Global and Targeted Proteomics of Secretome and Exosomes Derived from Primary and Metastatic Prostate Cancer Cell Lines

原発及び転移性前立腺癌細胞株由来の分泌型タンパク質とエクソソームの網羅的標的プロテオミクス

＜要約＞

Graduate School of Pharmaceutical Sciences, Tohoku University
Life and Pharmaceutical Science

ARMANIA NURDIN

東北大学大学院薬学研究科

生命薬科学専攻

アーマニア　ナーディン

Abstract of thesis presented to the Graduate School of Pharmaceutical Sciences, Tohoku University, in fulfillment of the requirement for the degree of Doctor of Philosophy

Global and Targeted Proteomics of Secretome and Exosomes Derived from Primary and Metastatic Prostate Cancer Cell Lines

原発及び転移性前立腺癌細胞株由来の 分泌型タンパク質とエクソソームの網羅的標的プロテオミクス

By

ARMANIA BINTI NURDIN

September 2016

[Purpose] The identification of cancer-related proteins that could be used as a biomarker to predict the aggressiveness of prostate cancer is still clinically challenging. Prostate specific antigen (PSA) is currently the only protein biomarker that is routinely used as a diagnostic tool for early detection and prostate cancer treatment monitoring. However, it remains questionable whether PSA-based screening can sensitively and selectively identify the presence and progressions status of primary and metastatic prostate cancers. Considering that the secreted proteins in the cancer secretome (all proteins released by cancer cells) enters the blood stream, there is growing interest in the study of the secretome to discover novel serum/plasma-based biomarkers. Since cancer cells-derived exosomes are believed to play an important role in cancer metastasis by facilitating the formation of pre-metastatic niche, exosomes have been also exploited as a source to identify tumor-specific biomarkers in biological fluids. To address the aggressiveness of prostate cancer, immortalized 22Rv1 cell line, established from androgen-dependent CWR22R xenograft, taken from a primary human prostatic carcinoma, was selected to represent the characteristics of primary prostate cancer. In metastatic prostate cancer, since the majority of castration-resistant prostate cancer (CRPC) patients developed bone metastasis, PC-3 cells, immortalized prostate cancer cells derived from bone metastasis, was selected to represent metastatic prostate cancer. As normal prostate model cells, PNT2 cells were selected. Therefore, the purpose of this thesis was to identify potential biomarker candidates in the secretome and exosomes of cultured primary and metastatic prostate cancer cells by using a combination of global and targeted proteomics. This study was also performed to validate the limitations and advantages of 2DICAL-data dependent acquisition (DDA) and SWATH-MS data independent acquisition (DIA) based quantitative proteomics analyses in biomarker discovery.

[Method] Secretome and exosomes released from androgen-responsive primary prostate cancer cells (P-22Rv1), androgen-irresponsive bone metastatic prostate cancer cells (M-PC-3), and non-cancerous prostate cells (N-PNT2) were collected under serum-free conditions and

further concentrated by lyophilization and a differential ultracentrifugation technique, respectively. To identify prostate cancer-related proteins, proteomic profiling of secretome and exosomes were performed by comparative 2DICAL-DDA and SWATH-MS DIA based quantitative proteomics analyses. Prior to peak alignment by 2DICAL analysis, the tryptic peptides were injected onto a nano-LC-ESI-TOF mass spectrometer in two different LC-gradients, shorter (115 min) and longer (175 min) LC-gradient lengths (n=5 injections/sample). Prior to SWATH-MS DIA, the spectral library database was constructed by repeated DDA-based analysis (n=7 injections/sample) of subcellular fractions of prostate cells-derived protein samples. SWATH-MS acquisition was carried out based on 12 Da isolation windows with a mass range of *m/z* 300-1008 (n=5 injections/sample). Protein identification was performed by assigning the resulting fragmented ions in SWATH-MS to their corresponding peptides and proteins through utilization of the constructed spectral library. Inclusion criteria for the identified peptides was applied to increase the reliability of protein quantification by global proteomic analysis. Prediction analysis, i.e. SignalP 4.1, SecretomeP 2.0 and ExoCarta database, were used to predict the cellular localization of the identified proteins. Identified proteins predicted to be secreted extracellularly were retained for further analysis. To identify potential biomarker candidates in the secretome/exosomes released from primary and metastatic prostate cancer cells, the candidates were selected from those proteins that were abundantly secreted in prostate cancer cells when compared with normal prostate cells (at least with a 2-fold increment). Selected biomarker candidates were validated by *in silico* selection selected reaction monitoring (SRM)-based targeted proteomic analysis.

[Results and Discussion] Global and targeted proteomics of prostate cancer cell secretome by 2DICAL and *in silico* selection SRM analyses: Based on the inclusion criteria, 141 and 234 proteins in total were selected as identified proteins in the secretome analysis by comparative 2DICAL-DDA analysis with shorter and longer LC-gradients, respectively. The prediction analyses revealed that more than 90.0% of these identified proteins secreted by primary and metastatic prostate cancer were predicted to be released into the conditioned medium (shorter: 130 (92.2 %); longer: 215 (91.9%)). Commonly elevated proteins in prostate cancer identified in both LC-gradients were further extracted and validated by SRM analysis. The analysis revealed that the mediator of the RNA polymerase II transcription subunit 13-like (MED13L), insulin-like growth factor binding protein 2 (IGFBP-2) and hepatocyte growth factor (HGF) were identified as highly secreted proteins from P-22Rv1 cells compared with N-PNT2 cells. Prostate-associated microseminoprotein (MSMP), proactivator polypeptide (PSAP), collagen-α-1 (VI) chain (COL6A1) and neuropilin-1 (NRP1) were identified as the predominantly secreted proteins in M-PC-3 cells compared with N-PNT2 cells. These proteins secreted in biological fluids are considered to be candidate biomarkers of primary and/or metastatic prostate cancer.

Here, the effect of LC-gradient time on protein identification and peak intensity was also examined. It was found that only 76 proteins (28.3%) were commonly identified in comparative 2DICAL analysis with both shorter and longer LC-gradients. The other proteins were specifically identified with either the shorter (54 proteins; 20.0%) or the longer (139 proteins; 51.7%) LC-gradient. It has been suggested that increase of gradient length would increase the peak capacity, resulting in better separation of eluted peptides on the column and enabling better protein identification. Higher signal intensities among commonly identified peptides (70.8%) in the secretome of P-22Rv1 were obtained with the shorter LC-gradient. For M-PC-3, two patterns of signal intensity were observed, and in the majority of cases (52.8%), a higher signal intensity was obtained with the longer LC-gradient. Similarly, the majority of the commonly identified peptides (71.5%) in the secretome of N-PNT2 showed higher signal

intensity with the longer LC-gradient. The result shows a significant decrease in the signal intensity of peptides identified in the secretomes of P-22Rv1 and N-PNT2 when measured with the longer and shorter LC-gradients, respectively. The decrease in the analyte signal could be caused by ion suppression resulting from matrix interference. As the analyte signal was decreased in both LC-gradients, the ion suppression was thought to not be entirely dependent on the gradient length, but may have also been partly due to differences in the complexity of the secretome samples.

Identification of secreted protein prostate cancer biomarker in secretome by SWATH-MS DIA and *in silico* selection SRM proteomics analyses: Since the precision and reproducibility of the protein identification in SWATH-acquisition is highly dependent on the reliability of a spectral library database, a higher coverage of spectral information should be established. Protein identification by ProteinPilot^{TM} revealed that 1087, 915, 736, 1005 and 1176 of proteins, with 99.0% confidence, were identified in the secretome, exosomes, cytosol fractions (CF), plasma membrane fractions (PMF) and whole cell lysate (WCL), respectively. After integrating all of the data obtained from 105 DDA data sets, 28253 peptides, which correspond to 2741 proteins, were identified and exported into an *in house* spectral library database, and used for targeted data analysis in SWATH-MS peak extraction. Secretome analysis by SWATH-MS DIA revealed that 305 unique proteins in total were identified in primary and metastatic prostate cancers with 97.4% (297 proteins) of them predicted to be secreted extracellularly. Among these secreted proteins, 43 of the proteins that yielded a prostate cancer/normal prostate ratio > 2, were selected for validation by SRM analysis. Quantification by *in silico* selection SRM analysis revealed that secreted IGFBP-2, HGF, growth/differentiation factor 15 (GDF-15) and vesicular integral-membrane protein VIP36 (VIP36) are potential biomarker candidates merely for primary prostate cancer. The elevated secretion of proteins, i.e. MSMP, PTX3, retinol-binding protein 4 (RBP4), COL6A1, collagen-α-2 (VI) chain (COL6A2), NRP1 and PSAP in M-PC-3 cells suggests that these proteins are potential biomarker candidates for metastatic prostate cancer. In comparison with 2DICAL-DDA analysis, secretome analysis by SWATH-MS DIA further highlighted another five secreted proteins as potential prostate cancer-associated biomarker candidates. These include GDF-15, VIP36, COL6A2, PTX3 and RBP4.

Although a higher number of proteins were identified by SWATH-MS DIA strategy (SWATH: 305 proteins; 2DICAL: 148 proteins (shorter gradient); 276 proteins (longer gradient), the present finding suggests that both acquisition strategies (SWATH and 2DICAL) are capable of identifying proteins that are abundantly secreted in the cancer secretome. In the identification of potential prostate cancer-associated tumor markers, secretome analysis revealed that MED13L and VIP36 (potential biomarker for primary prostate cancer) were selectively identified by 2DICAL-DDA and SWATH-MS DIA proteomics analyses, respectively. Both of these proteins are recognized as newly identified prostate cancer-associated tumor markers. The study shows that no secreted PSA was detected in the secretome released from prostate cancer cells, either primary cancer (P-22Rv1) or metastatic cancer (M-PC-3). In the case of metastatic cancer, this result is consistent with the previous finding that bone metastatic androgen-irresponsive prostate cancer cell line M-PC3 did not express PSA. Although P-22Rv1 cells were reported to express PSA mRNA, PSA protein secretion per cell was quite low, and in this study P-22Rv1 cells also showed low protein secretion ability when compared to other prostatic cells.

Discovery of prostate cancer biomarker candidates in exosomes by 2DICAL and *in silico* selection SRM proteomics analyses: Comparative 2DICAL-DDA analysis with shorter and

longer LC-gradients revealed that 883 and 874 exosomal proteins, respectively, were identified. Subcellular localization shows that the majority of the identified exosome proteins released by prostate cells (i.e. primary and metastatic prostate cancer) are localized in the cytoplasmic region (50.0%) and at least 90.0% of them were reported as exosomal proteins in the ExoCarta database (shorter: 792 (89.7%); longer: 793 (90.7%)). This indicates that exosomal proteins were successfully isolated from the secretome released by prostate cells using a differential ultracentrifugation technique. Among the identified exosomal proteins in both LC-gradients, 29 exosomal proteins (prostate cancer cells/normal prostate cells ratio > 10-fold) were selected for validation by SRM analysis. The analysis revealed that exosomal proteins, i.e. protein kinase C and casein kinase substrate in neurons protein 3 (PACSIN3), protein kinase C and casein kinase substrate in neurons protein 2 (PACSIN2), polyadenylate-binding protein 4 (PABP4), sorting nexin-9 (SNX9) and rho GTPase-activating protein 27 (RHG27) are potential biomarker candidates merely for primary prostate cancer. The elevated amount of secreted exosomal proteins, i.e. EGF-like repeat and discoidin I-like domain-containing protein 3 (EDIL3), integrin α-1 (ITGα1), ephrin-B1 (EFNB1), PSAP and NRP1 by metastatic cancer cells suggests that these exosomal proteins are potential biomarker candidates for metastatic prostate cancer.

Identification of biomarker candidates in the exosomes of prostate cancer cells by SWATH-MS DIA and *in silico* selection SRM proteomics analyses: The spectral library database constructed prior to secretome analysis was used in the exosome analysis. Exosome analysis by SWATH-MS DIA revealed that 1083 unique proteins were identified in the exosomes released from primary and metastatic prostate cancers. Of those proteins, 64 exosomal proteins (prostate cancer cells/normal prostate cells ratio > 50) were selected for validation by SRM analysis. Further quantification by *in silico* selection SRM analysis revealed that exosomal proteins, i.e. PACSIN3, PACSIN2, PABP4, SNX9, tetraspanin-8 (TSPAN8), protein MAL2 (MAL2), HGF and collagen-α-3 (VI) chain (COL6A3) are potential biomarker candidates merely for primary prostate cancer. In metastatic cancer cells, the elevated amount of secreted exosomal proteins, i.e. EDIL3, ITGα1, COL6A1, and tenascin-X (TNX) are noted and suggests that these exosomal proteins are potential biomarker candidates merely for metastatic prostate cancer. In comparison with 2DICAL-DDA, exosomes analysis by SWATH-MS DIA further highlighted another six exosomal proteins, i.e. TSPAN8, MAL2, HGF, COL6A3, COL6A1 and TNX, as potential prostate cancer-associated biomarker candidates.

Among these exosomal proteins, RHG27 was the only protein that selectively identified in the comparative 2DICAL-DDA analysis, with the rest of them being commonly identified in both acquisition strategies (2DICAL and SWATH method). Six exosomes proteins were recognized as newly identified prostate cancer-associated tumor markers based on exosomes analysis. These include PACSIN3, SNX9, RHG27, EDIL3, EFNB1 and TNX. The identification of common protein compositions, i.e. exosomes markers TSG101 and PDCD6IP in the cancer exosomes suggest that exosomal proteins were successfully isolated through a differential centrifugal technique used in the present study. The enrichment of secreted proteins, i.e. HGF, NRP1, PSAP and COL6A1 in prostate cancer-derived exosomes suggests that their release into extracellular space could be through an exosome-mediated secretory pathway.

In conclusion, a combination of global and targeted proteomics analyses of secretome and exosomes suggests a total of 12 secreted proteins (primary cancer: MED13L, IGFBP-2, HGF, GDF-15, and VIP36; metastatic cancer: MSMP, NRP1, COL6A1, COL6A2, PSAP, RBP4, and PTX3) and 16 exosomal proteins (primary cancer: PACSIN3, PACSIN2, PABP4, SNX9, RHG27, TSPAN8, MAL2, HGF, and COL6A3; metastatic cancer: EDIL3, ITGα1, PSAP, NRP1, EFNB1, TNX, and COL6A1), respectively, as a panel of potential prostate cancer

biomarker candidates. The distinct expression level of these proteins in different prostate cancer cell lines suggests that they could be utilized as potential biomarker candidates to predict the aggressiveness of prostate cancer.