

東海大学大学院令和二年度博士論文

Comprehensive Serum Glycopeptide
Spectra Analysis (CSGSA):
A Potential New Tool for
Early Detection of Ovarian Cancer

(網羅的血清糖ペプチド解析を用いた
卵巣癌早期診断法)

指導 三上 幹男 教授

東海大学大学院医学研究科
先端医科学専攻

林 優

This doctoral thesis was written on the basis of the original article published in “Cancers”.

I got permissions for the submission of the doctoral thesis to the institutional repository of Tokai University, in addition to reuse of part of my article in my doctoral thesis.

SUMMARY

Objectives: To conduct a comprehensive glycopeptide spectra analysis of serum between cancer and non-cancer patients to identify early biomarkers of epithelial ovarian cancer (EOC).

Methods: Approximately 30,000 glycopeptide peaks were detected from the digested serum glycoproteins of 39 EOC patients (23 early-stage, 16 advanced-stage) and 45 non-cancer patients (27 leiomyoma and ovarian cyst cases, 18 endometrioma cases) by liquid chromatography mass spectrometry (LC–MS). The differential glycopeptide peak spectra were analyzed to distinguish between cancer and non-cancer groups by employing multivariate analysis including principal component analysis (PCA), orthogonal partial least squares discriminant analysis (OPLS-DA) and heat maps.

Results: Examined spectral peaks were filtered down to 2281 serum quantitative glycopeptide signatures for differentiation between ovarian cancer and controls using multivariate analysis. The OPLS-DA model using cross-validation parameters R² and Q² and score plots of the serum samples significantly differentiated the EOC group from the non-cancer control group. In addition, women with early-stage clear cell carcinoma and endometriomas were clearly distinguished from each other by OPLS-DA as well as by PCA and heat maps.

Conclusions: Our study demonstrates the potential of comprehensive serum glycoprotein analysis as a useful tool for ovarian cancer detection.

BACKGROUND AND INTRODUCTION

Epithelial ovarian cancer (EOC) is the leading cause of gynecologic cancer mortality, and is difficult to detect at an early stage. Over 70% of women with EOC are diagnosed with advanced stage disease with a 5-year relative survival of less than 50%. CA 125 is a standard biomarker for diagnosis of EOC, however CA 125 has various problems such as false positives in patients with endometrioma, during menstruation and pregnancy. A large randomized trial found that screening using both CA 125 and TVUS at annual intervals did not reduce mortality in the general population. Human epididymis 4 (HE-4) is also a biomarker for diagnosis of EOC. However, it is suggested that HE4 provided no additional value when used alongside both ultrasound and CA 125. Current screening methods lack sufficient accuracy and thus, there is an urgent need to develop new strategies for detection of early stage EOC to improve patient survival.

We previously identified a C4-binding protein with fully-sialylated N-glycans (FS-C4BP = A2160) as a novel EOC marker. The FS-C4BP particularly distinguished early-stage ovarian clear cell carcinoma (OCCC) from endometrioma more reliably than CA-125 and higher pretreatment levels of A2160 predicts poor survival outcome. Endometriosis is a common gynecologic disease that affects 5 to 10% of women of reproductive age in the United States and associated with increased risks for EOC of endometrioid carcinoma and clear cell carcinoma types. Therefore, tumor markers that can distinguish between endometrioid cyst and ovarian clear cell carcinoma are important. In addition, A2160, as a FS-C4BP peptide with a fully sialylated and fucosylated triantennary glycan and fully sialylated biantennary glycan exhibited superior sensitivity and specificity as an EOC marker.

Common strategies for the discovery of new cancer biomarkers include metabolomics, identification of genetic and genomic markers and gene mutations by PCR and/or next generation sequencing, anomalous alternative mRNA splicing products of an expressed gene, as well gene re-expression following silencing during normal differentiation. The use of liquid chromatography with mass spectrometry (LC-MS) to identify glycoproteins released into the circulation provides a new approach

to the discovery of cancer markers. With this novel technology, we investigated not only sugar chain alterations but also combinations of protein and glycan changes.

The analysis of serum glycoproteins remains a difficult and exhausting procedure due to the large amount of data generated from the numerous peaks produced by the LC–MS method. High inter-individual variability and complexity in glycan and glycopeptide profiles makes the visual comparisons of these spectra impractical and multivariate analyses are required to determine consistent variations between data sets. Some of the statistical methods required for the multivariate analysis of the instrumental spectral results involve bilinear and multiway models, cluster analysis, scatter plots, principal component analysis (PCA), linear and non-linear partial least squares, and pattern recognition computations through the newly developed application of orthogonal partial square discrimination analysis (OPLS-DA) in scatter and scores plots.

In our previous study, we identified the glycopeptide A2160 cancer marker from among more than 100,000 glycopeptide peaks detected in the blood sera of ovarian cancer patients. During the course of detecting, isolating and identifying the A2160 glycopeptide we developed filtering procedures to find thousands of MS glycopeptide peaks that showed statistically significant differences between ovarian cancer patients and control patients. In this study, we investigated the comprehensive glycopeptide spectra analysis (CSGSA) of serum between cancer and non-cancer patients using the basic t-test, scatter plots, PCA, OPLS-DA, and heat maps in order to use the MS spectra as a possible novel approach for the early detection of ovarian cancer.

DISCUSSION

In a previous study, we used LC–MS for comparative profiling of serum glycoproteins of ovarian cancer cases and non-cancer controls and found that there were more than 30,000 glycopeptide peaks that could be used potentially as serum biomarkers for the early diagnosis of ovarian cancer. The glycopeptide A2160 was one of the first glycopeptides that we identified as a potential cancer marker and peptide sequencing revealed that it was a component of the fully-sialylated alpha-chain of the complement 4-binding protein. Moreover, we showed that the serum levels of A2160 exhibited superior sensitivity and specificity to CA-125 as an EOC marker. A2160 can be potentially applied for the monitoring of endometrioma patients, making possible the early diagnosis of OCCC.

Glycopeptide A2160 was identified by screening over 30,000 glycopeptide peaks. However, many other glycopeptide peaks showed a statistically significant difference between ovarian cancer patients and control patients, suggesting that they also could be used as markers to identify cancer patients.

The aim of the present study was to validate a LC–MS profiling method that we had used previously to determine and visually quantitate multiple serum glycopeptides as a useful tool for the detection and diagnosis of EOC. We investigated the LC–MS quantitative profiling of 2281 glycopeptides selected on the basis of statistical significance following an initial t-test of 30,000 glycopeptides detected in the serum of cancer and non-cancer patients. We used the same validated statistical and visualization methods to quantitate the signature glycopeptides in the ovarian cancers compared to two separate non-cancer controls, the endometrioma control group as well as the leiomyoma and ovarian cyst control group. In both cases the signature glycopeptides in the cancer group were highly differentiated from the control groups using various measures including PCA, OPLS-DA scatter and scores plots, and heat maps.

In this study, we did not characterize or identify individual glycopeptides. Instead we used them

together as a single group to demonstrate the potential marker signature of ovarian cancer. It is clear from the present study that there are thousands of potential glycopeptide biomarkers for ovarian cancer, but it will take considerable time to isolate, sequence and identify all of them. In the meantime, we have devised an analytical method to measure the individual concentrations of thousands of glycopeptides using a single assay, which substantially reduces analytical time and sample size. In contrast to classical biochemical approaches using immunological methods such as ELISA, RIA and EIA that mainly focus on single targets, our CSGSA involves the collation of quantitative results for a broad series of glycopeptides to reveal an overall change in blood sera of ovarian cancer patients compared to controls.

As part of our statistical and visual analysis to differentiate between the cancer and non-cancer groups, we used various measures including the volcano plot, PCA, OPLS-DA scatter and scores plots, and heat maps. The volcano plots were used initially to plot significance (t-statistic) versus fold-change on the Y and X axes respectively to compare the cancer and non-cancer samples and identify changes in the large glycopeptide data set generated by the LC-MS. From these volcano plots, we identified glycopeptides candidate EOC glycopeptide markers. The unsupervised PCA identified a set of unique spectral patterns that generally separated the cancer and non-cancer data into relatively distinct clusters. However, some of the cancer spectral data overlapped with the non-cancer data.

In comparison to PCA, the supervised OPLS-DA scatter plots clearly forced a greater scores-space separation that grouped the cancer and non-cancer data into two significantly distinct clusters. Worley and Powers used Monte Carlo simulations to demonstrate that the PCA and the OPLS-DA analyses should be performed together on the same samples to confirm reliability because the OPLS-DA analysis alone can yield statistically unreliable group separation. They found that when the PCA failed to expose group separation, OPLS-DA continued to separate the groups at the expense of model

reliability. Therefore, we used both the PCA and the OPLS-DA analyses to test and confirm the reliability of our results. We further validated the reliability of our results by computing OPLS-DA scatter and scores plots in association with the OPLS-DA prediction model.

Heat maps are another efficient way to visualize complex data sets organized as matrices. Our study is the first to report on the MS-based glycopeptides quantitative levels in ovarian cancer and non-cancer individual patients using statistical comparative analysis with heat maps. We used heat maps to visualize the quantitative glycopeptide data in individual patients as a single graphic that allowed us to easily identify the signature patterns of many glycopeptides and their quantities across multiple subjects in highly complex MS datasets comparing cancer and non-cancer patients. The relative concentrations of the glycopeptides were statistically filtered between the patients and controls and the quantitative results were displayed for each individual to reveal the most effective glycopeptide signatures. Although most of the statistically significant glycopeptides showed higher levels in the patients with cancer, some individuals had levels similar to those in the non-cancer group. These inconsistencies may have arisen from the fact that individual ratios were based on the average glycopeptide levels in the non-cancer group and not on the potentially variable glycopeptide base-line levels of each individual. However, the heat maps seem to reflect the trends of the PCA results more closely than the OPLS-DA. These sample outliers would need to be followed up in greater detail to better understand the inconsistencies between the cancer and non-cancer groups and between the PCA and OPLS-DA results.

We have used the mass spectra profiles of serum glycopeptides as multi-biomarkers for the purpose of cancer screening. Kim et al. previously had undertaken a serum multi-biomarker MS screening of ovarian cancers, but they screened for N-glycans instead of glycopeptides. No study has previously used the CSGSA for the detection of ovarian cancer. However, like Kim et al., we focused on diagnosis

rather than marker discovery and even though at this stage we do not understand the biological context of the changes, the MS signals exhibited differences between the cancer and non-cancer groups that could be exploited to improve current cancer screening procedures.

The 5-year survival rates of advanced stage of EOC patients who have metastasis in the pelvis or abdomen is less than 30%. Therefore, tumor markers that can distinguish between early stage of EOC and non-cancer patients are important. CA 125 and HE4 are common biomarker for EOC, but in our study the diagnostic utility of CA-125 and HE4 to distinguish between early stage of EOC and non-cancer.

Our study shows that CSGSA is a potentially new tool for the early detection of ovarian cancer using glycopeptides within blood serum samples that could also be applied for the detection and diagnosis of other cancers. We believe that the use of multiple biomarkers can lead to far better diagnostic classification than current single-marker approaches that depend on CA-125 or HE4. Our analysis is early and many more comprehensive studies are required to determine if the over-expressed blood serum glycopeptide markers are specific to certain cancers such as EOC or if they also markers for various other cancer and pre-cancer types. Thus, we will need to ascertain the specificity of our markers to ovarian cancer by comparing our current results in patients with ovarian cancer to those with other cancer types. We must also further address how these markers change with different stages of precancerous and cancerous lesions.

Another limitation in our study is small sample size. Generally, high-grade serous ovarian cancer is the most common histology type of ovarian cancer, and therefore, ovarian cancer screening to distinguish stage I high-grade serous ovarian cancer from benign ovarian tumors is the most clinically meaningful comparison. However, this comparison was not feasible as there were too few cases of this

subgroup to analyze in our dataset. Moreover, comparisons between histology types were also not feasible. In Japan, OCCC is a particularly common histology, presenting typically at an early-stage. As women with endometrioma possess an increased risk of OCCC, comparison of early-stage OCCC to endometrioma is more relevant to the Japanese population.

CONCLUSIONS

The Comprehensive Serum Glycopeptide Spectra Analysis has the potential to be a useful tool for the early detection of EOC. CSGSA described in this study may provide important signatures and biomarkers for ovarian cancer patients. We would like to develop an early diagnosis system for ovarian cancer and develop a follow-up system of endometrioma which are considered to be the home of ovarian clear cell carcinoma.