 Ci of [γ-^{32}P]ATP (GE Healthcare) for 30 min at 30 °C. The reaction was terminated by addition of SDS sample buffer and boiled for 5 min prior to SDS-PAGE. The gel was dried and autoradiographed with intensifying screens at room temperature. OTSSP167 (final concentration of 10 nM) was dissolved in DMSO and added to kinase buffer before the incubation.

Western blot analysis and immunocytochemistry

Cells were lysed with RIPA buffer containing protease inhibitor cocktail and phosphatase inhibitor cocktail (Calbiochem). The proteins were separated by electrophoresis using 10% or 7.5% SDS-PAGE gel and transferred onto nitrocellulose membrane. The membranes were incubated with the first antibody, respectively: anti-PSMA1 antibody (Epitomics), anti-DBNL antibody, anti-Myc (Santa Cruz Biothechnology), anti-HA (Roche), anti-Oct3/4 (Santa Cruz Biothechnology) or anti-ACTB. We generated mouse anti-MELK monoclonal antibodies using partial recombinant MELK protein (264-601 amino acids of MELK) as an immunogen by the methods as described previously[41]. For immunocytochemistry, MCF-7 cells were seeded onto glass slide-chamber and transfected with expression vector(s) as described above. After 48 hours of incubation, cells were fixed with 4% paraformaldehyde and permeabilized with 0.1% Triton X-100 in PBS for 1 min at room temperature. Non-specific binding was blocked by treatment with PBS containing 3% bovine serum albumin (BSA) for 30 min at room temperature. Cells were incubated for 60 min at room temperature with anti-HA or anti-DBNL antibody diluted at 1:200 by PBS containing 3% BSA. After washing with PBS, cells were stained by Alexa fluor-conjugated secondary antibody (Invitrogen) for 60 min at room temperature, and visualized with Spectral Confocal Scanning Systems (Leica).

In vivo phosphorylation assay

DBNL expression vector was transfected into cells with or without MELK expression vector. After 48 hours of incubation, cells were treated with 100 nM Okadaic acid (Calbiochem) and incubated for 6 hours. The cells were lysed after the treatment with Okadaic acid and the lysed samples were loaded into 7.5% SDS-PAGE gel. The proteins were transferred onto nitrocellulose membrane (GE Healthcare). The membrane was incubated with anti-DBNL antibody (Abnova) or anti-β-actin (ACTB) (Sigma-Aldrich). ACTB served as a loading control.

Matrigel invasion assay and mammosphere formation assay

NIH3T3 cells transfected with plasmids expressing MELK (pCAGGSnHc-MELK), DBNL (pcDNA3.1-Myc-His-DBNL) or both were grown to near confluence in DMEM containing 10% FBS. After the incubation of 24 hours, the cells were harvested by trypsinization, washed in DMEM without addition of serum, and suspended in serum-free DMEM. The cells ($1x10^4$ cells) were seeded onto the Matrigel matrix chamber (BD Biosciences) and incubated for 22 hours.The cells invading to Matrigel were stained by Giemsa (Merck) and counted. For sphere formation assay, $1x10^3$ cells of MCF-7 cells which transiently over-expressed wild-type MELK, kinase-dead MELK, PMSA1, PSMA1 and wild-type MELK, or PMSA and kinase-dead MELK were seeded onto Ultra-Low attachment plate (Corning). For knockdown experiments, MDA-MB-231 cells ($1x10^3$ cells) which seeded onto Ultra-Low attachment plate were transfected with oligo siRNA for EGFP, MELK or PSMA1 as described above. For examination of sphere formation under treatment of MELK inhibitor OTSSP167, $1x10^3$ MCF-7 cells were seeded with 0.01, 0.02, 0.04, 0.08, or 0.16 μM of OTSSP167, respectively. DMSO alone was used as a control. Following incubation for two weeks, cell viability was measured by using Cell-counting kit-8 (DOJINDO).

In vivo xenograft study

MDA-MB-231 cells were injected into the mammary fat pads of NOD.CB17-*Prkdc*^{scid}/J mice (Charles River Laboratory). A549, MIAPaCa-2 and PC-14 cells (1×10^7 cells) were injected subcutaneously in the left flank of female BALB/cSLC-nu/nu mice (Japan SLC, Inc.). DU145 cells were injected subcutaneously in the left flank of male BALB/cSLC-nu/nu mice (Japan SLC, Inc.). When MDA-MB-231, A549, DU145, MIAPaCa-2, and PC-14 xenografts had reached an average volume of 100, 210, 110, 250, and 250 mm^3, respectively, animals were randomized into groups of 6 mice (except for PC-14, for which groups of 3 mice were used). For oral administration, compounds were prepared in a vehicle of 0.5% methylcellulose and given by oral garbage at the indicated dose and schedule. For intravenous administration, compounds were formulated in 5% glucose and injected into the tail vein. An administration volume of 10 ml per kg of body weight was used for both administration routes. Concentrations were indicated in main text and Figures. Tumor volumes were determined every other day using a caliper. The results were converted to tumor volume (mm^3) by the formula length x width^2 x 1/2. The weight of the mice was determined as an indicator of tolerability on the same days. The animal experiments using A549 xenografts were conducted by contract with KAC Co., Ltd. (Shiga, Japan) in accordance with their Institutional Guidelines for the Care and Use of Laboratory Animals. The other animal experiments were conducted at OncoTherapy Science, Inc. in accordance with their Institutional Guidelines for the Care and Use of Laboratory Animals. Tumor growth inhibition (TGI) was calculated according to the formula $\{1 - (T - T_0) / (C - C_0)\} \times 100$, where T and T_0 are the mean tumor volumes at day 14 and day 0, respectively, for the experimental group, and $C - C_0$ are those for the vehicle control group.

Statistical analysis

All values were presented as means ± SD. Statistical significance was computed using student's t-test, and the level of significance was set at $p<0.05$.

Detailed methods are described in the Supplementary Methods.

ACKNOWLEDGMENTS

We thank S. Hisada for *in vitro* experimental designs for the inhibitor discovery and manuscript preparation, and K. Toyota for technical support in *in vitro* experiments on compounds. This work was supported in part by the Innovation Promotion Program of the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

Author contributions

Y.N. planned and supervised the entire project; Y.N. and S.C. designed the study for identification of MELK substrate and functional analysis; S.C. performed most of the experiments and summarized the functional analysis results; S.C and K.K. performed subcloning and protein purification; A.T. and K.U performed mass spectrometry; Y.M. contributed to the planning of the novel MELK inhibitor discovery research and the compound design; H.S. and T.M. performed *in vitro* experiments and data analyses for the inhibitor discovery; N.T. performed *in vivo* xenograft studies; S.C., Y.N. and Y.M wrote the manuscript; Y.N obtained funding for the study.

Competing Interests

H.S., T.M., N.T., and Y.M. are employees of OncoTherapy Science, Inc. Y.N. is a stock holder and a scientific advisor of OncoTherapy Science, Inc. The other authors declare no competing interests.

REFERENCE

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Murray T, Thun MJ. Cancer statistics, 2008. CA Cancer J Clin. 2008; 58:71-96.
2. Forouzanfar MH, Foreman KJ, Delossantos AM, Lozano R, Lopez AD, Murray CJ, Naghavi M. Breast and cervical cancer in 187 countries between 1980 and 2010: A systematic analysis. Lancet. 2011; 378:1461-1484.
3. Sotiriou C, Pusztai L. Gene-expression signatures in breast cancer. N Engl J Med. 2009; 360:790-800.
4. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, El-Sayed ME, Benhasouna A, Brunet JS, Akslen LA, Evans AJ, Blamey R, Reis-Filho JS, Foulkes WD, Ellis IO. Triple-negative breast cancer: Distinguishing between basal and nonbasal subtypes. Clin Cancer Res. 2009; 15:2302-2310.
5. Griffiths CL, Olin JL. Triple negative breast cancer: A brief review of its characteristics and treatment options. J Pharm Pract. 25:319-323.
6. Cleator S, Heller W, Coombes RC. Triple-negative breast cancer: Therapeutic options. Lancet Oncol. 2007; 8:235-244.
7. Foulkes WD, Smith IE, Reis-Filho JS. Triple-negative breast cancer. N Engl J Med. 2010; 363:1938-1948.
8. Heyer BS, Warsowe J, Solter D, Knowles BB, Ackerman SL. New member of the snf1/ampk kinase family, melk, is expressed in the mouse egg and preimplantation embryo. Mol Reprod Dev. 1997; 47:148-156.
9. Lin ML, Park JH, Nishidate T, Nakamura Y, Katagiri T. Involvement of maternal embryonic leucine zipper kinase (melk) in mammary carcinogenesis through interaction with

bcl-g, a pro-apoptotic member of the bcl-2 family. Breast Cancer Res. 2007; 9:R17.

10. Gray D, Jubb AM, Hogue D, Dowd P, Kljavin N, Yi S, Bai W, Frantz G, Zhang Z, Koeppen H, de Sauvage FJ, Davis DP. Maternal embryonic leucine zipper kinase/murine protein serine-threonine kinase 38 is a promising therapeutic target for multiple cancers. Cancer Res. 2005; 65:9751-9761.
11. Bianchini G, Iwamoto T, Qi Y, Coutant C, Shiang CY, Wang B, Santarpia L, Valero V, Hortobagyi GN, Symmans WF, Gianni L, Pusztai L. Prognostic and therapeutic implications of distinct kinase expression patterns in different subtypes of breast cancer. Cancer Res. 2010; 70:8852-8862.
12. Sutter R, Yadirgi G, Marino S. Neural stem cells, tumour stem cells and brain tumours: Dangerous relationships? Biochim Biophys Acta. 2007; 1776:125-137.
13. Hebbard LW, Maurer J, Miller A, Lesperance J, Hassell J, Oshima RG, Terskikh AV. Maternal embryonic leucine zipper kinase is upregulated and required in mammary tumor-initiating cells in vivo. Cancer Res. 2010; 70:8863-8873.
14. Rich JN. Cancer stem cells in radiation resistance. Cancer Res. 2007; 67:8980-8984.
15. Dean M, Fojo T, Bates S. Tumour stem cells and drug resistance. Nat Rev Cancer. 2005; 5:275-284.
16. Kemper K, Grandela C, Medema JP. Molecular identification and targeting of colorectal cancer stem cells. Oncotarget. 2010; 1:387-395.
17. Curtin JC, Lorenzi MV. Drug discovery approaches to target wnt signaling in cancer stem cells. Oncotarget. 2010; 1:563-577.
18. Nakano I, Masterman-Smith M, Saigusa K, Paucar AA, Horvath S, Shoemaker L, Watanabe M, Negro A, Bajpai R, Howes A, Lelievre V, Waschek JA, Lazareff JA, Freije WA, Liau LM, Gilbertson RJ, et al. Maternal embryonic leucine zipper kinase is a key regulator of the proliferation of malignant brain tumors, including brain tumor stem cells. J Neurosci Res. 2008; 86:48-60.
19. Sportsman JR, Gaudet EA, Boge A. Immobilized metal ion affinity-based fluorescence polarization (imap): Advances in kinase screening. Assay Drug Dev Technol. 2004; 2:205-214.
20. Larbolette O, Wollscheid B, Schweikert J, Nielsen PJ, Wienands J. Sh3p7 is a cytoskeleton adapter protein and is coupled to signal transduction from lymphocyte antigen receptors. Mol Cell Biol. 1999; 19:1539-1546.
21. Mise-Omata S, Montagne B, Deckert M, Wienands J, Acuto O. Mammalian actin binding protein 1 is essential for endocytosis but not lamellipodia formation: Functional analysis by rna interference. Biochem Biophys Res Commun. 2003; 301:704-710.
22. Kessels MM, Engqvist-Goldstein AE, Drubin DG, Qualmann B. Mammalian abp1, a signal-responsive f-actin-binding protein, links the actin cytoskeleton to endocytosis via the gtpase dynamin. J Cell Biol. 2001; 153:351-366.
23. Cohen P, Holmes CF, Tsukitani Y. Okadaic acid: A new probe for the study of cellular regulation. Trends Biochem Sci. 1990; 15:98-102.
24. Yamaguchi K, Hata K, Wada T, Moriya S, Miyagi T. Epidermal growth factor-induced mobilization of a ganglioside-specific sialidase (neu3) to membrane ruffles. Biochem Biophys Res Commun. 2006; 346:484-490.
25. Deng S, Zhou H, Xiong R, Lu Y, Yan D, Xing T, Dong L, Tang E, Yang H. Over-expression of genes and proteins of ubiquitin specific peptidases (usps) and proteasome subunits (pss) in breast cancer tissue observed by the methods of rfdd-pcr and proteomics. Breast Cancer Res Treat. 2007; 104:21-30.
26. He X, Gonzalez V, Tsang A, Thompson J, Tsang TC, Harris DT. Differential gene expression profiling of cd34+ cd133+ umbilical cord blood hematopoietic stem progenitor cells. Stem Cells Dev. 2005; 14:188-198.
27. Zhang Z, Stiegler AL, Boggon TJ, Kobayashi S, Halmos B. Egfr-mutated lung cancer: A paradigm of molecular oncology. Oncotarget. 2010; 1:497-514.
28. Zhang J, Yang PL, Gray NS. Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer. 2009; 9:28-39.
29. Schmidt-Kittler O, Zhu J, Yang J, Liu G, Hendricks W, Lengauer C, Gabelli SB, Kinzler KW, Vogelstein B, Huso DL, Zhou S. Pi3kalpha inhibitors that inhibit metastasis. Oncotarget. 2010; 1:339-348.
30. Liu-Sullivan N, Zhang J, Bakleh A, Marchica J, Li J, Siolas D, Laquerre S, Degenhardt YY, Wooster R, Chang K, Hannon GF, Powers S. Pooled shrna screen for sensitizers to inhibition of the mitotic regulator polo-like kinase (plk1). Oncotarget. 2011; 2:1254-1264.
31. Hantschel O, Grebien F, Superti-Furga G. Targeting allosteric regulatory modules in oncoproteins: «Drugging the undruggable». Oncotarget. 2011; 2:828-829.
32. Crespo A, Zhang X, Fernandez A. Redesigning kinase inhibitors to enhance specificity. J Med Chem. 2008; 51:4890-4898.
33. Morachis JM, Huang R, Emerson BM. Identification of kinase inhibitors that target transcription initiation by rna polymerase ii. Oncotarget. 2011; 2:18-28.
34. Le Bras S, Foucault I, Foussat A, Brignone C, Acuto O, Deckert M. Recruitment of the actin-binding protein hip-55 to the immunological synapse regulates t cell receptor signaling and endocytosis. J Biol Chem. 2004; 279:15550-15560.
35. Friedl P, Locker J, Sahai E, Segall JE. Classifying collective cancer cell invasion. Nat Cell Biol. 2012; 14:777-783.
36. Li X, Wolf ME. Visualization of virus-infected brain regions using a gfp-illuminating flashlight enables accurate and rapid dissection for biochemical analysis. J Neurosci Methods. 2011; 201:177-179.
37. Hermeking H. The 14-3-3 cancer connection. Nat Rev

Cancer. 2003; 3:931-943.

38. Murata S, Yashiroda H, Tanaka K. Molecular mechanisms of proteasome assembly. Nat Rev Mol Cell Biol. 2009; 10:104-115.

39. Bose S, Stratford FL, Broadfoot KI, Mason GG, Rivett AJ. Phosphorylation of 20s proteasome alpha subunit c8 (alpha7) stabilizes the 26s proteasome and plays a role in the regulation of proteasome complexes by gamma-interferon. Biochem J. 2004; 378:177-184.

40. Vilchez D, Boyer L, Morantte I, Lutz M, Merkwirth C, Joyce D, Spencer B, Page L, Masliah E, Berggren WT, Gage FH, Dillin A. Increased proteasome activity in human embryonic stem cells is regulated by psmd11. Nature. 2012; 489:304-308.

41. Fukukawa C, Hanaoka H, Nagayama S, Tsunoda T, Toguchida J, Endo K, Nakamura Y, Katagiri T. Radioimmunotherapy of human synovial sarcoma using a monoclonal antibody against fzd10. Cancer Sci. 2008; 99:432-440.

A Comprehensive Peptidome Profiling Technology for the Identification of Early Detection Biomarkers for Lung Adenocarcinoma

Koji Ueda[1]*, Naomi Saichi[1], Sachiko Takami[2], Daechun Kang[1,3], Atsuhiko Toyama[1,3,4], Yataro Daigo[3], Nobuhisa Ishikawa[5], Nobuoki Kohno[5], Kenji Tamura[6], Taro Shuin[6], Masato Nakayama[7], Taka-Aki Sato[4], Yusuke Nakamura[3], Hidewaki Nakagawa[1]*

1 Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, Yokohama, Japan, **2** CSK Institute for Sustainability, Ltd., Tokyo, Japan, **3** Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan, **4** Shimadzu Corporation, Kyoto, Japan, **5** Department of Molecular and Internal Medicine, Hiroshima University, Hiroshima, Japan, **6** Department of Urology, Kochi University School of Medicine, Nankoku, Japan, **7** Toppan Printing Co., Ltd., Tokyo, Japan

Abstract

The mass spectrometry-based peptidomics approaches have proven its usefulness in several areas such as the discovery of physiologically active peptides or biomarker candidates derived from various biological fluids including blood and cerebrospinal fluid. However, to identify biomarkers that are reproducible and clinically applicable, development of a novel technology, which enables rapid, sensitive, and quantitative analysis using hundreds of clinical specimens, has been eagerly awaited. Here we report an integrative peptidomic approach for identification of lung cancer-specific serum peptide biomarkers. It is based on the one-step effective enrichment of peptidome fractions (molecular weight of 1,000–5,000) with size exclusion chromatography in combination with the precise label-free quantification analysis of nano-LC/MS/MS data set using Expressionist proteome server platform. We applied this method to 92 serum samples well-managed with our SOP (standard operating procedure) (30 healthy controls and 62 lung adenocarcinoma patients), and quantitatively assessed the detected 3,537 peptide signals. Among them, 118 peptides showed significantly altered serum levels between the control and lung cancer groups ($p<0.01$ and fold change >5.0). Subsequently we identified peptide sequences by MS/MS analysis and further assessed the reproducibility of Expressionist-based quantification results and their diagnostic powers by MRM-based relative-quantification analysis for 96 independently prepared serum samples and found that APOA4 273–283, FIBA 5–16, and LBN 306–313 should be clinically useful biomarkers for both early detection and tumor staging of lung cancer. Our peptidome profiling technology can provide simple, high-throughput, and reliable quantification of a large number of clinical samples, which is applicable for diverse peptidome-targeting biomarker discoveries using any types of biological specimens.

Citation: Ueda K, Saichi N, Takami S, Kang D, Toyama A, et al. (2011) A Comprehensive Peptidome Profiling Technology for the Identification of Early Detection Biomarkers for Lung Adenocarcinoma. PLoS ONE 6(4): e18567. doi:10.1371/journal.pone.0018567

Editor: Richard C. Willson, University of Houston, United States of America

Received November 16, 2010; **Accepted** March 4, 2011; **Published** April 12, 2011

Copyright: © 2011 Ueda et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by Japanese Ministry of Education, Culture, Sports, Science and Technology (Wakate-A, 22680064, http://kaken.nii.ac.jp/ja/p/22680064). This study was also funded by Shimadzu Corporation, CSK Institute for Sustainability, Ltd., Toppan Printing Co., Ltd. As employers of ST, AT, TS, or MN in this study, these funders did play a role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: ST is an employee of CSK Institute for Sustainability, Ltd. AT and TS are employees of Shimadzu Corporation. MN is an employee of Toppan Printing Co., Ltd. They contributed to the technical support and data analysis of this manuscript. The companies listed above also funded for this study, since this collaborative work was performed in "the academic-industrial alliance project for the development of lung cancer early detection system" among RIKEN, the University of Tokyo, Shimadzu Corporation, Toppan Printing Co., Ltd., and CSK Institute for Sustainability, Ltd. This does not alter the authors' adherence to all the PLoS ONE policies on sharing data and materials.

* E-mail: k-ueda@riken.jp (KU); hidewaki@ims.u-tokyo.ac.jp (HN)

Introduction

Lung cancer is the leading cause of cancer death worldwide [1]. Smoking is still the leading risk factor for lung cancer, but recently the proportion of never smoker-related lung cancer is significantly increasing, although its cause or other risk factor(s) is unknown [2]. Lung cancer patients show the poor prognosis with an overall 5-year survival rate of only 15% [3]. One of the reasons for this dismal prognosis is no effective tools to detect it at an early stage and in fact only 16% of patients are diagnosed at their early stage of the disease [3]. Current screening methods such as chest X-ray or cytological examination of sputum have not yet shown their effectiveness in the improvement of mortality of lung cancer, whereas low dose helical CT have been proved to possess a potential to detect early-stage lung cancer and demonstrate 20% lower lung cancer mortality rate compared to chest X-ray screening [4]. On the other hand, serum biomarkers for lung cancer have been investigated to achieve early detection of the disease and improve clinical management of patients [5]. Nonetheless, their present clinical usefulness remains limited [6,7]. CEA (carcinoembryonic antigen) and CYFRA (cytokeratin 19 fragment) are elevated in sera in a subset of lung cancer patients, and are clinically applied to monitor the disease status and evaluate the response to treatments. However, they are not recommended to use in clinical diagnosis and screening [8] because they are also elevated in certain non-cancerous conditions

such as smoking and lung inflammation as well as in patients with other types of cancers. It is obvious that CEA and CYFRA do not have the sufficient power to apply for the screening of early-stage lung cancer. Hence, development of novel serum/plasma biomarkers applicable for lung cancer diagnosis is urgently required.

Recently monitoring the protein expression pattern in clinical specimens by proteomics technologies has offered great opportunities to discover potentially new biomarkers for cancer diagnosis. Various proteomic tools such as 2D-DIGE, SELDI-TOF-MS, protein arrays, ICAT, iTRAQ and MudPIT have been used for differential analysis of biological samples including cell lysates and blood to better understand the molecular basis of cancer pathogenesis and the characterization of disease-associated proteins [9]. In order to explore putative biomarkers in complicated biological samples, focused proteomics or targeted proteomics technologies have been utilized such as; phosphoprotein enrichment technologies IMAC [10], the cell-surface-capturing (CSC) technology [11,12], glycan structure-specific quantification technology IGEL [13]. Most recently, to identify novel lung cancer biomarkers, Ostroff *et al.* reported the aptamer-based proteomic technology targeting 813 known proteins. Finally they selected 12 proteins which discriminated NSCLC from controls with 89% sensitivity and 83% specificity [14]. Thus targeted proteomics technologies such as the aptamer method would be applicable for the measurement of already known proteins, however could not be applied for the discovery of biomarkers targeting unknown proteins, post translational modifications, or biologically-processed polypeptides.

These methods can circumvent the technological limitations that currently prohibit the sensitive and high-throughput profiling of, in particular, blood proteome samples because of its high complexity and large dynamic range of proteins. The peptidome profiling technology addressed in the present study is one of the focused proteomics approaches targeting on biosynthetic fragments of proteins/peptides in blood, involving bioactive peptides and those non-specifically degraded by proteases or peptidases [15,16].

So far more than 500 proteases/peptidases are known to be expressed in human cells [17,18]. They function at almost all locations in the body including intracellular region, extracellular matrices, and in blood, involved in activation of other protein functions, degradation of cellular proteins, and notably tumor progression or suppression [19,20,21]. Indeed many matrix metalloproteases are overexpressed in various types of tumor cells, that facilitate construction of favorable micro-environment for tumor cells or promotion of metastasis[21]. Definitely these protease/peptidase activities should result in the production of digested peptide fragments well reflecting the tumor progression or tumor-associated responses. Thus peptidomic profiling of human serum or plasma is a promising tool for the discovery of novel tumor markers.

In this article, we extracted peptidome fractions (molecular weight <5,000 Da) from 92 individuals using the well-established and reproducible one-step peptidome enrichment method based on size exclusion chromatography (SEC) [22,23] and provided them to the label-free mass spectrometric quantification analysis combined with the statistical analyses on Expressionist proteome server platform. Our rapid and simple peptidome enrichment procedure can circumvent both less reproducible peptidome extraction by such as ultrafiltration spin filters and prolonged sample preparation including immuno-depletion column chromatography, denaturing proteins, buffer exchange, ultrafiltration, and so on [16]. After quantitative comparison of 3,537 serum peptides among 92 cases in the lung cancer biomarker discovery, we further evaluated the accuracy of quantification results by another more reliable quantification method MRM (multiple reaction monitoring) technology using independently prepared 96 serum samples.

Materials and Methods

Serum samples

All human serum samples were obtained with informed consent from 122 patients with lung adenocarcinoma (stage I to IV) at Hiroshima University Hospital at the examination on admission. Serum samples as normal controls were also obtained with informed consent from 30 healthy volunteers who received medical checkup at Hiroshima NTT Hospital and 36 from Kochi University Hospital. Each consent above was given in writing. To circumvent undesirable degradation of proteins and peptides, all serum samples were collected and stored under unified SOP. Briefly, all venous blood specimens were collected with vacuum blood collection tubes TERUMO VP-P070K (TERUMO, Tokyo, Japan). After staying upright at ambient temperature for 60 minutes, serum fractions were separated with centrifugation at 1500 x *g* for 15 min (4°C) and immediately stored at −80 °C. One freeze-and-thaw procedure was permitted for any serum samples used in the present study. This study was approved by individual institutional ethical committees; The Ethical Committee of Yokohama Institute, RIKEN (Approval code: Yokohama H20-12), The Ethical Committee of Hiroshima University Hospital, and The Ethical Committee of Kochi University Hospital.

Heat inactivation of sera and subsequent peptidome enrichment

All serum samples were freezed and thawed once and immediately incubated at 100 °C for 10 minutes after 4 times dilution with proteomics grade water. Following filtration with Spin-X 0.45 μm spin filters (Corning Incorporated, Corning, NY, USA), samples were loaded into 10/300 Superdex peptide column (GE Healthcare UK Ltd., Buckinghamshire, England) coupled with Prominence HPLC system (Shimadzu Corporation, Kyoto, Japan). The peptidome fraction was collected from 22 to 34 minutes in the constant flow of 100 mM ammonium acetate at 0.5 ml/min flow rate. The collected fractions were dried-up with Vacuum Spin Drier (TAITEC Co., Ltd., Saitama, Japan).

LC/MS/MS analysis for the screening study

The dried peptide samples were resuspended in 2% acetonitrile with 0.1% trifluoroacetic acid and analyzed by QSTAR-Elite mass spectrometer (AB Sciex, Foster City, CA, USA) combined with UltiMate 3000 nano-flow HPLC system (DIONEX Corporation, Sunnyvale, CA, USA). Samples were separated on the 100 μm×200 mm tip-column (GL Sciences Inc., Tokyo, Japan), in which L-Column beads (Chemicals Evaluation and Research Institute, Tokyo, Japan) were manually loaded, using solvent A [0.1% formic acid, 2% acetonitrile] and solvent B [0.1% formic acid, 70% acetonitrile] with the multistep linear gradient of solvent B 5 to 55% for 95 minutes and 55 to 95% for 10 minutes at a flow rate 200 nl/min. The elute was directly analyzed with the 1 second MS survey (m/z 400–1800) followed by three MS/MS measurements on the most intense parent ions (30 counts threshold, +2 - +4 charge state, and m/z range 50–2000), using the "smart exit" setting (SIDA = 3.0, max accumulation time = 2.0 sec.). Previously targeted parent ions were excluded from repetitive MS/MS acquisition for 40 seconds (100 mDa mass tolerance). The other parameters on QSTAR-Elite were shown as follows: DP = 60, FP = 265, DP2 = 15, CAD = 5, IRD = 6,

IRW = 5, Curtain gas = 20, and Ion spray voltage = 1600 V. The mass of each run was calibrated using three typical polysiloxane-derived background peaks: m/z = 445.12003, 519.13882, and 667.17640. The resolution of mass spectra was around 20,000 at m/z = 400. The primary data files (formatted as wiff and wiff.scan) from 92 clinical samples are available in a public repository site Proteome Commons (https://proteomecommons.org/). The MASCOT database search was performed on the Analyst QS 2.0 software (AB Sciex, Foster City, CA, USA). The MS/MS data was searched against the human protein database from SwissProt 57.4 (20,400 sequences) using the search parameters: Taxonomy = Homo sapiense, Enzyme = None, Fixed modifications = None, Variable modifications = Oxidation (Met), MS tolerance = 50 ppm, and MS/MS tolerance = 0.1 Da, with Mascot Automatic Decoy Search. Although Matrix Science recommends to use the Homology threshold for less-stringent criteria or Identity threshold providing almost same protein identification numbers with the criteria Expectation value <0.05 (http://www.matrixscience.com/help/interpretation_help.html), we accepted peptide identifications that satisfied both the false discovery rate (FDR) of peptide matches above identity threshold less than 5% and the Expectation value <0.05 in order to obtain more reliable identification of individual peptides than that from Mascot default criteria.

Alignment of MS chromatogram planes and peak detection on Expressionist RefinerMS

The raw data files from QSTAR-Elite (.wiff and wiff.scan formatted) were directly loaded onto the Genedata Expressionist modules (Genedata AG, Basel, Switzerland). Genedata Expressionist worked on the in-house server system HP-DL380-G5 (Hewlett-Packard Development Company, Palo Alto, CA, USA) equipped with 16 GB memory, (72 GB×2) + (146 GB×25) RAID 0+1 hard disks, and SUSE Linux Enterprise Server 10 SP2 operating system, installed with Oracle 10 g ver. 10.2.0.4. software (Oracle Corporation, Redwood Shores, CA, USA). All MS chromatograms were smoothed with RT Window = 3 scans in the Chromatogram Chemical Noise Subtraction Activity. To remove the background noise, a peak intensity is defined as follows.

$$\text{Intensity}_{\text{subtracted}} = \max(\text{Intensity}_{\text{original}} - \text{Quantile} - \text{Threshold}, 0)$$

Here, values Quantile = 50%, Intensity Threshold = 15 cps were used. Furthermore signals satisfying at least one of the following criteria were considered as noise peaks and subtracted: RT Window >50 scans, Minimum RT Length = 4 scans, or Minimum m/z Length = 8 data points. Then MS chromatogram planes derived from 92 serum samples were accurately aligned using parameters: m/z Window = 0.1 Da, RT Window = 0.2 min, Gap Penalty = 1, and RT Search Interval = 5 min in the Chromatogram RT Alignment Activity. Next, the Summed Peak Detection Activity detected the peaks on a temporary averaged chromatogram with parameters as follows: Summation Window = 5 scans, Overlap = 50, Minimum Peak Size = 4 scans, Maximum Merge Distance = 10 data points, Gap/Peak Ratio = 1, Method = curvature-based peak detection, Peak Refinement Threshold = 5, Consistency Filter Threshold = 0.8, Signal/Noise Threshold = 1. Finally the two steps Summed Isotope Clustering Activity identified isotope patterns among 2D peaks, in which peaks identified as belonging to the same isotope pattern of a molecule were grouped into peak clusters. The first clustering was performed with the following criteria: Minimum Charge = 1, Maximum Charge = 10, Maximum Missing Peaks = 0, First Allowed Gap Position = 3, Ionization = protonation, RT Tolerance = 0.1 min, m/z Tolerance = 0.05 Da, Isotope Shape Tolerance = 10.0, and Minimum Cluster Size Ratio = 1.2. The second clustering was performed with the same setting above, except for Minimum Cluster Size Ratio = 0.6 and Reuse Existing Clusters = true. The information of all detected cluster peaks, including m/z, retention time, and intensity, was exported as ABS files.

Label-free quantification and statistical analysis on Expressionist Analyst

The ABS files were loaded on the Expressionist Analyst module (Genedata AG, Basel, Switzerland). The peak intensity variation among 32 samples was normalized by fixing the median intensity of each sample at 10,000. Using the normalized intensity data, Student's t-test was performed between the normal group (n = 30) and lung cancer patients group (n = 62). The candidate biomarker peaks were extracted which showed $p<0.01$ and fold-change >5.0 between two groups. The candidates were further selected by Absent/Present Search to identify peaks with all-or-nothing detection pattern, which were detectable in 15 or all of 16 samples in one group and 1 or none of 16 samples in another group.

Multiple Reaction Monitoring

Serum samples were processed with Superdex peptide column chromatography as described above before mass spectrometric analyses. The dried peptide samples were resuspended with 1 fmol/μl BSA tryptic digest solution in 2% acetonitrile, 0.1% trifluoroacetic acid and analyzed by 4000 QTRAP mass spectrometer (AB Sciex, Foster City, CA, USA) combined with Paradigm MS4 PAL nano-flow HPLC system (AMR Inc., Tokyo, Japan). Peptides were separated on the 100 μm×100 mm tip-column (GL Sciences Inc., Tokyo, Japan), in which L-Column ODS beads (Chemicals Evaluation and Research Institute, Saitama, Japan) were manually loaded. Using solvent A [0.1% formic acid, 2% acetonitrile] and solvent B [0.1% formic acid, 90% acetonitrile], the linear gradient of solvent B 2 to 100% for 10 minutes was configured at a flow rate 200 nl/min. 19 targeted peptide ions and 5 BSA-derived peptide ions were simultaneously monitored by the MRM mode in Analyst 1.5 software (AB Sciex, Foster City, CA, USA) in duplicate. The MRM transitions are shown in **Table S4**. The acquired MRM chromatograms were then smoothened and quantified with MultiQuant software (AB Sciex, Foster City, CA, USA). MRM peak areas in each sample were normalized as follows:

$$\text{Area}_{\text{Normalized}} = \text{Area}_{\text{Raw data}} / (\text{summed area of 5 BSA standards}) \times 1000$$

Box plot analysis and ROC curve analysis

The averaged area of the duplicated MRM chromatogram peak corresponding to 19 candidate biomarker peptides was used to create box plot with R algorithm. For each study the box represents the middle half of the distribution of the data points stretching from the 25th percentile to the 75th percentile. The line across the box represents the median. The lengths of the lines above and below the box are defined by the maximum and minimum datapoint values, respectively, that lie within 1.5 times the spread of the box. Results of Student's t-test were included on the box plot. ROC curves were also depicted by R. 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 are under the curve (AUC) were shown on each graph.

Results

The efficient enrichment of peptidome fractions from sera

Since reproducible and accurate separation of the peptidome fraction from serum was essential for the effective screening of biomarkers, we optimized a simple gel filtration chromatography method (Fig. 1) and evaluated the peptide recovery. To avoid uncontrolled degradation of serum components arising from intact proteases and peptidases, all serum samples were immediately heated at 100 °C for 10 min after only one freeze-thaw procedure. Four-fold dilution of serum with water could eliminate the protein aggregation during heat inactivation even though the samples appeared slightly cloudy. Figure 2*A* shows the merged gel filtration HPLC chromatograms from 16 individual serum samples using the Superdex Peptide 10/300 column. The spectra illustrated highly reproducible separation of serum proteins and peptides. Then, the accuracy of size exclusion chromatography was assessed by analyzing 10 fractions (2 min each from retention time for the period of 14–34 min, Fig. 2*B*) with the MALDI-TOF-TOF mass spectrometer (Fig. 2*C*). As shown by the continuous MS spectra in Figure 2*C*, our gel filtration chromatography procedure allowed precise separation of serum proteins and peptides based on their molecular weights. Consequently we defined the fraction numbers 5 to 10 (corresponding to molecular weight 1,000 to 5,000) that should be focused in the further biomarker screening and validation studies.

Label-free quantification-based peptide biomarker screening for lung cancer

To explore serum peptides which could be applied for early detection of lung cancer, we acquired quantitative peptidome

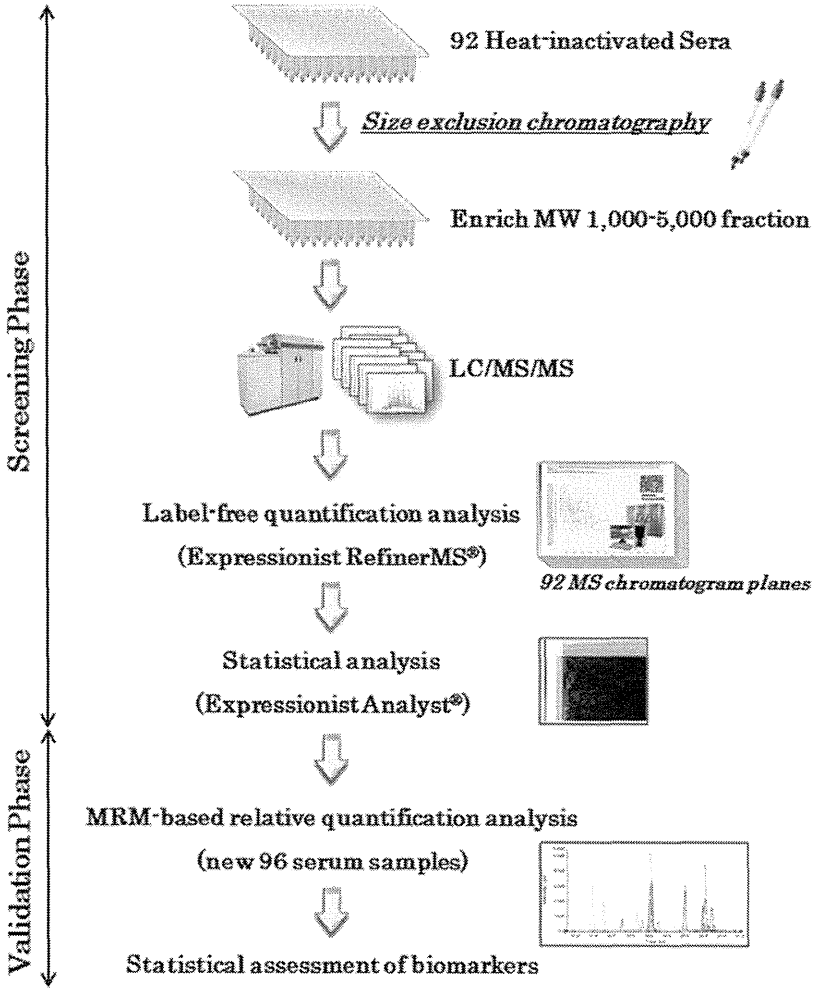

Figure 1. Schematic view of peptidome biomarker development workflow. In the screening phase, 92 serum samples were initially heat inactivated. The peptidome fractions enriched with gel filtration chromatography were analyzed with QSTAR-Elite LC/MS/MS. Following LC/MS data processing and label-free quantification analysis on the Expressionist RefinerMS module, candidate biomarkers were statistically extracted by the Expressionist Analyst module. In the validation phase, MRM experiments were performed to assess the applicability of 19 biomarker candidates using additional 96 serum samples.
doi:10.1371/journal.pone.0018567.g001

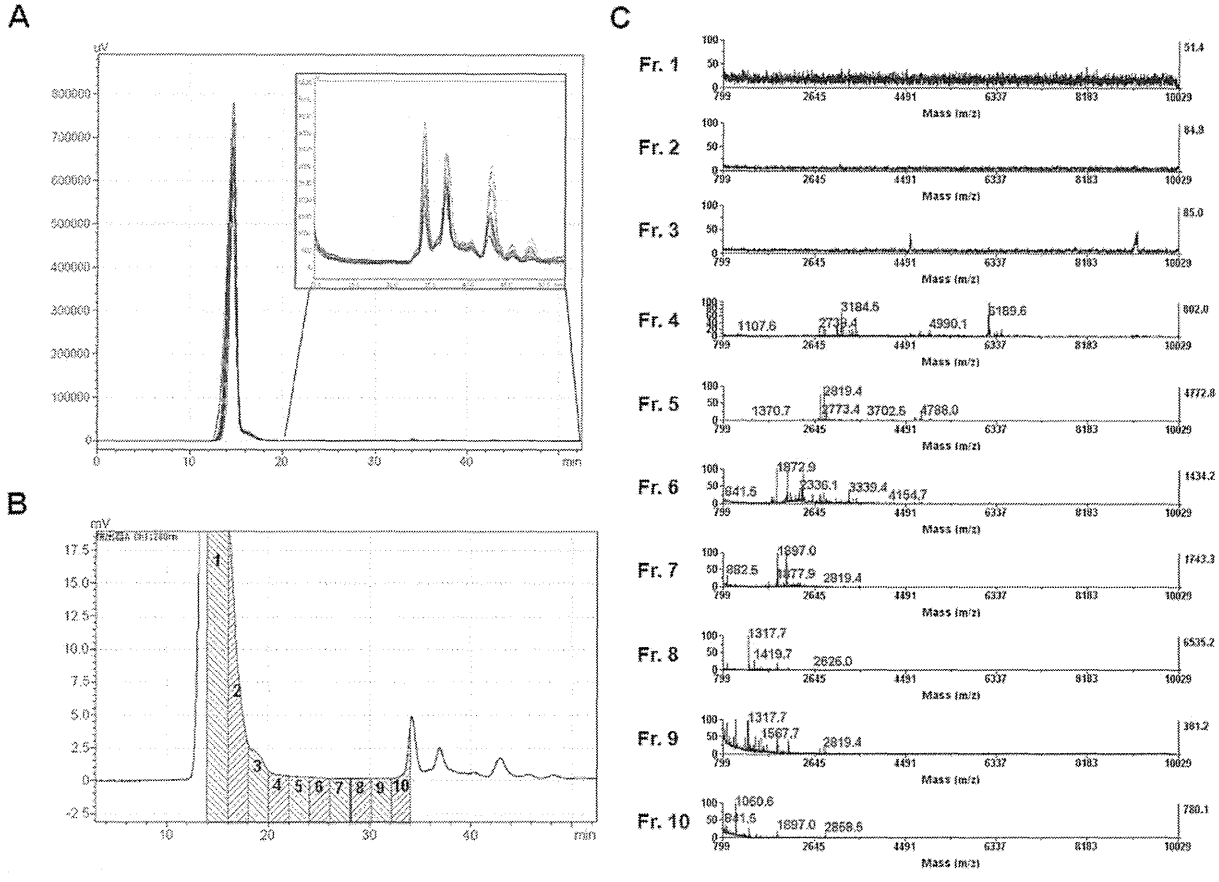

Figure 2. Simple and efficient enrichment of serum peptidome fractions by gel filtration chromatography. (*A*) The merged display of 16 independent spectra of gel filtration chromatography (280 nm UV absorbance). 10 μl each of serum sample was loaded. The upper right box shows the magnified view of the retention time range from 20 to 50 minutes. (*B*)(*C*) To evaluate the fractionation efficacy of Superdex Peptide 10/300 column, the elute was separated into 10 fractions and analyzed with MALDI-TOF mass spectrometer. The numbers of fractions in *B* correspond to the spectra numbers in *C*.
doi:10.1371/journal.pone.0018567.g002

profiles from 92 individuals (**Table S1**) including 62 lung cancer patients that consisted of 32 patients with an operable lung cancer (stage-I: n = 10, stage-II: n = 10, stage-IIIa: n = 12) and 30 lung cancer patients at an advanced stage (stage-IIIb: n = 15, stage-IV: n = 15) to identify candidate serum biomarkers for lung cancer. The serum samples were purified using gel filtration chromatography as described above and individually subjected to LC/MS/MS analyses using QSTAR-Elite mass spectrometer (Fig. 1). Subsequently 92 MS raw data were loaded and processed on the Expressionist RefinerMS module (Fig. 3*A*). Genedata Expressionist is an enterprise system for omics data management comprised of integrated software modules, which support the complete R&D processes involving data processing, statistical analysis, data management and result reporting. The technology-dependent modules for microarray data (Refiner Array), mass spectrometry (Refiner MS, used in the present study) and genomic profiling (Refiner Genome) allow highly-sophisticated data processing, quantification, visualization, and result exporting in any generally-used formats. Once all data are quantified and summarized, they can be seamlessly analyzed with the Genedata Analyst module employing various statistical analyses. This system initially made the MS chromatogram planes as shown in Figure 3*C*, and subtracted the instrument specific noises and chemical noises effectively. At the fourth step of the workflow in Figure 3*A*, the retention time (RT) grids on each MS chromatogram plane were perfectly aligned among these 92 samples (Fig. 3*B*), which allowed the solid quantification analysis of multiple samples. Subsequently, peaks were detected from temporarily averaged m/z-RT planes by the Chromatogram Summed Peak Detection Activity in order to avoid missing peak-location information even if the peaks were not detectable in particular planes. The detected isotopic peaks belonging to the same peptide signals were grouped into individual clusters that are displayed as colored rectangles in Figure 3*C*. A total of 12,396 non-redundant isotopic peak clusters with charge state +1 to +10 were detected from 92 serum samples. We then utilized 3,537 clusters with charge stage +2 to +10 for further statistical considerations in the Expressionist Analyst module, since singly-charged ions might include substantial proportion of non-peptide components such as chemicals.

Student's t-test was applied to investigate the differences in their serum levels between the normal group (n = 30) and the lung cancer group (n = 62) (Fig. 4*A*). This analysis identified 118 candidate biomarker peptides ($p<0.01$ and fold-change of >5.0). Since the criteria of t-test were variable for the purpose of candidate selection, we used the threshold above just in order to define the highest priority group. The intensity distributions of these peptides were

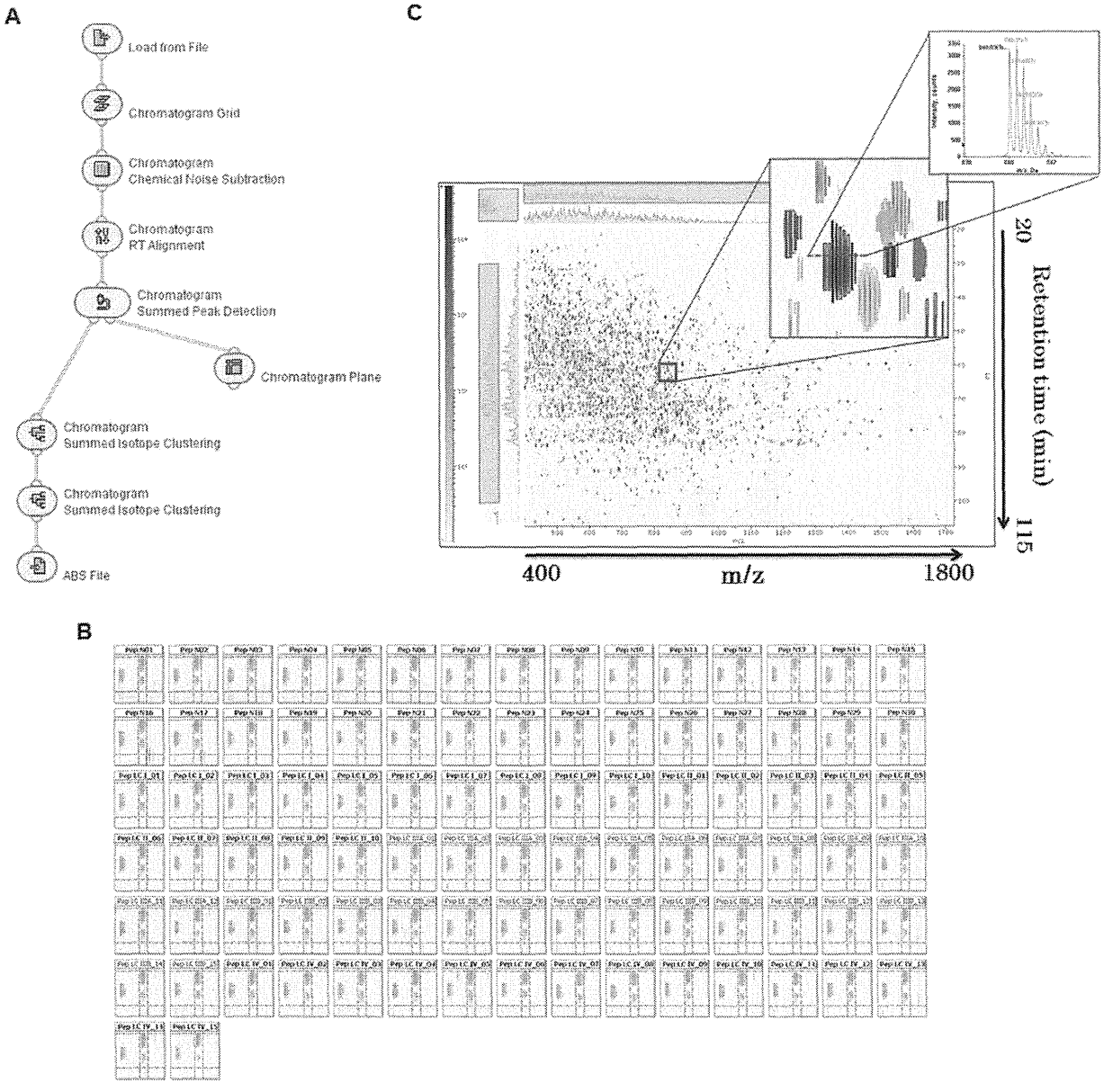

Figure 3. Rapid and accurate data processing for label-free quantification on the Expressionist RefinerMS module. (*A*) The total workflow used in the Expressionist RefinerMS module. Only 3 hours were needed to complete entire steps in this workflow on 92 LC/MS/MS data (each with 120 minutes LC gradient). (*B*) The representative area of m/z - retention time planes after RT alignment of 92 LC/MS/MS data. In each panel, three isotopic clusters and grid lines were displayed, showing highly exact alignments. (*C*) The MS chromatogram plane in which all data processing were completed. Finally, isotopic clusters derived from a single peptide were grouped into a colored cluster as shown in the middle panel. The far right panel shows the MS spectrum corresponding to the horizontal section view of a representative cluster.
doi:10.1371/journal.pone.0018567.g003

visualized with bar charts in Figure S1. The subsequent principal component analysis demonstrated that the values of 118 candidate biomarker peptides could explicitly separate control and lung cancer groups on the 3D plot using principal component 1, 2, and 3 (Fig. 4*B*). The proportion of variance described by the principal component 1, 2, or 3 was 66.9%, 15.0%, or 4.4%, respectively, indicating that illustrated components 1 to 3 could reflect 86.3% (the cumulative proportion) of quantitative information in this mass spectrometric screening analysis.

Identification of peptide sequences by LC/MS/MS

Alongside the label-free quantification-based biomarker screening described above, the comprehensive peptide sequencing was performed by a combination of QSTAR-Elite LC/MS/MS analysis and MASCOT database search. Among 230,657 MS/MS queries from 92 serum samples, 5,382 peptides were successfully sequenced with MASCOT expectation value <0.05 (FDR of peptide matches above Identity threshold was 1.49%). After examining redundancy, 424 unique peptides were

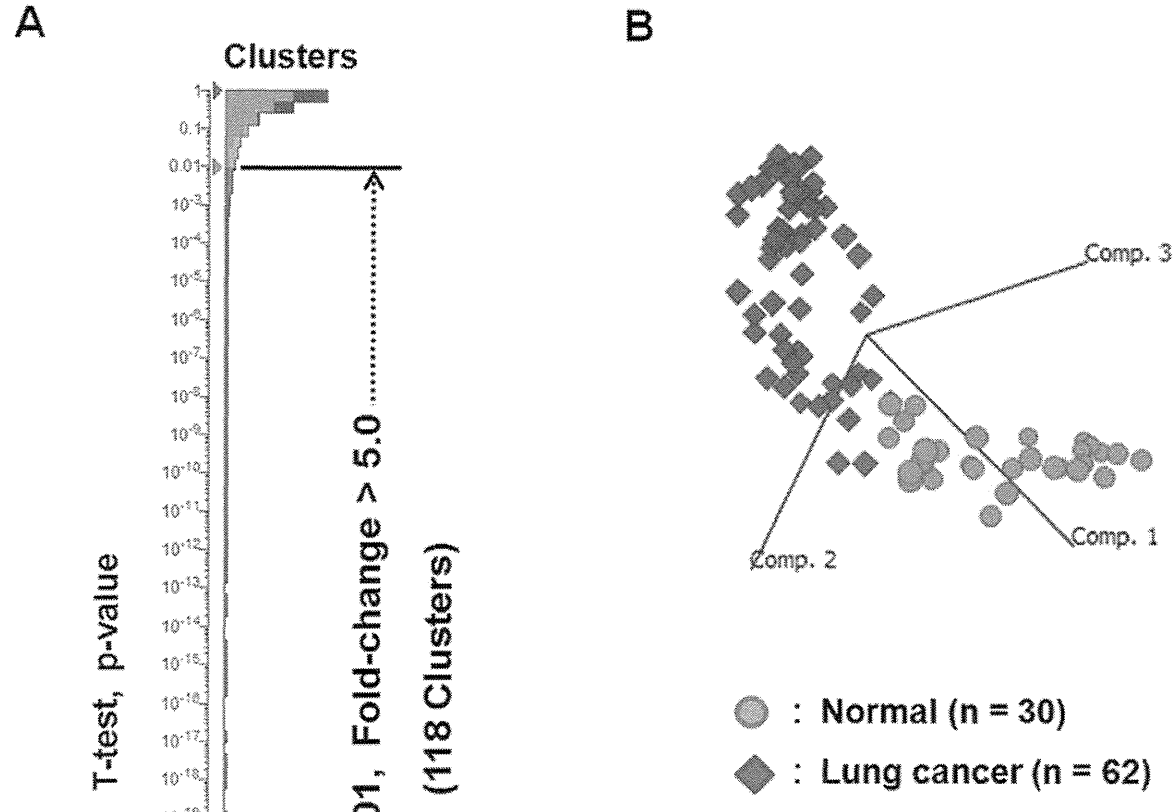

Figure 4. Statistical identification of candidate biomarkers for lung cancer. (*A*) The hierarchy chart of clusters (peptides) according to Student's t-test p-values (normal group vs. lung cancer group). 118 peptides satisfied the criteria of $p<0.01$ and fold change >5.0. (*B*) Principal component analysis using the values of 118 candidate biomarker peptides showed clear separation between control and lung cancer groups on the 3D plot. The proportion of variance described by the principal component 1, 2, or 3 was 66.9%, 15.0%, or 4.4%, respectively.
doi:10.1371/journal.pone.0018567.g004

identified that corresponded to 106 proteins (**Table S2**). Regarding the 118 candidate peptides, 19 peptides were uniquely identified; 12 of them were found to be derivatives from fibrinogen alpha chain (FIBA), 4 from apolipoprotein A-IV (APOA4), and the remaining three peptides were turned out to be a fragment of amiloride-sensitive cation channel 4 (ACCN4), apolipoprotein E (APOE), and limbin (LBN) (**Table 1**).

MRM-based validation experiment for 19 candidate biomarker peptides

To assess the quantitative reproducibility of the label-free quantification results in our single-run screening analysis, as well as the clinical usefulness of the 19 candidate biomarkers, we conducted further validation studies by multiple reaction monitoring (MRM) technology using 96 additional serum samples (**Table S1**). For designing the optimum MRM transitions specific

Table 1. 19 lung cancer biomarker candidates.

Expressionist [a]					MASCOT [b]			
Cluster ID [c]	m/z	RT	z	t-test p-value [d]	Acc. [e]	start	end	Peptide sequence
Cluster_3187	551.8	64.1	2	1.54E-15	ACCN4	613	624	CPSLGRAEGGGV
Cluster_3858	750.9	60.3	2	7.85E-04	APOA4	271	283	ELGGHLDQQVEEF
Cluster_3444	629.8	52.2	2	9.41E-07	APOA4	268	284	GGHLDQQVEEF
Cluster_3661	689.8	75.3	2	8.52E-08	APOA4	260	284	GNTEGLQKSLAELGGHLDQQVEEFR
Cluster_3498	643.3	65.7	2	6.08E-05	APOA4	288	304	SLAELGGHLDQQVEEFR
Cluster_2454	756.4	65.6	3	2.93E-03	APOE	194	214	TVGSLAGQPLQERAQAWGERL
Cluster_248	768.8	53.0	2	6.41E-23	FIBA	1	16	ADSGEGDFLAEGGGVR
Cluster_126	432.7	62.6	2	3.07E-22	FIBA	7	15	DFLAEGGGV
Cluster_159	510.7	49.9	2	5.75E-25	FIBA	7	16	DFLAEGGGVR
Cluster_240	733.3	56.5	2	3.80E-15	FIBA	2	16	DSGEGDFLAEGGGVR
Cluster_166	525.7	62.8	2	2.99E-25	FIBA	5	15	EGDFLAEGGGV
Cluster_3342	603.8	50.7	2	8.17E-27	FIBA	5	16	EGDFLAEGGGVR
Cluster_2872	461.2	61.8	2	3.31E-12	FIBA	6	15	GDFLAEGGGV
Cluster_174	539.3	52.3	2	4.10E-15	FIBA	6	16	GDFLAEGGGVR
Cluster_180	554.2	63.5	2	5.37E-22	FIBA	4	15	GEGDFLAEGGGV
Cluster_207	632.3	52.1	2	1.98E-21	FIBA	4	16	GEGDFLAEGGGVR
Cluster_196	597.8	63.2	2	4.44E-24	FIBA	3	15	SGEGDFLAEGGGV
Cluster_221	675.8	52.3	2	2.22E-22	FIBA	3	16	SGEGDFLAEGGGVR
Cluster_135	453.2	39.0	2	2.81E-24	LBN	306	313	FLLSLVLT

[a]Information acquired from the Expressionist RefinerMS or the Analyst module.
[b]Information acquired from MASCOT database search.
[c]Each ID corresponds to that in the bar chart (Fig. S1).
[d]Shown is the p-value of t-test between normal group and lung cancer group.
[e]UniProt Accession Number.
doi:10.1371/journal.pone.0018567.t001

to the 19 candidate peptides, the m/z values of precursor ions detected in the screening phase were set as Q1 channels and those of four most intense fragment ions were selected from each MS/MS spectrum for Q3 channels (Fig. S2 and Table S3). Hence, a total of 76 MRM transitions were simultaneously monitored by 4000 QTRAP mass spectrometry using a serum peptidome sample (Fig. 5). We then determined the specific eluting retention time for each candidate peptide and selected the optimum MRM transitions showing the highest MRM chromatogram peak out of four transitions for each peptide (Table S4). In our observations, only two peptides (FIBA 3–16 and FIBA 5–16) showed the identical orders of fragment ion intensities between QSTAR-Elite and 4000 QTRAP systems as shown in Figure 5. We further performed MRM-based relative quantification analysis using 36 normal controls and 60 lung cancer samples in duplicated experiments. The serum levels of 19 candidate biomarker peptides were calculated on the basis of normalized and averaged MRM chromatogram peak areas and displayed with box plots (Fig. 6*A*). To evaluate the efficacy of these candidates for early detection of lung cancer, we compared the earlier-stage lung cancer group (stage-I, II, and IIIa) with the normal group by Student's t-test. The results revealed that 15 out of 19 candidate peptides showed significant differences in their serum levels between the two groups, while 4 peptides (FIBA 4–15, FIBA 5–15, FIBA 7–15, and FIBA 7–16) showed no significant differences. Concerning the comparison between the normal group and the advanced-stage lung cancer group (stage-IIIb and IV), similarly 4 peptides (APOA4 268–284, APOA4 271–283, FIBA 5–15, and APOE 194–214) did not satisfy the criterion of $p<0.05$. Hence, we considered that the remaining 12 peptides are likely to be more promising biomarkers for lung cancer diagnosis. We next assessed the sensitivity and specificity of the 19 biomarkers for lung cancer diagnosis by ROC curve analysis (Fig. 6*B* and Fig. S3). The cut-off value was set at the point whose distance from the (sensitivity, specificity) = (1, 1) reached the minimum. Given the value of sensitivity to detect lung cancer at an earlier stage, FIBA 6–15 (87.1%), APOA4 273–283 (61.3%), FIBA 5–16 (58.1%), and LBN 306–313 (58.1%) appeared to be the good biomarker candidates. However although the specificity of APOA4 273–283, FIBA 5–16, and LBN 306–313 were remarkably higher (88.9%, 94.4%, and 100%, respectively, Fig. 6*B*), FIBA 6–15 showed relatively lower specificity (44.4%) and the area under the curve (0.641). By integrating the results from t-test and ROC curve analysis, the 3 candidates shown in Figure 6*B* were considered as the most promising peptide biomarkers for early detection of lung cancer.

Discussion

Even though recent mass spectrometry instruments have allowed measurements of peptide mixtures at high sensitivity [24], enrichment of targeted proteins/peptides is still indispensable to achieving detection and identification of serum components in limited amounts of biological materials. In this sense, the methodology to purify preanalytical samples without loss of targeted components is crucial. From this point of view, the previous peptidome profiling technologies, such as SELDI-TOF-MS coupled with ProteinChip arrays or MALDI-TOF-MS