Proteomics Approaches for the Discovery of Potential Enzymatic Biomarkers for Early Diagnosis of Breast Cancer †

Yingxi Li 1,‡, Nico Hüttmann 2,‡, Zoran Minic 2,* and Maxim V. Berezovski 1,2,*

1 Department of Chemistry and Biomolecular Sciences, University of Ottawa, Ottawa, ON K1N 6N5, Canada; yli840@uottawa.ca
2 John L. Holmes Mass Spectrometry Facility, Faculty of Science, University of Ottawa, Ottawa, ON K1N 6N5, Canada; nhutt069@uottawa.ca
* Correspondence: zminic@uottawa.ca (Z.M.); maxim.berezovski@uottawa.ca (M.V.B.)
‡ These authors contributed equally to this work.

Abstract: Breast cancer (BC) is one of the leading causes of death in Canadian women, with an average survival rate of 5 years after diagnosis. Early detection of BC can greatly improve patient outcomes and survival. However, a non-invasive BC detection method is not currently available in clinics. Recent studies suggest that proteins from small extracellular vesicles (sEVs) could be promising biomarkers for non-invasive BC early-stage diagnosis. sEVs are membrane-enclosed vesicles secreted by cells that drive different stages of carcinogenesis in BC. The purpose of this work was to analyze different published proteomics datasets to identify enzymes that could be potentially used as diagnostic biomarkers. Three cell line studies were compared, and overlapping BC proteins were highlighted with proteins found in sEVs from blood and plasma. In total, 106 proteins were selected based on the cell line studies, of which 40 have been identified in blood/plasma sEVs. These 106 proteins were mostly enriched with cell–cell signaling and DNA repair terms based on GO analysis. Furthermore, these 40 proteins contained 11 enzymes that can be explored as potential BC biomarkers. Future validation of enzymes using cancer cell lines and blood from BC patients remains to be determined.

Keywords: breast cancer; biomarkers; enzymes; proteomics; extracellular vesicles

1. Introduction

Extracellular vesicles (EVs) are small, membrane-bound particles that are released by most cells in the body. They contain a variety of molecules, such as proteins, metabolites, and nucleic acids, and are thought to play important roles in cell–cell communication [1]. EVs can be classified into several subtypes based on their size, biogenesis, and cargo, including small extracellular vesicles (sEVs), microvesicles, and apoptotic bodies [1]. Research has shown that EVs can be involved in a range of physiological and pathological processes, such as immune regulation, tumor progression, and neurodegenerative diseases. EVs also have potential as diagnostic and therapeutic tools, as they can be isolated from biological fluids such as blood and urine, and their cargo can be manipulated for targeted delivery of therapeutic agents [2].

sEVs have emerged as promising biomarkers for cancer due to their ability to reflect the molecular signature of their cell of origin [3–5]. Cancer cells release sEVs that contain a variety of molecules, such as proteins, which can be detected in blood, urine, or other types of bodily fluids. By analyzing the proteins of these sEVs, specific biomarkers associated with cancer development, progression, and treatment can be identified. Breast cancer (BC) is one of the leading causes of death in women [3]. Early detection of BC can greatly improve patient outcomes and survival. However, a non-invasive BC detection method...
is not available in clinics yet [4]. Recent studies suggest that proteins in EVs could be promising biomarkers for non-invasive BC in early-stage diagnosis [5].

Various studies have been performed with the aim of discovering potential BC biomarkers in EVs, and MS-based proteomic analysis is one of the main approaches. For example, Risha et al. isolated EVs from a highly metastatic BC cell line and MCF-10A, a non-cancerous epithelial breast cell line [6]. The proteomic analysis of the isolated sEVs revealed 726 proteins unique to the BC cell line [6]. Minic et al. isolated sEVs and enriched phosphopeptides and performed phosphoproteomic analysis using LC-MS/MS to expand the known proteome of such sEVs. The profiling of the phosphoproteome resulted in the identification of 2003 phosphopeptides that were mapped onto 855 different proteins, encompassing a broad range of functionalities [5]. Rontogianni et al. isolated EVs from nine subtypes of BC cell lines and MCF-10A, which were used to perform both quantitative proteomic analysis and phosphoproteomic analysis. Their results reveal that EVs are subtype-specific in BC cell lines, which indicates EVs could potentially play an important role in BC subtyping in clinical diagnostics [1]. In addition to BC cell lines, researchers also isolated EVs from human plasma and serum to perform proteomic analysis. For example, Muraoka et al. isolated EVs via the affinity capture isolation method from EDTA plasma and serum and successfully identified a total of 4079 proteins by quantitative proteomics [7]. This presents the deepest proteomic study of plasma and blood EVs besides the medium EV (mEV) study by Kverneland et al. [8].

Enzyme biomarkers can be very efficient for early detection, diagnosis, therapeutic treatment, and monitoring disease recurrence of cancer patients [9]. Although enzyme biomarkers can be efficiently used for precise measurement of cancer progression, a limited number of clinically approved cancer biomarkers are available for early diagnosis [9]. Advancement of proteomic technologies enables the identification of potential biomarkers using different human cancer cell lines and blood.

The purpose of this work was to find potential BC-associated proteins identified from cell line sEV studies that can be isolated from blood or plasma sEVs and may serve as diagnostic biomarkers.

2. Methods
2.1. Data Source

The data were obtained from the supplemental material from the studies presented in Table 1.

<table>
<thead>
<tr>
<th>Publication</th>
<th>Sample Types</th>
<th>Breast Cancer Subtypes</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risha et al. [6]</td>
<td>BC cell lines</td>
<td>MDA-MB-231</td>
<td>MCF-10A</td>
</tr>
<tr>
<td>Minic et al. [5]</td>
<td>BC cell lines</td>
<td>MDA-MB-231, MCF-7</td>
<td>MCF-10A</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MCF-7, Hs578T, BT549, MDA-MB-231, LM2, HCC1954,</td>
<td></td>
</tr>
<tr>
<td>Rontogianni et al. [1]</td>
<td>BC cell lines</td>
<td>HCC1419, JIMT1, SKBR3</td>
<td>MCF-10A</td>
</tr>
<tr>
<td>Muraoka et al. [7]</td>
<td>Plasma and serum</td>
<td>Healthy human subjects</td>
<td></td>
</tr>
</tbody>
</table>

2.2. Bioinformatics Analysis

The protein groups identified from all datasets were compared based on the first protein accession ID. Contaminant and reverse proteins were removed if still present [5]. In the cell line studies, proteins identified in the control MCF10A cell line, which was used in all datasets, were pooled and subtracted from all proteins identified in more than 50% of replicates for any given BC cell line. When only quantitative data were available [1], proteins were considered to not be identified in MCF10A if their log2 LFQ intensity was one SD below the average with a significant ANOVA test [1]. Then, only proteins identified in BC cell lines were compared between the datasets and only selected if present in two out
of three studies. The resulting proteins were compared with proteins identified in blood or plasma.

Gene Ontology enrichment analysis was performed with the clusterProfiler R package using the 106 overlapping proteins against all identified proteins in cell lines. Enzyme commissions were retrieved from the org.Hs.eg.db R annotation package.

3. Results and Discussion

Usually, proteins are first identified in both non-cancerous and cancerous cell lines, and cancer-specific proteins are selected as potential biomarkers. Then, those biomarker candidates are matched in serum/plasma samples, which are among the most accessible biological samples from patients. EVs can be detected in different types of biofluids, such as blood and urine, making them an ideal subject for our study on the use of EVs for the early diagnosis of BC. In this work, we compared the proteins obtained from different studies on EVs derived from human BC cell lines. Three of them, Risha et al. [6], Minic et al. [5], and Rontogianni et al. [1], used EVs isolated from breast cancer cell lines, and Muraoka et al. [7] used EVs isolated from blood and plasma.

The three cell lines studies identified a total of 5309 proteins from both the non-cancerous epithelial breast cell line and various breast cancer cell lines reflecting different BC subtypes. Proteins that were identified in MCF-10A were removed from the total identified proteins, resulting in 831 proteins unique to the BC cell lines. This led to the identification of 106 proteins present in at least two out of the three cell line studies (Figure 1 and Table S1).

![Figure 1](image)

Figure 1. The overlap between identified sEV proteomes from three breast cancer cell line studies [1,5,6]. In total, 106 proteins were identified in at least two out of the three studies.

These 106 proteins were then compared to the proteomic data from the blood and plasma study. In total, 40 proteins overlapped, from which 11 enzymes were identified as enzymes by their enzyme commission number (Table 2). Four of these enzymes belong to the family of protein kinases: unc-51-like kinase 3, protein kinase C beta, G-protein-coupled receptor kinase 2, and glycogen synthase kinase 3 alpha. Several reports found that targeted inhibition of unc-51-like kinase 3 contributed to the inhibition of metastasis and tumor...
growth in a diverse range of cancers [10–12]. It has been found that the expression of protein kinase C beta (PKCβ) and G-protein-coupled receptor kinase 2 promotes tumorigenesis in BC [13,14]. Using MCF7 cell lines, it has been demonstrated that glycogen synthase kinase-3 protects estrogen receptor alpha from proteasomal degradation and is required for full transcriptional activity of the receptor [15]. This estrogen receptor is known to play a significant role in the formation of BC [16,17]. Two enzymes are associated with ligase in the ubiquitin pathway: thyroid hormone receptor interactor 12 and WW domain containing E3 ubiquitin protein ligase 2. Ubiquitination-related proteins (URGs) have been proposed as important biomarkers and therapeutic targets in cancer, including BC [18]. The WW domain containing E3 ubiquitin protein ligase 2 has been proposed to play a central role in tumourigenesis and has potential as a prognostic marker and molecular therapeutic target [19]. Other identified enzymes include galactosidase alpha, dipeptidyl peptidase 4, peptidylprolyl cis/trans isomerase—NIMA-interacting 1, acyl-CoA synthetase short-chain family member 2, and CTP synthase 2. All these enzymes have been reported to have different roles in cancer biology [20–24].

Table 2. List of enzymes identified in EVs from blood and plasma suggested as potential BC biomarkers. Eleven enzymes were identified in sEVs from both BC cell lines and blood and plasma. T indicates that the enzymes were identified in the proteomic data from the published BC cell lines studies.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Q14669</td>
<td>6.3.2.19</td>
<td>TRIP12</td>
<td>Thyroid hormone receptor interactor 12</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Q6PH2</td>
<td>2.7.11.1</td>
<td>ULK3</td>
<td>Unc-51-like kinase 3</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P05771</td>
<td>2.7.11.13</td>
<td>PRKCB</td>
<td>Protein kinase C beta</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P25098</td>
<td>2.7.11.15</td>
<td>GRK2</td>
<td>G-protein-coupled receptor kinase 2</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P49840</td>
<td>2.7.11.26</td>
<td>GSK3A</td>
<td>Glycogen synthase kinase 3 alpha</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P06280</td>
<td>3.2.1.22</td>
<td>GLA</td>
<td>Galactosidase alpha</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P27487</td>
<td>3.4.14.5</td>
<td>DPP4</td>
<td>Dipeptidyl peptidase 4</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>P13526</td>
<td>5.2.1.8</td>
<td>PIN1</td>
<td>Peptidylprolyl cis/trans isomerase, NIMA-interacting 1</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Q9NR19</td>
<td>6.2.1.1</td>
<td>ACSS2</td>
<td>Acyl-CoA synthetase short-chain family member 2</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>O00308</td>
<td>6.3.2.19</td>
<td>WWP2</td>
<td>WW domain containing E3 ubiquitin protein ligase 2</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>Q9NRF8</td>
<td>6.3.4.2</td>
<td>CTPS2</td>
<td>CTP synthase 2</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>

The functional enrichment analysis of the selected 106 proteins (Table S2) revealed many enriched terms, including cell–cell signaling (Table S3) and DNA repair. EVs are known to play an important role in mediating cell–cell signaling, immune response, and tumor metastasis [25]. Additionally, altered DNA repair systems are often related to an increase in the cancer rate [26]. Our findings suggest that EVs play a role in DNA repair during cancer progression.

4. Conclusions

In conclusion, the results of this study and previously reported data strongly suggest that 11 identified enzymes may be potential candidates as biomarkers for the early diagnosis of BC. In our previous report [5], we also identified and validated some enzymatic biomarkers from sEVs derived from BC cell lines. All these candidates, including those reported here, provide a feasible panel of potential biomarkers that can be tested using plasma/blood from BC patients.
Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ECB2023-14099/s1. Table S1: the overlap between identified sEV proteomes from three breast cancer cell lines studies identified in at least two out of the three studies; Table S2: gene ontology enrichment result using Biological Process terms for 106 overlapping BC proteins vs. all identified cell line proteins. Table S3: proteins that are related to cell–cell signaling function according to Gene Ontology analysis.

Author Contributions: Study design, Z.M. and M.V.B.; methodology, Y.L., N.H. and Z.M.; data analysis and bioinformatics, N.H.; writing, Z.M., Y.L. and N.H.; review and editing, M.V.B.; funding, M.V.B. All authors have read and agreed to the published version of the manuscript.

Funding: The study was supported by the John L. Holmes Mass Spectrometry Facility for the sample preparation and MS analysis and granted by Natural Sciences and Engineering Research Council of Canada (M.V.B.; grant # RGPIN-2020-05775).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.