University of Szeged
Albert Szent-Györgyi Medical School
Doctoral School of Interdisciplinary Medicine

INVESTIGATING CIRCULATING EXTRACELLULAR VESICLES FOR ENHANCED BRAIN TUMOR DIAGNOSIS AND PROGNOSIS

Thesis booklet

Gabriella Dobra

Supervisor:
Krisztina Buzás, PhD

Szeged,
2024

Introduction

Diagnosis of central nervous system (CNS) tumors

In CNS tumor diagnostics, computed tomography (CT) and magnetic resonance imaging (MRI) scans are used to determine disease status and evaluate treatment response; however, these techniques have well-known limitations. A further common method for tumor profiling is the tissue biopsy which requires invasive surgical procedures to collect tumor samples. In addition to the high risk of complications, the disadvantages of such invasive procedures include the difficulty of obtaining tumor samples from highly heterogeneous or inaccessible locations, the need for multiple biopsies in the event of metastasis, and the inability to monitor tumor response or relapse. These challenges are exacerbated if the tumor is located within the skull.

Due to these shortcomings, attention has shifted towards the molecular investigation of body fluids, also known as liquid biopsy (LB). Every single cell, including neoplastic cells, releases molecular information into the circulation, accumulating in urine, cerebrospinal fluid, saliva, and blood, thus, these body fluids can serve as a rich source of tumor-associated molecules. liquid biopsy comprises the process of detecting and measuring distinct tumor-associated materials, such as circulating tumor DNAs (ctDNA), circulating tumor cells (CTCs), circulating tumor proteins, and tumor-derived or other extracellular vesicles (EVs). Identifying these materials can be of great importance in patient care (Figure 1).

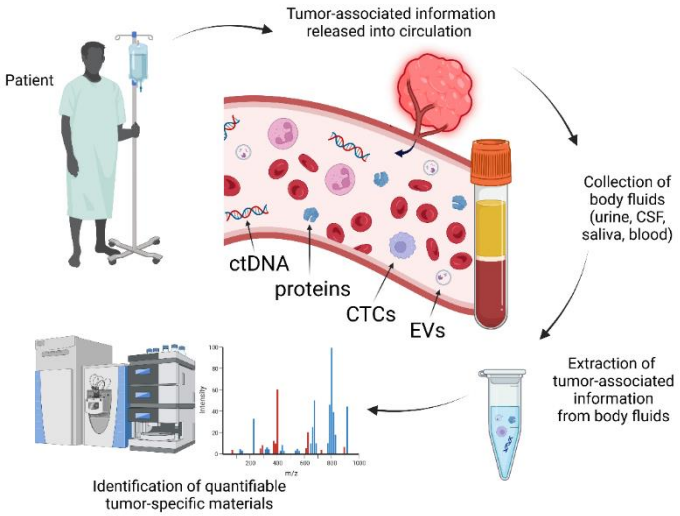

Figure 1 Schematic representation of liquid biopsy.
Abbreviations: ctDNA: circulating tumor DNA, CTCs: circulating tumor cells, proteins: tumor proteins, EVs: extracellular vesicles, CSF: cerebrospinal fluid

In our view, EVs are the most prominent LB targets, since they possess various advantages, like entering the circulatory system and offering a diverse range of diagnostically significant content. EVs have a nanometer-scale dimension, cells actively release substantial amounts of them; they are stable in all body fluids, and they can cross the blood-brain barrier (BBB). These valuable additional features make them ideal for liquid biopsies of brain tumors.

Extracellular vesicles, and their role in tumor diagnosis and prognosis

The term extracellular vesicles (EVs) endorsed by the International Society for Extracellular Vesicles (ISEV) as a "generic term for particles naturally released from the cell that is delimited by a lipid bilayer and cannot replicate, i.e., do not contain a functional nucleus". EVs contain a sample of the cytosolic milieu of the donor cells, including abundant DNA, RNA, proteins, and other analytes, while outwardly resembling the donor cells.

In the last few decades, EVs have become the target of intensive research, including their biological characteristics and physiological functions; and their role was also described in cellular senescence, immunity, and communication between cells. In the field of tumor biology, EVs have been extensively investigated and found to play a role in various aspects of the disease. These include tumor growth, development, invasion, inflammatory responses, immune suppression, epithelial-to-mesenchymal transition, angiogenesis, pre-metastatic niche formation, immunomodulation, tumor progression, and therapy resistance.

There is increasing evidence that small extracellular vesicles (sEVs) contain a wealth of information that can be used for cancer diagnosis and prognostic evaluation. Many studies focus primarily on nucleic acid content, but the EV proteome is also gaining increasing attention in cancer diagnostics. Considering the number of proteins that can be produced by cells (which varies by time and stress), it seems that studying the proteome provides a greater understanding of the complexity of the tumor tissue and the evolution of tumors than studying the genetic material alone.

MMPs in tumor invasion

The lack of effectiveness in treating certain aggressive brain tumors, such as glioblastoma, can be attributed to the challenging task of completely removing the tumor due to the presence of peritumoral infiltration, and invasion of malignant cells into the surrounding healthy tissue.

Tumor invasion is predominantly facilitated by extracellular matrix (ECM) molecules (e.g. brevican, cadherin, fibronectin, neurocan, versican, tenascin-C, and matrix metalloproteinases (MMPs); consequently, the ECM is significantly involved in the

tumor progression. MMPs are zinc-dependent endopeptidases that play an essential role in cell physiology by degrading the ECM components such as collagens, which are pivotal for normal physiological processes like wound healing, embryonic development, and tissue remodeling.

In the context of pathology, MMPs are implicated in numerous processes. They enable cancer cells to breach the basal membrane barriers, facilitating metastasis and promoting cancer cell migration and invasion. Additionally, MMPs are involved in angiogenesis, which is essential for tumor growth as it provides the necessary blood supply for expanding tumors.

MMPs are, therefore, key players in the process of tumorigenesis, although, in addition to it, several proteins are supposed to be involved in tumor progression. The identification of these tumor-associated molecules by LB can be of great value in the diagnosis and monitoring of brain tumor patients.

Justification of the research

Non-invasive diagnostic tests are of great importance, especially in CNS tumors, because of their minimal burden and risk to the patient, their repeatability, low cost, high information content, and easy accessibility. Various papers report gene or protein expression analysis of CNS tumor tissue (mainly gliomas) allowing the identification of biomarkers secreted into the blood, however, there is no reliable circulating protein biomarker for CNS tumors in clinical practice.

Since previous attempts to find definite serum markers for brain tumors failed when they were based on a single or only a few candidate factors, the solution may lie in identifying 10-20 candidate markers associated with brain tumors and creating a protein panel as a characteristic fingerprint of CNS diseases.

Another issue is that the potential biomarker candidates are difficult to detect because of their low concentrations in the blood, and the high levels of abundant serum proteins which may confound the analytical measurements. Hence, it is necessary to refine the sample to increase the concentration of potential biomarker candidates beyond the detection limit. This biomarker enrichment may be achieved by sEV isolation, but it is necessary to compare the two sample types (whole serum and serum-derived sEV) to determine which is more suitable to distinguish between patient groups.

As the content of EVs varies based on the disease progression, the feasibility of utilizing sEVs for prognostic purposes is also an important issue. Hence, it is important to examine whether any of the proteins exhibiting the most significant differences among patient groups are appropriate for prognostic markers.

Aims

We aimed to identify circulating protein markers for liquid biopsy-based diagnosis, prognosis, and monitoring of central nervous system tumors utilizing patient-derived sera (later referred to as whole serum) and serum-derived extracellular vesicles (later referred to as serum sEV).

Therefore, we performed comprehensive quantitative and qualitative analysis on the protein content of whole serum and serum sEV using LC-MS, ELISA, and comparative statistical analyses. Four groups of patients with a disease affecting the central nervous system were included in the tests (glioblastoma, brain metastasis of non-small cell lung cancer, meningioma, and lumbar disc herniation as controls), a total of 222 patients were examined during this proteomic study.

Our specific aims were:

1. To compare the protein content of the whole serum and the serum sEV of four patient groups and identify non-invasive protein biomarkers by establishing characteristic protein panels that are suitable for classifying CNS tumors;
2. To compare the potential of the two proteome and the established protein panels in tumor classification and biological characterization of the disease;
3. To explore the factors underlying the differences in classification efficiency between the two sample types;
4. To select the most sufficient sEV protein for distinguishing the patient groups;
5. To examine the factors (gender, age, and clinical history of the patient) influencing the concentration of the protein biomarker candidates;
6. To determine the clinical potential of the most appropriate protein as a prognostic biomarker.

.

Materials and Methods

Patient cohorts

Blood samples of 222 patients treated in the Department of Neurosurgery, University of Debrecen were analyzed. Samples were obtained from patients with primary glioblastoma multiforme (GBM), meningioma (M), and single brain metastasis originating from non-small-cell lung cancer (BM). Control samples (CTRL) were collected from patients with spinal disc herniation without evidence of cancer.

Blood samples were stored by the Neurosurgical Brain Tumor and Tissue Bank of Debrecen according to the criteria of the National Research Ethics Committee. An informed consent form was signed by each patient; the study was conducted in accordance with the Declaration of Helsinki. Studies were carried out according to two ethical approvals, namely 51450-2/2015/EKU (0411/15), 121/2019-SZTE.

Preparation of serum samples

Blood samples were collected into BD Vacutainer SST II Advance Tubes (Becton, Dickinson and Company), allowed to clot for at least 1 h at room temperature (RT), and centrifuged for 20 min at 3000 × *g*, 10 °C.

sEV isolation, characterization

sEVs were isolated by differential centrifugation (30 min at 10,000 × *g*, 4 °C and 70 min, at 100,000 × *g*, 4 °C). sEV morphology was measured by atomic force microscopy (Asylum MFP-3D) and transmission electron microscopy analysis (Tecnai G2 20 X-Twin type instrument). Particle distribution was analyzed by nanoparticle tracking analysis (NanoSight NS300). The presence of sEVs was confirmed by Western blot analyses (XCell SureLock Mini-Cell System for Alix, CD81, CD5L, and calnexin markers). To characterize proteins expressed on the surface of the serum-derived sEVs, we used bead-based multiplex EV analysis by flow cytometry (MACSPlex Exosome Kit, human, Miltenyi Biotec).

Proteomic Analysis by liquid chromatography-mass spectrometry (LC-MS)

The proteome of 96 whole serum samples and 96 serum sEV samples was analyzed by LC-MS (nanoAcquity UPLC, Waters ACQUITY UPLC M-Class Peptide C18 column; Q Exactive Plus quadrupole-orbitrap hybrid MS). Each group contained 24 individuals, six-sample-pools were created from the individuals. The quantitative measurements were performed in DIA mode, and the analysis was conducted in Encyclopedia 0.81.

MMP-9 Analysis by Enzyme-Linked Immunosorbent Assay (ELISA)

MMP-9 content of 222 serum sEV samples was measured by LEGEND MAX Human MMP-9 ELISA Kit according to the manufacturer's protocol. sEV isolates were measured individually, and the absorbance was read on a Multiskan RC microplate reader.

Statistical analysis of LC-MS and ELISA data

Pearson's correlation analysis was used to investigate the outlier samples. Contaminating proteins and missing values were excluded from the MS data. Data were log-transformed; Cohen's d effect size was calculated to measure the difference between the protein intensity means (Cohen d $\geq$ 2). Pairwise ROC analysis was used to find proteins that can separate a group from another (AUC = 1). Principal component analysis (PCA) with k-means clustering was performed to reduce a large number of variables. Two-tailed Welch's *t*-test was performed to identify the significantly enriched or depleted proteins in sEV samples. Statistical analyses were performed using R program (version 3.6.), Python (version 3.8), and MaxQuant Perseus. Values of $p < 0.05$ were considered significant. GraphPad Prism 8 was used for visualization.

MMP-9 level was normalized to the protein content of serum sEV, meaning that MMP-9 concentration (ng/mL) was divided by the protein concentration of the sEV-enriched isolates (ng/mL) for every sample. The results of all the analyses are notated in parts-per-million (ppm), describing the individual values of MMP-9 in patients. Outliers from the analyzed groups were always excluded (ROUT method), and normal distribution and homogeneity of variances were examined (Shapiro–Wilk, F and the Brown–Forsythe test). The MMP-9 ppm concentrations were compared with Welch's test, paired t-test, and One-way ANOVA. Linear regression analyses were conducted on the logarithmized dataset. The diagnostic potential of the MMP-9 ppm was evaluated using ROC analyses. Kaplan–Meier analyses were used to compare the overall survival rates of different groups. GraphPad Prism 8.3.4 was used for the analyses.

In Silico Analysis of LC-MS Data

Protein data derived from the LC-MS were analyzed by the Ingenuity Pathway Analysis (IPA, Qiagen Bioinformatics).

Data Availability

We have submitted all relevant data of our experiments to the EV-TRACK knowledgebase. EV-TRACK ID: EV200080 and EV230005

.

New Findings

From the comprehensive quantitative and qualitative analyses on four patient groups with a disease affecting the central nervous system (glioblastoma, brain metastasis of non-small cell lung cancer, meningioma, and lumbar disc herniation) by comparing the protein content of their whole serum and serum sEV, we can make the following conclusions:

1. Instead of one or two molecules, only a panel of 10-20 proteins can distinguish CNS tumors precisely.

We aimed to identify the differences between the four patient groups to reveal the characteristic protein profiles associated with the CNS tumors in point. Using an intensity ratio of >2 or <0.5 with Cohen's d effect size of ≥ 2 as a cut-off, we investigated which proteins show reliable intensity difference and which proteins can separate at least one group from the others based on a receiver operating characteristic (ROC) analysis. Moreover, utilizing principal component analysis (PCA) with k-means clustering, we were able to compare the suitability of the two different sample types to distinguish between the CNS tumors in point.

Proteomics analyses by LC-MS were performed on whole serum and sEV samples obtained from patients with GBM, BM, M, and CTRL. Individual samples (n = 24) in each group were arranged into 4 pools to eliminate individual variances, reduce sample number, shorten the time of LC-MS measurements, and reduce the need for materials. The Data independent acquisition (DIA) mode constructed spectral library revealed 311 proteins. Based on Pearson's correlation analyses, one of the sEV control samples had to be excluded from further statistical analyses. After excluding unreliable proteins, as well as proteins with missing values, a total of 262 proteins remained for the final analysis.

Following basic processing, up- and down-regulated protein discovery resulted in 41 whole serum proteins and 45 sEV proteins. In addition to comparing each CNS tumor group to CTRL, between-group differences among the CNS tumor groups were also assessed in the protein selection process. As clinically relevant incidence is an important consideration for selecting the proteins identified, Cohen's d effect size was adopted as an indicator of between-group difference. Cohen's d effect size analysis with a threshold of $d \geq 2$ yielded 10 and 21 proteins in the whole serum and sEV samples, respectively.

In the ROC analyzes 10 whole serum proteins (MMP-9, HSPB1, CASP14, HBG1, IGHG4, DEFA1, VWF, HNRNPA1, S100A8, TLN1) and 17 sEV proteins (MMP-9 , HSPB1, CASP14, HBG1, FGB, GGCT, PF4, S100A7, FN1, ANPEP, FLG2, HSPA8, IGLL1, MMRN1, S100A14, SBSN, SPRR2E) were found to meet the AUC = 1 selection criteria. These proteins were most successful in identifying samples belonging to tumor groups The whole serum and the serum sEV panels shared four significantly altered proteins, namely MMP-9, CASP14, HBG1, and HSPB1.

The best classification results were achieved with the 10-membered serum panel and the 17-membered sEV panel. The number of proteins that comprise the panels cannot be further reduced, as this would lead to a decrease in classification efficiency.

2. The protein content of serum sEV is more suitable for the classification of CNS tumors than the protein content of whole serum.

Following protein selection, PCA and k-means clustering were performed. On the whole serum, the PCA biplot resulted in 3 inhomogeneous or incomplete clusters with calculated cluster homogeneity and completeness scores are 0.56 and 0.73, respectively. In contrast to whole serum samples, the clustering of sEV samples formed homogeneous and complete clusters, with homogeneity and completeness scores of 1. The results of the PCA analyses and k-means clustering indicate considerable differences between the whole serum and sEV samples, of the two sample types, the sEV samples proved to be more efficient in the classification.

In conclusion, the accuracy of distinguishing between CNS tumors can be increased using a protein panel from serum-derived sEVs, compared to analyzing whole serum samples.

3. The quantitative and qualitative differences between the protein content of whole serum and serum sEV may affect the suitability for providing biomarkers.

Pairwise statistical comparison was used to identify proteins significantly enriched or depleted in sEV samples compared to whole serum samples. Sixty-five proteins were significantly enriched in sEV samples, while 129 proteins were significantly depleted.

Among the 17 proteins of the sEV marker panel only 6 were significantly enriched in the sEV samples and 5 of the 10 proteins comprising the specific serum panel had higher abundance in the whole serum.

The presence of sEVs was confirmed by characterization as detailed in the methods section, but the quantitative changes between sEV and total serum proteome indicated the possible presence of lipoprotein and serum protein contamination. The level of apolipoproteins was decreased in sEV-enriched samples (sEV/serum mean ratio is 0.66), however, this fraction could not be completely eliminated. Besides, well-known high-abundance serum proteins (e.g., albumin) dominated the protein content of sEV-enriched samples too. However, the enrichment of non-tissue specific (ITGA2B, ITGB3, LGALS3BP), epithelial cell (CD5L), and platelet related (STOM, TSPAN9) EV marker proteins confirms sEV enrichment (sEV/serum mean ratio is 26.58).

A comparison of sEV and total serum samples showed that there are quantitative differences in the proteome of the two sample types, which may affect the suitability of providing biomarkers for monitoring CNS tumors. These findings suggest that the better suitability of sEV-enriched samples to serve as a biomarker source is not explained by a total increase in the abundance of specific proteins but by the changes in their ratio.

To gain insight into the biological background of the obtained proteomics data, IPA was applied. We performed 'Core Analyses' for whole serum and sEV data separately, yielding a list of significantly influenced 'Diseases and Functions' in each patient group. Using 'Comparison Analysis,' we were able to develop heatmaps covering the relevant systemic and tumor-related functions, as well as the activated or inhibited immune functions. Regarding whole serum samples, many of the significantly influenced functions identified are related to CNS involvement and active immune regulatory processes but the patient groups are not clearly distinguished on the heatmaps. In contrast, on the sEV proteome-based heatmaps M was separated from the malignant tumor groups, where tumor progression-related functions were detected to be highly activated and the activated immune functions predominate over inhibited immune functions.

Since the panel members are not individually appropriate for separation and we could not assign relevant biological functions to them individually, we attempted to specify the common biological role of the protein panels. Therefore, we elaborated two networks containing the selected 10 and 17 proteins identified based on whole serum and the sEV data, respectively. Using the 'Grow tool,' the top ten influenced 'Diseases and Functions' were integrated into the networks. In the case of the whole serum network, nine different related 'Diseases and Functions' were identified, including viral infection, apoptosis, necrosis, cell movement of phagocytes and myeloid cells, and only one was cancer-related.

In contrast, the top ten influenced diseases identified on the sEV network were all tumor-associated, suggesting their potential involvement in the pathophysiology of cancers.

Based on these findings we can conclude that the biological background also might be responsible for the increased suitability of sEV samples in distinguishing CNS tumors.

4. Among the proteome of serum EVs the vesicular MMP-9 showed the greatest potential for distinguishing between patient groups.

As a result of the statistical analyses on LC-MS measurements (96 patients), a 17-membered sEV protein panel was constructed from the identified proteins which were able to separate the four groups with 100% efficacy. After that, we investigated which of the 17 sEV proteins were the most suitable for distinguishing the patient groups.

Based on ROC analysis of MS data, we found matrix metalloproteinase-9 (MMP-9) to be the most significant (p = 0.0065, multi ROC AUC = 0.86).

5. The vesicular MMP-9 level is independent of gender and age, but there is a significant difference in the case of the original tumor and the recurrent tumor, and also in the case that the patient had already received therapy at the time of sampling.

We performed ELISA measurements on the serum sEV samples of 222 patients. The large sample size allowed us to investigate factors influencing the MMP-9 level and is suitable for examining whether there was a measurable, significant difference in the sEV MMP-9 level associated with different diseases and patient survival outcomes. Measuring the MMP-9 concentrations (ELISA) of serum sEV instead of MMP-9 intensities (LC-MS) enabled the determination of the cut-off values required in the clinic, as well as the specificity and sensitivity of the test.

The MMP-9 level was normalized to the total protein content of serum sEV, meaning that MMP-9 concentration (ng/mL) was divided by the protein concentration of the sEV-enriched isolates (ng/mL) for every sample. The results of all the analyses are notated in parts-per-million (ppm) describing the individual values of MMP-9 in patients.

Age and gender analyses were carried out in the CTRL group. Correlation analysis revealed that there is no distinct relationship between age and MMP-9 levels, and there is no significant difference between male and female patients.

Effects of surgical resection on the MMP-9 level of serum sEV were examined in the GBM and BM groups. The MMP-9 levels were similar (p = 0.1843, fold change = 15%)

before and after the surgical resection in the case of primary GBM patients, while BM samples showed a marked increase (p = 0.0065, fold change = 209%) after resection.

Further examination of preoperative GBM samples found a distinct difference between the original tumor and the recurrence based on MMP-9 levels of serum sEV, and the recurrence showed a lower level of MMP-9 on average.

Determining the influence of the administered therapy, MMP-9 levels were also compared based on whether or not GBM patients had received treatment at the time of sampling. Our result indicates that therapy might decrease the MMP-9 levels of the circulating sEVs.

In conclusion, surgical resection, recurrence, and treatment, apart from disease type, might influence the MMP-9 level of the serum sEV. Due to these findings, all subsequent analyses were conducted exclusively on samples obtained prior to surgical resection and therapy administration.

6. There are significant differences between the vesicular MMP-9 content of the malignant and the benign CNS tumors and the malignant tumor and the control groups, and it has prognostic value regarding the probability of survival of glioblastoma patients.

Further statistical analyses were performed to identify if there was a difference between the vesicular MMP-9 levels of the patient groups.

As a first step, Welch's test was used to compare the control group with all the tumor patients. Based on ROC analysis, the two groups were significantly distinguishable using a cut-off point of 16 MMP-9 ppm with 74% sensitivity, 61% specificity, and an AUC of 0.70.

The tumor patients then were divided into M, GBM, and BM, and we found that the differences remained significant between control and malignant tumors, as well as between the benign and malignant tumors. The separate comparison resulted in increased specificity, sensitivity values, and AUC scores in the case of control-malignant tumor comparisons at the expense of control-benign comparisons.

In the last step, the patients were divided into further subgroups based on histopathology. In the subgroup analysis, primary and secondary GBM, patients with grade I and II meningiomas, and brain metastases from patients with adenocarcinoma and carcinoma planocellulare were distinguished. The comparisons of the patient groups showing significant differences in MMP-9 concentrations resulted in AUC scores up to 0.77.

Our data indicate that patients with malignant, but not benign brain tumors can be distinguished from CTRL patients based on the MMP-9 level of serum sEV, and the MMP-9 level of serum sEV shows a positive correlation with tumor aggressiveness.

After comparing the patient groups, we aimed to determine whether vesicular MMP-9 levels correlate with disease progression/patient survival. To assess the prognostic value of MMP-9 levels in serum sEV, we analyzed the preoperative serum sEV samples from 27 GBM patients. Patients in this study were not administered any treatment at the time of sampling.

To reveal the prognostic value of serum sEV's MMP-9 level, subjects were divided into three groups based on their survival time (short-, medium- and long-term) using 0–2, 3-8, and 10–23 months as the cut-offs. Long-term survival was found to be associated with a significantly lower MMP-9 level compared to the MMP-9 levels of the other two groups. These differences represented a high specificity and sensitivity of 80–89% at a threshold of 28 MMP-9 ppm, and AUC values of 0.83–0.87 in ROC analyses allowed efficient distinction of these survival groups.

These results suggest that there should be a correlation between the MMP-9 levels of serum sEV and patient survival, so we performed a subsequent analysis in which we approached the question from the other direction, associating survival time with MMP-9 levels. To determine the extent of the influencing effect of MMP-9 levels on survival, patients were separated into two groups, with the previously established threshold of 28 ppm. Based on the Kaplan–Meier chart, patients with low MMP-9 level (<28 ppm) presented with a significant survival benefit (HR 2.401, 95%CI 1.095 to 5.261, $p = 0.0063$), which represents an eight-month increase in the median survival.

The probability of survival also decreased with age; therefore, we repeated the examination with a cut-off of 65 years. According to the Kaplan–Meier chart, patients under the age of 65 had a five-month increase (HR 2.037, 95%CI 0.8524 to 4.869, $p = 0.0340$) in median survival. It is crucial to note that age and MMP-9 level are independent; we can conclude that the MMP-9 level of serum sEV may influence survival regardless of age.

To summarize the main findings, analysis of samples taken prior to surgical resection and the administration of therapy revealed a negative correlation between higher MMP-9 levels and survival, and long-term (10–23 months) survival was found to be associated with low MMP-9 levels (<28 ppm).

Taken together, the high MMP-9 level of serum sEV might be a negative prognostic marker for overall survival in glioblastoma patients.

Discussion

Small extracellular vesicles isolated from serum may serve as signal-enhancers for the monitoring of CNS tumors

Our investigations aimed to identify protein markers for liquid biopsy-based diagnosis, prognosis, and monitoring of CNS tumors utilizing whole sera and serum sEVs.

Covering the most common CNS tumors, we analyzed a total of 222 serum samples of patients with GBM, BM, M, and controls. Particles were isolated by differential centrifugation from serum samples, the isolated sEVs were characterized by AFM, TEM, and NTA, as well as by examining sEV markers (CD81, Alix, CD5L, and calnexin) by WB. As the quantitative description of the isolation efficiency is required for the proper interpretation of the analytical results, we provided a detailed description of the sEV/serum mean ratio of sEV-marker proteins, marked the remained contaminants (high-abundance serum proteins and lipoproteins), and also submitted all relevant data from our experiments to the EV-TRACK knowledgebase.

We aimed to identify the differences between the four patient groups to reveal the characteristic protein profiles associated with the CNS tumors in point. Using an intensity ratio of >2 or <0.5, Cohen's d effect size of 2 as a cut-off, and AUC = 1 from ROC analysis as selection criteria, we investigated which proteins can separate at least one group from the others. Our research revealed that it is more effective to employ a protein panel as opposed to a single biomarker protein. Using the 17-membered sEV protein panel, we were able to discriminate not only between the tumorous and control patients but between four different patient groups, pointing the way toward the establishment of a general CNS biomarker panel.

Following protein selection, PCA and k-means clustering were used to compare the potential of the established protein panels in tumor classification. Our study revealed that the utilization of a serum sEV protein panel provides a more accurate way of differentiating among different types of CNS tumors than using the protein panel of whole serum samples.

There is increasing evidence, that serum sEV proteins can serve as reliable biomarkers for cancer detection and cancer type determinations. However, the effectiveness of the blood-based and sEV-based biomarkers has not been compared. To the best of our knowledge, we are the first group to quantitatively compare the proteome of whole serum and serum-derived sEVs. Our findings confirm the significance of sEVs in identifying CNS tumors and underline their enhanced capability in categorizing CNS tumors compared to the whole serum.

Statistical and IPA analysis of the proteomic data revealed that quantitative changes and biological background might be responsible for the increased suitability of sEV samples in distinguishing CNS tumors.

On one hand, sEV enrichment increased the relative abundance of proteins innately present in higher concentrations within sEVs and reduced the masking effect of the uninformative protein fraction released from sources not specific to the target disease. The increased signal-to-noise ratio was beneficial for the quantitative LC-MS analysis of proteins of interest.

On the other hand, IPA analysis revealed that many functions related to tumor progression, such as angiogenesis, proliferation, and migration of tumor cells, were significantly activated on the sEV proteome-based heatmaps. The IPA heatmaps and networks showed that the proteome of sEV samples may offer more specific information on the tumor-type and immune reactions in patient groups compared to whole serum.

According to Anderson *et al.*, serum sEV is a dual source of biomolecular information on cancer, as it contains the molecules released by cancer cells as well as those released during tumor-specific immune responses. It can therefore be concluded that the differences observed in the serum vesicles isolated from different patient groups may reflect not only tumor-specific processes but also those related to the associated immune responses.

MMP-9 levels of serum sEV can be used as a prognostic marker for brain tumors

From the determined sEV panel, MMP-9 was found to be the most successful protein in separating the patient groups. To the best of our knowledge, we were the first ones who performed comparative analyses on the MMP-9 content of serum sEVs. We performed ELISA on 222 individual serum sEV samples and compared the results with the gender, age, and clinical history of the patients. Since there is no study on MMP-9 content of circulating vesicles in the literature, our findings can be compared only with blood-based ones.

Numerous studies discuss the clinical significance of MMP-9 content of serum in inflammation, and various cancers, such as breast-, ovarian-, colorectal-, lung-, bladder cancer, and melanoma. Regarding CNS tumors, only a few serum MMP-9 investigations have been published with contradictory findings. Hormingo *et al*. found that YKL-40 and MMP-9 could be monitored in patients' serum and help confirm the absence of active disease in glioblastoma. However, after five years, the same group (Iwamoto *et al*.) conducted a longitudinal prospective study of MMP-9 as a serum marker in gliomas, and the larger cohort could not confirm the previous findings. In contrast, Ricci *et al*. found that MMP-9 levels are significantly higher in high-grade glioma, in low- and high-grade meningioma samples, and in metastasis specimens compared to healthy

individuals. The latter study on serum MMP-9 levels aligns with our findings based on serum sEVs.

There is no consensus in the literature about what factors could influence the MMP-9 level of serum. In our study on serum-derived sEVs, there was no distinct relationship between age and MMP-9 levels, and there was no significant difference between male and female patients in terms of the MMP-9 level. In relation to the surgical resection, the MMP-9 level of serum sEVs were similar before and after the surgical resection in the case of primary GBM patients, while BM samples showed a marked increase after resection. Determining the influence of the administered therapy, MMP-9 levels have to also be compared based on whether or not GBM patients had received treatment at the time of sampling. Our result indicated that therapy might decrease the MMP-9 level of circulating sEVs. Our findings support that the vesicular MMP-9 levels are partly influenced by some patient characteristics. Therefore, when conducting comparative analyses, it is crucial to consider factors such as the surgical resection and the administered therapy rather than gender and age.

We discovered that there are significant differences between the vesicular MMP-9 content of the malignant and the benign CNS tumors and the malignant tumor and the control groups. In other words, sEVs of more aggressive tumors had higher levels of MMP-9. Although our findings are indirect, we were the first to investigate this phenomenon at the serum vesicular level. Numerous studies analyzing tumor tissue. have demonstrated the relationship between tumor aggressiveness and the MMP-9 levels. The higher level of MMP-9 in aggressive tumors compared to benign ones reflects its multifaceted role in promoting tumor progression. Our research on sEVs has led us to the same conclusion. However, a great advantage is a single blood test is sufficient to measure the MMP-9 content of serum vesicles eliminating the need for invasive tissue biopsies through the skull.

Furthermore, we revealed a negative correlation between higher MMP-9 levels and survival. In the test population, long-term (10–23-month) survival was associated with a low MMP-9 level. Moreover, we established a threshold of 28 ppm to determine the patient's probability of survival. Our results support that the high level of MMP-9 in serum-derived sEVs might be a negative prognostic marker of the probability of survival in GBM patients. Only a few studies have addressed the relationship between MMP-9 and survival in CNS tumors. These serum-based studies have conflicting results, only Jiguet-Jiglaire *et al*. reported that reduced MMP-9 levels of plasma were linked to improved overall survival, which aligns with our findings based on serum sEV.

Conclusions

The comparative proteomic analysis showed that sEVs may be more suitable for investigating tumor-related molecular patterns because these molecules are present in higher concentrations in sEV samples than in whole serum samples and contain less "noise" that could introduce bias into the analytical results.

The research discovered that the number of proteins used for monitoring cannot be reduced to a few single molecules, instead, a specific protein panel is required for perfect differentiation.

The *in silico* analyses on the proteome of the two different samples (whole serum and serum sEV) revealed that the biological background of the sEV-based characteristic protein profile is more accurately associated with the tumor types than the whole serum-based protein profile. Samples enriched in sEVs can be a source of biomarkers that can provide us with amplified relevant information. The information thus available represents not only the specific tumor-type but also the associated immune responses. This finding provides further evidence that sEVs may be more effective than whole serum samples for monitoring CNS tumors. It also highlights the advantages of introducing EV-based diagnostics into clinical practice.

Furthermore, the dissertation demonstrates the correlation between the survival of GBM patients and their MMP-9 levels of serum-derived sEVs. As MRI status has limited value in certain clinical situations for predicting tumor spread, and the average survival rate for glioblastoma patients is only 5.5% five years after diagnosis, the prognostic value of vesicular MMP-9 as a biomarker for glioblastoma is critical.

Therefore, given the above findings and the ease with which sEVs can be analyzed, incorporating sEV analysis into clinical practice could open new perspectives in the diagnosis and prognosis of CNS tumors.

Acknowledgments

I would like to start by expressing my gratitude to my supervisor, Krisztina Buzás, for inviting me to join her group and recognizing my potential and ability for scientific work. I am grateful to fate for bringing us together again on a playground ten years after our university days together. This allowed me to hear about the fascinating and significant work that Krisztina and her research group are engaged in. I am grateful to her for this exceptional opportunity, her guidance, and the excellent scientific milieu she has created in her research group.

My gratitude extends to my colleagues, Mária Harmati, Edina Gyukity-Sebestyén, Mátyás Bukva, Lilla Pintér, and Tímea Böröczky. I am thankful for their help and endless support, as well as the opportunity to learn from them. We are connected by countless collaborative research projects and experiences; without them, writing articles and dissertation would not have been feasible.

I would like to thank Professor Dr. Márta Széll, for my admission to the Microbiology, Antimicrobial Immunity Programme of the Doctoral School of Interdisciplinary Medicine.

I am grateful to Péter Horváth, head of the Institute of Biochemistry and the Laboratory of Image Analysis and Machine Learning in the Biological Research Centre, for the infrastructural support of my project.

Our cooperating partners, who are eminent representatives of their field of science, are to be thanked for their professional assistance. I am grateful to Álmos Klekner for having faith in our research group and providing valuable clinical samples and advice. I am grateful to Zoltán Szabó for taking on the challenge and showing enthusiasm for the examination of the previously unknown sample type, extracellular vesicles, as well as for performing the mass spectrometry measurements with great care and proficiency.

Furthermore, I appreciate the financial support, as my research was funded by the National Brain Research Program NAP 2.0; GINOP-2.3.2-15-2016-00015; GINOP-2.2.1-15-2017-00052; OTKA-K143255; and TKP-2021-EGA-09.

I am especially grateful to my parents for their love, and support. I always knew that you believed in me and wanted the best for me.

Last, but most certainly not least, my deepest gratitude is reserved for my husband and my three amazing children. I hope that my tenacity has been an inspiration to them as they have watched me patiently throughout this journey.

The dissertation is based on the following publications:

I. **Dobra, G**.; Bukva, M.; Szabo, Z.; Bruszel, B.; Harmati, M.; Gyukity-Sebestyen, E.; Jenei, A.; Szucs, M.; Horvath, P.; Biro, T.; et al. Small Extracellular Vesicles Isolated from Serum May Serve as Signal-Enhancers for the Monitoring of CNS Tumors. Int. J. Mol. Sci. 2020, 21, 5359, doi:10.3390/ijms21155359.
IF: 5.924; Q1

II. **Dobra, G**.; Gyukity-Sebestyén, E.; Bukva, M.; Harmati, M.; Nagy, V.; Szabó, Z.; Pankotai, T.; Klekner, Á.; Buzás, K. MMP-9 as Prognostic Marker for Brain Tumors: A Comparative Study on Serum-Derived Small Extracellular Vesicles. Cancers 2023, 15, 712, doi:10.3390/cancers15030712. IF: 5.064; Q1

Other publications related to the subject of the dissertation:

III. Bukva, M.; **Dobra, G**.; Gomez-Perez, J.; Koos, K.; Harmati, M.; Gyukity-Sebestyen, E.; Biro, T.; Jenei, A.; Kormondi, S.; Horvath, P.; et al. Raman Spectral Signatures of Serum-Derived Extracellular Vesicle-Enriched Isolates May Support the Diagnosis of CNS Tumors. Cancers 2021, 13, 1407, doi:10.3390/cancers13061407. IF: 6.575; Q1