“VICTOR BABEȘ” UNIVERSITY OF MEDICINE AND PHARMACY FROM TIMISOARA

FACULTY OF MEDICINE

Department IV – BIOCHEMISTRY AND PHARMACOLOGY

NITUŞCĂ DIANA-GABRIELA

PHD THESIS

The diagnostic value of non-coding RNA species and plasma metabolites in prostate adenocarcinoma

A B S T R A C T

Scientific Coordinator:

PROF. UNIV. DR. MARIAN CĂTĂLIN

T i m i ș o a r a

2 0 2 4

A B S T R A C T

An increasing incidence of prostate cancer has been observed in the last decade (prostate cancer cases increased by 3% per year from 2014 to 2019, with a higher incidence rate among African-American men when compared to Caucasian men), representing in 2023 the 2nd most common form of cancer among men worldwide. In addition, this type of cancer has an incidence rate of almost 60% in men over 65 years old, who, due to their advanced age, are immunologically more susceptible to the development of complications that will lead to an increase in mortality and associated morbidity for this type of cancer.

The increased incidence and mortality of this urological pathology can be explained, on the one hand, due to its asymptomatic or non-specific nature in the early stages of development, which leads to a late diagnosis in most cases, and on the other hand due to the lack of sufficiently specific, sensitive and economically accessible diagnostic biomarkers. The most commonly used molecular marker in clinical practice for the diagnosis and monitoring of prostate cancer is prostate-specific antigen (PSA, ng/ml), and its utility has been the focus of controversy and debate in the scientific community over the past decade due to its increased sensitivity, its low specificity (PSA can also have elevated values in other non-malignant pathologies associated with the prostate, inflammation, weight gain, etc.) and the minimal effect on mortality caused by prostate cancer, a fact that led to overdiagnosis, administration of excessive treatments and performing invasive biopsies that were not necessary. Therefore, further research is needed for the discovery, development and validation of novel, minimally invasive diagnostic biomarkers, which could complement existing ones for the detection and monitoring of prostate cancer, or which could even surpass the performance of current markers.

Therefore, given the motivation, importance and actuality of the chosen doctoral topic, the objectives of this study were:

O1. Determination of plasma long non-coding RNA (lncRNA) profiles belonging to a group of 30 subjects, of which 15 belong to the group of patients with confirmed prostate cancer with different PSA levels and Gleason scores, and 15 belong to the group of healthy subjects. The objective of this stage is to demonstrate

the ability of lncRNA to distinguish between these two groups, through differential expression, which will suggest their role as diagnostic biomarkers in prostate cancer. This analysis is practically performed by real-time polymerase chain reaction (qPCR) techniques, after the extraction of total RNA with the help of specific kits purchased from established manufacturers. In addition, for comparison, within objective O1, lncRNA profiles are determined from biopsies available from 8 subjects (4 patients and 4 controls) by the microarray technique.

O2. Identification of relevant, common, differentially expressed lncRNAs in the two types of biological samples (plasma and tissue), by comparing the data obtained from the two distinct analyses performed (qPCR and microarray).

O3. Validation step: determining the expression level of the common lncRNA identified from the panel selected at O1, in a case-control study of prostate cancer, comprising the previous 15 cases and 15 controls, plus 37 patients and 23 additional controls, to replicate and validate specific lncRNAs as prostate cancer diagnostic biomarkers. Within this objective, the analysis is performed by qPCR using specific primers for the lncRNA identified at O2, and the statistically significant difference between the expression level of the lncRNA in the patient samples compared to the controls will be expressed by the statistical value of p, which must be smaller than 0.05 to be considered statistically significant. The data are also compared with the relevant and current scientific literature. Last but not least, the diagnostic value of the respective lncRNA is evaluated through Receiver Operating Characteristics (ROC) statistical analysis and through the value of the area under the curve (AUC).

O4. Separation and identification of all circulating metabolites (< 550 Da) by modern techniques of ultra high-performance liquid chromatography coupled with mass spectrometry (UHPLC-QTOF-(ESI+)-MS) from plasma samples from patients with histopathologically confirmed prostate cancer and healthy individuals (71 samples in total – 48 patients and 23 controls, the same from the previous objectives plus other additional samples received along the way from the Urology Department of the "Pius Brînzeu" County Emergency Clinical Hospital of Timisoara).

O5. Comparison and analysis of differentially expressed metabolites in prostate cancer samples compared to controls; multivariate statistical analysis for discrimination and separation of groups; univariate statistical analysis for evaluating the diagnostic performance of differentially expressed metabolites; non-targeted metabolomic analysis (with the identification of all molecules by consulting

established databases) and targeted that was carried out by selecting the most relevant and studied metabolites, discovered in the non-targeted phase (and corroborated with the literature available internationally until 2023): spermine, acetylcarnitine, tryptophan, proline, lysophosphatidylcholine C18:2.

O6. Correlation of metabolites with different concentrations, statistically significant ($p<0.05$), with the serum PSA level and with the cancer stage for the evaluation of prognostic and staging potential.

O7. Carrying out a meta-analysis (in silico) to evaluate the diagnostic value of a microRNA intensively studied in the literature (miR-375) as a potential diagnostic biomarker in prostate cancer. To achieve this objective, the specialized literature is studied to identify the most relevant microRNAs associated with this type of cancer and which currently do not benefit from a global study of this kind, and then, for the chosen microRNA, the statistical tests necessary to evaluate its value as a potential diagnostic biomarker are performed, by putting together all the statistical results obtained individually by international researchers who studied the same microRNA and who met the inclusion criteria.

In study 1, following the comparison between FFPE tissues and plasma samples, we observed only one common lncRNA, significantly differentially expressed, with increased expression in cancer, namely lncRNA NEAT1 ($p<0.05$ in both analysis groups). In the next step, NEAT1 was individually validated in plasma and tissue samples from both groups, as its expression was commonly and significantly upregulated in cancer.

In the plasma samples from Group 1 (referred to as TM), although NEAT1 had an almost two-fold increased expression (FC = 1.836), unfortunately, the threshold of statistical significance was not reached ($p = 0.351$), probably due to the small cohort. In contrast, in Group 2 (named CJ), where the sample size was larger, NEAT1 was significantly increased in cancer (FC = 2.101, $p = 0.009$).

ROC analysis for NEAT1 in Group 2 (CJ) revealed a very good area under the curve (AUC) value of 0.7298 (95%CI = 0.5812–0.8785), therefore suggesting a potential high biomarker for this type of lncRNA.

In study 2, of the six total articles included in the meta-analysis, I identified a total of 422 patients with diagnosed prostate cancer and 212 controls (70 healthy subjects and 142 BPH patients). Of all included studies, three used serum, two used plasma, and two used urine as biological samples. QUADAS-2 showed that the vast majority of included research studies had a low risk of bias.

Since after testing heterogeneity between studies we found significant heterogeneity (I2 = 93.68% and 86.12% for sensitivity and specificity respectively), we decided to use the random effects model. The meta-analysis showed a pooled sensitivity of 0.76 (95% CI: 0.55–0.89) and a pooled specificity of 0.83 (95% CI: 0.63–0.94) and an AUC value of 0.87 (95% CI: 0.83–0.89), indicating an overall good diagnostic accuracy for miR-375.

Based on the calculated pooled estimates of sensitivity and specificity, we calculated the average likelihood ratios of positive and negative test results and found that the LR+ and LR- for miR-375 were 4.6 (95% CI: 2.30–9, 30) and 0.29 (95% CI: 0.16–0.51), respectively. In addition, we obtained a mean DOR value of 16 (95% CI: 10–26).

Next, we investigated possible sources of heterogeneity in both sensitivity and specificity by performing a specimen-based meta-regression analysis and found that the types of biological specimens used could represent a potential source of heterogeneity in specificity ($p < 0.001$). However, given the small-scale design of the study, further reports are needed to fully elucidate the source of heterogeneity.

Analysis of pretest and posttest probabilities suggests a somewhat high diagnostic value of miR-375 as a promising future biomarker for this urological disease. At a pretest probability of 25%, the posttest probability positivity would increase to 60% with an LR+ of 5, while the posttest probability negativity would decrease to 9% with an LR− of 0.29.

Finally, we performed the pooled sensitivity and specificity analyses after removing each report and found that the final results did not differ much from the initial results. Correlated with Deeks' lack of significance for the asymmetry test ($p = 0.58$), these data strongly suggest a lack of significant publication bias on the reports and underline the stability and credibility of the results.

In study 3, a total of 296 molecules with molecular weights below 550 Da were identified. These molecules were subsequently identified using various international databases such as the Human Metabolome Database (http://www.hmdb.ca), Lipid Maps (http://www.lipidmaps.org), PubChem (https://pubchem.ncbi.nlm.nih.gov) and the Heidelberg database (https://www.msomics.com/).

The vast majority of metabolites identified were either amino acids (AA - such as proline, L-tryptophan, valine, threonine, aspartic acid, ornithine, lysine, glutamic acid, etc.) or were products derived from AA metabolism. Other metabolites were amines (spermine, glutamine, serotonin) or, in general, high-nitrogen compounds, while a substantial part consisted of metabolites derived from lipid metabolism (phospholipids, lysophospholipids, fatty acids). A small fraction were nucleotide metabolites such as uridine, guanosine and inosine. Interestingly, these preliminary data are consistent with other findings in the literature regarding free AA and lipid profiles.

Next, we performed an analysis to assess biomarker potential that included receiver operating characteristic (ROC) curve analysis, with corresponding area under curve (AUC) calculation, and metabolites were classified based on AUC values. The most representative molecules to discriminate group C from group P and to be considered as potential biomarkers (with AUC values greater than 0.7) are AA (such as L-tryptophan, tyrosine, pro-line and their derivatives), phosphatidylcholines, amines (spermine), N1-acetylspermidine, L-cystathionine, D-sphingosine, hexacosanoylcarnitine and some fatty acids and esters. Acylated AAs (acetyltyrosine) were also important for discrimination.

Furthermore, according to the previous untargeted analysis and consistent with other findings in the literature, we selected five different molecules (proline, spermine, acetylcarnitine, L-tryptophan and LPC 18:2 lysophosphatidylcholine) for the targeted metabolomics study.

Determination of linear ranges (calibration curves and equations including R^2 values), limits of detection (LOD) and limits of quantification (LOQ) of each standard were calculated. Correlation coefficients (R^2) were greater than 0.898 for all standards in their linear range, showing good linear relationships in the linear ranges. All LOD values were in the range of 0.3–4 μM and LOQ values were in the range of 0.9–5.5 μM.

Also, graphs were constructed for the calibration curves, from which the equations of the lines and the R^2 values were generated, for each selected metabolite separately.

It is noted that all metabolites had low levels in group P, regardless of their stage, compared to group C. For L-proline, we observed a gradual decrease in concentration from stage I to stage III, with a slight increase in stage IV. In addition, spermine levels decreased more than 10-fold in the P group (all stages) compared to the C group, with the most significant change among all metabolites. For acetylcarnitine, there was a gradual increase in concentrations that was directly proportional to stages II-IV, while the reverse occurred for L-tryptophan (stages I-III), which had a similar profile to L-proline.

In addition, to assess the staging potential of these five molecules, ordinary one-way ANOVA tests were performed on their concentration values (μM) for controls and patients grouped according to their AJCC stage (I, II, III and IV). Statistically significant differences were observed for spermine, acetylcarnitine and L-tryptophan ($p<0.0001$), but not for L-proline ($p = 0.209$) or LPC 18:2 ($p = 0.159$).

Unpaired t-tests showed that there were statistically significant differences between patients (all stages) and controls for all five molecules analyzed. Spermine, acetylcarnitine and L-tryptophan showed p-values below 0.0001, while L-proline and LPC 18:2 had p-values of 0.02 and 0.01, respectively. All five molecules had significantly lower concentrations in the patient group compared to the control group.

According to these data, the molecule with the highest diagnostic potential is represented by L-tryptophan (AUC value of 0.981), suggesting a very good diagnostic potential for detecting prostate cancer. The same is true for acetylcarnitine and spermine (AUC values of 0.923 and 0.922, respectively), while L-proline and LPC 18:2 showed only moderate diagnostic value, both with AUC values below 0.7.

In conclusion, this Ph.D. thesis demonstrates that the research, in a multidisciplinary manner (using various techniques and targeting molecules from various biochemical classes) of new biomarkers for the early and specific detection of prostate cancer (with an increasing incidence), represents a step forward for modern, minimally invasive, personalized medicine, based on solid evidence, which could represent a real advantage both for specialists and for patient comfort. It seems that the liquid biopsy has enormous potential for elucidating the processes of carcinogenesis that take place at the cellular level. The investigation of different

metabolites or genetic material from the circulating biological fluids, originating from the circulatory stream directly from the tumor sites, could help in a deeper understanding of cancer mechanisms and could also represent new reliable biomarkers for the early and specific diagnosis of this malignancy.

Thus, we demonstrated (through study 1) that the lncRNA species NEAT1 is statistically significantly overexpressed in biological samples from patients with confirmed prostate cancer, compared to healthy controls, and that the diagnostic value is very high, suggesting the potential of NEAT1 as a biomarker for the detection of this pathology.

Also, the meta-analysis performed (in study 2), corroborated the existing data in the literature and statistically confirmed that miR-375 has a satisfactory diagnostic accuracy in differentiating prostate cancer from healthy individuals.

Last but not least and without limiting to the study of non-coding genetic material, it has been demonstrated (through study 3) that metabolomics represents an emerging technique for the simultaneous discovery of circulating metabolites with significantly altered levels in prostate cancer. In this regard, we observed a decrease in various classes of molecules such as amino acids (tryptophan, proline), amines (spermine), lysophosphatidylcholine and acetylcarnitine, in the samples of cancer patients compared to healthy individuals.

Therefore, this study shows that the various molecular biomarkers, together with the clinical and histopathological features of the tumors and the imaging techniques will have to be corroborated and used in a complementary rather than competitive manner, as each presents specific peculiarities and provides crucial information for an early and effective diagnosis, which could lead not only to a decrease in the incidence of prostate cancer, but also to an increase in the quality of life of these patients, which will eventually lead to a reduction in the morbidity and mortality caused by prostate cancer.