

Biomarkers From Discovery to Clinical Application: In Silico Pre-Clinical Validation Approach in the Face of Lung Cancer

Biomarker Insights
Volume 19: 1–9
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/11772719241287400



Medi Kori¹, Esra Gov², Kazim Yalcin Arga³ and Raghu Sinha⁴

¹Department of Medical Biotechnology, Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul, Türkiye. ²Department of Bioengineering, Faculty of Engineering, Adana Alparslan Türkeş Science and Technology University, Adana, Türkiye. ³Department of Bioengineering, Marmara University, Istanbul, Türkiye. ⁴Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, PA, USA.

ABSTRACT

BACKGROUND: Clinical biomarkers, allow better classification of patients according to their disease risk, prognosis, and/or response to treatment. Although affordable omics-based approaches have paved the way for quicker identification of putative biomarkers, validation of biomarkers is necessary for translation of discoveries into clinical application.

OBJECTIVE: Accordingly, in this study, we emphasize the potential of in silico approaches and have proposed and applied 3 novel sequential in silico pre-clinical validation steps to better identify the biomarkers that are truly desirable for clinical investment.

DESIGN: As protein biomarkers are becoming increasingly important in the clinic alongside other molecular biomarkers and lung cancer is the most common cause of cancer-related deaths, we used protein biomarkers for lung cancer as an illustrative example to apply our in silico pre-clinical validation approach.

METHODS: We collected the reported protein biomarkers for 3 cases (lung adenocarcinoma-LUAD, squamous cell carcinoma-LUSC, and unspecified lung cancer) and evaluated whether the protein biomarkers have cancer altering properties (i.e., act as tumor suppressors or oncoproteins and represent cancer hallmarks), are expressed in body fluids, and can be targeted by FDA-approved drugs.

RESULTS: We collected 3008 protein biomarkers for lung cancer, 1189 for LUAD, and 182 for LUSC. Of these protein biomarkers for lung cancer, LUAD, and LUSC, only 28, 25, and 6 protein biomarkers passed the 3 in silico pre-clinical validation steps examined, and of these, only 5 and 2 biomarkers were specific for lung cancer and LUAD, respectively.

CONCLUSION: In this study, we applied our in silico pre-clinical validation approach the protein biomarkers for lung cancer cases. However, this approach can be applied and adapted to all cancer biomarkers. We believe that this approach will greatly facilitate the transition of cancer biomarkers into the clinical phase and offers great potential for future biomarker research.

PLAIN LANGUAGE SUMMARY

Biomarkers, which are routinely used in clinics, allow better classification of patients according to their disease risk, prognosis, and/or response to treatment. Although affordable omics-based approaches have paved the way for quicker identification of putative biomarkers, validation of biomarkers is necessary for translation of discoveries into clinical application. This research article highlights the challenges of translating cancer biomarkers into clinical practice and summarizes feasible step toward “in silico pre-clinical validation” using the example of lung cancer types. Accordingly, protein biomarkers proposed for lung cancer are being investigated using the “in silico pre-clinical validation” approach to determine whether they have cancer altering properties (i.e., oncoprotein, tumor suppressor, and cancer hallmark), are expressed in body fluids (i.e., plasma/serum, saliva, urine, and bronchoalveolar lavage) and can be targeted with FDA-approved drugs. We believe that the step of in silico pre-clinical validation is the future of biomarker research for all professionals involved in clinical, biological, epidemiological, biostatistical and health research, and that it will greatly facilitate the transition of biomarkers to the clinical phase.

KEYWORDS: Protein biomarkers, in silico pre-clinical validation step, lung cancer, oncoprotein, tumor suppressor, cancer hallmark, body fluids, approved drugs

RECEIVED: May 15, 2024. **ACCEPTED:** September 10, 2024.

TYPE: Research Article

CORRESPONDING AUTHORS: Medi Kori, Department of Medical Biotechnology, Institute of Health Sciences, Acibadem Mehmet Ali Aydinlar University, Istanbul 34752, Türkiye. Email: medi.kori@acibadem.edu.tr

Raghu Sinha, Department of Biochemistry and Molecular Biology, Penn State College of Medicine, Hershey, PA 17033-2360, USA. Email: rsinha@pennstatehealth.psu.edu

Introduction

Cancer is an uncontrolled cell division process that leads to cell transformation due to the occurrence of various genetic changes.¹ It is one of the most common causes of death and the most frequently diagnosed disease in almost every country in

the world. According to Global Cancer Statistics, 20 million new cancer cases and 9.7 million cancer deaths were observed in 2022. Among the different types of cancer, lung cancer is the most common cause of cancer death according to Global Cancer Statistics.² Recently, the American Cancer Society



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

estimated that lung cancer will be responsible for 234 580 new cancer cases and 125 070 cancer deaths in 2024.³

The diagnosis of early-stage lung cancer is around 15% and the cancer is usually only diagnosed in the late stages. In addition, the prognosis in the late stage is poor and is less than 5%.⁴ Smoking is one of the main causes of lung cancer. In addition, genetic factors and exposure to environmental pollutants increase the risk of developing lung cancer.⁵ Histopathologically, lung cancer is divided into 2 main types, small cell lung cancer and non-small cell lung cancer.⁶ The prevalence of non-small cell lung cancer is much higher than that of small cell lung cancer. For example, small cell lung cancer accounts for around 13% to 15% of all lung cancer cases. In addition, non-small cell lung cancers are essentially categorized into different subtypes. The most common forms are lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC).⁷

Statistics on the incidence and high mortality rate of cancers such as lung cancer show that new clinical strategies are needed, including the development of newer biomarkers or the discovery of novel biomarkers in diagnostics or drug development. For a trait to be called a biomarker, it must be measurable with valid tests, instruments, or assays, indicate differences between different phenotypes and be used specifically for medical practice (i.e., for diagnosis, prognosis, or response to a treatment).⁸ But are these properties sufficient to indicate that these biomarkers have clinical utility and are ready to be analyzed in various clinical qualification studies? In other words, once biomarkers have been discovered, should the analytical and clinical testing of biomarkers begin immediately? In our opinion, they should not. Many conditions can be investigated or qualified before promising biomarkers can be transferred to clinical validation, which we have termed “in silico pre-clinical validation” in this study. With the advent of next-generation technologies, omics approaches (genomics, transcriptomics, proteomics, metabolomics, lipidomics, and microbiomics), bio-informatics, and big data, biomarker discovery is now moving quite rapidly toward clinical translation, and we believe that any in silico pre-clinical validation initiative can help solve the current challenges in validating biomarkers^{9,10} and pave the way for their use in the clinic.

Although each method has advantages and disadvantages depending on the application, the advantages of in silico approaches can be listed as follows: (i) it is a high-throughput method that allows researchers to screen large amounts of data efficiently, quickly, and simultaneously, (ii) since it is a high-throughput method, the time and resources required are much less than compared to other approaches, (iii) combining in silico studies with other approaches is much easier than with other approaches.^{11,12} Since promising biomarkers that are discovered should have specific characteristics and stringent properties, these properties can be efficiently analyzed with the in

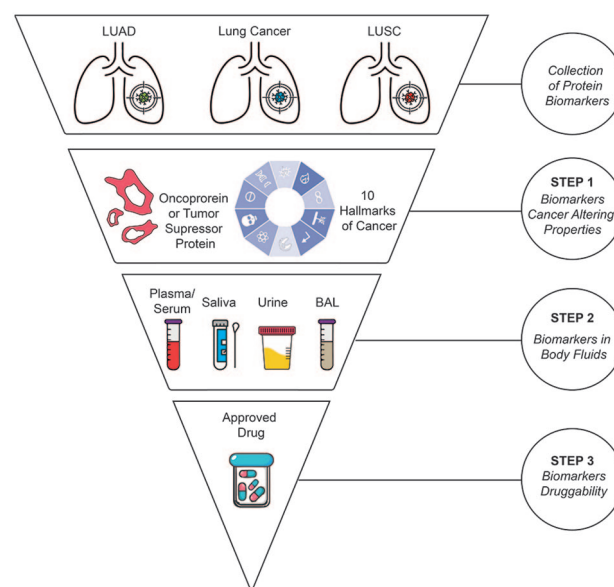


Figure 1. The overall approach implemented for the lung cancer types in the study.

Abbreviations: BAL, bronchoalveolar lavage; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

silico step (in silico pre-clinical validation step) and then the effects on human health outcomes should be approved experimentally (i.e., with clinical trials) for routine use in clinics.

Generally, the priority in the field of biomarkers is to validate the discovered biomarkers in an appropriate clinical context and to integrate them into the clinic. Considering the advantages of in silico methods and the requirements in the field of biomarkers, our main objective in this study is to apply and illustrate 3 different sequential in silico pre-clinical validation steps in lung cancer cases that we have proposed to help the possibility of translating cancer biomarkers into the clinic. In brief, we collected the reported protein biomarkers for 3 lung cancer types and performed 3 in silico pre-clinical validation steps: (i) evaluating whether the biomarkers function as tumor suppressors or oncoproteins and act as cancer hallmarks (i.e., cancer altering properties), (ii) evaluating whether they are expressed in body fluids, and (iii) evaluating whether they can be targeted with FDA-approved drugs (Figure 1). Overall, with this study we have proposed 3 novel sequential in silico pre-clinical validation steps that can be easily adapted for all cancer biomarkers to improve the quality of biomarkers and give them the confidence to qualify for clinical trials.

Materials and Methods

Collection of lung cancer protein biomarkers

We have collected the reported protein biomarkers for lung cancer, LUAD, and LUSC from the BIONDA database (BIOmarker and biomarker caNdidates DAtabase) (as of May 2023). This database provides researchers with molecular biomarkers such as gene, protein, and miRNA biomarkers as well

as biomarker candidates, which are published in scientific article abstracts and updated monthly using text mining.¹³ We focused on protein biomarkers in this study and collected only the protein biomarkers. When we compare the different types of molecular biomarkers, we can say that mRNAs are less stable than DNAs due to the differences in molecular structure, which is important for the reproducibility of biomarker measurements.¹⁴ However, proteins, which are among the most important functional molecules, can be highly dysregulated under different conditions. Therefore, the differential expression of proteins in cells under different conditions is important for identifying and understanding biological information. For these reasons, we believe that protein biomarkers can take a step forward compared to other types of molecular biomarkers in the clinic.

In addition, the reported protein biomarkers for lung cancer collected include all forms of lung cancer without specifying subtypes and may therefore include cases of LUAD and LUSC. Since it was hypothesized that the biomarkers for lung cancer currently used or recommended for use in the clinic should be different depending on the histological subtype of lung cancer,¹⁵⁻¹⁸ we included the unspecified lung cancer cases, LUAD, and LUSC cases in the study to investigate the similarities and differences of the reported biomarkers and the extent of their overlap.

Moreover, in order to understand the molecular functions of the collected protein biomarkers for 3 types, the PANTHER (protein analysis through evolutionary relationships) classification database¹⁹ was used.

Evaluation of protein biomarkers cancer altering properties

Cancer is a multi-stage development characterized by genetic changes that eventually lead to abnormal cell growth. Oncogenes and tumor suppressors are the 2 different classes of genes that can promote the development of malignant cells through different mechanisms such as gain of function and loss of function.²⁰ While oncogenes are genes whose activation contributes to the development of cancer, tumor suppressor genes are genes whose suppression contributes to the development of cancer.²¹ Since the presence or absence as well as the expression level and pattern of oncogenes and/or tumor suppressors may correlate with malignant transformation (i.e., disease development and progression) or treatment, their contribution to clinical outcome is evident. Therefore, to assess the cancer altering properties of the biomarkers, we first used the Catalog of Somatic Mutations in Cancer (COSMIC) database (v.95)²² and obtained a list of proteins that have tumor suppressor or/and oncogene functions.

In 2000, researchers wondered that while there are many types of cancer, how many or whether the same regulatory mechanisms or signals affect normal cells and turn them into

cancer. After pondering these questions, the researchers proposed the concept of “cancer hallmarks,” which states that these hallmarks play a crucial role in cancer development and are probably present in all types of cancer. The 10 cancer hallmarks approved today include (i) proliferative signaling, (ii) suppression of growth, (iii) escaping immune response to cancer, (iv) cell replicative immortality, (v) tumor promoting inflammation, (vi) invasion and metastasis, (vii) angiogenesis, (viii) genome instability and mutations, (ix) escaping programmed cell death, and (x) change of cellular energetics.²³

Mostly, the hallmarks of cancer genes are considered as driver genes of tumorigenesis. These genes can orchestrate for the most basic phenotypic features of tumor initiation and progression. Focusing on genes that play an important role in carcinogenesis and considering confirmed associations with cancer, in short, using genes that have been presented as cancer hallmarks as measurable indicators, such as biomarkers, can be very beneficial for the diagnosis and treatment of cancer. Therefore, it can be assumed that biomarkers that exhibit cancer hallmark features allow the identification of the cancer, the characterization of the cancer subtype or the detection of its biological characteristics in a more effective way. Accordingly, it can be assumed that a cancer biomarker with cancer hallmark properties will increase the quality of biomarkers and take them a step further in terms of their reliability in clinical studies. In this study, ten cancer hallmark genes were retrieved from a publicly available biological repository, the Cancer Hallmark Genes (CHG) database.²⁴ GeneCards: The Human Gene Database²⁵ was used to determine the proteins encoded by the cancer hallmark genes.

Evaluation of protein biomarkers presence in biological body fluids

Biological body fluids are the fluids that the human body either excretes or secretes. Biological fluids such as blood, plasma, urine, and saliva have high potential for non-invasive biomarker measurements compared to tissue, organ samples or cell extracts.^{26,27} It can be assumed that the collection of body fluids is non-injurious, simple, fast and inexpensive and that the examination of biomarkers from biological fluids in clinics is an important feature and advantage that should be considered. To discover the invasiveness of the culminated protein biomarkers with cancer altering features, the expression of proteins in plasma/serum, saliva, urine and bronchoalveolar lavage (BAL) was investigated using HBFP: a new repository for human body fluid proteome source.²⁸

Evaluation of protein biomarkers druggabilities

Once biomarkers have been discovered, the next step would be to find out if these biomarkers are druggable, that is, if the biomarker can be targeted with a drug, especially an FDA-approved

drug. In this way, oncologists can offer their patients an appropriate choice of drugs and/or drug combinations, and patients can be assured that the drugs they use are carefully controlled and approved for use.^{29,30} Therefore, to investigate whether protein biomarkers can be used as drug targets, we used the Pharos database³¹ to identify the presence of approved drug targets of biomarker proteins.

Results

Extraction of protein biomarkers

A total of 3407 different protein biomarkers for lung cancer (3008 protein biomarkers), LUAD (1189 protein biomarkers), and LUSC (182 protein biomarkers) were extracted from the database. To reveal the potential of the biomarkers for stratifying lung cancer patients into different types, we also comparatively analyzed the identified biomarkers. Of the 3407 protein biomarkers, 105 were identified for all analyzed types. In addition, 364 and 16 protein biomarkers specific to LUAD and LUSC, respectively, were identified. Also 701 common protein biomarkers for lung cancer and LUAD, 42 common protein biomarkers for lung cancer and LUSC and 19 common protein biomarkers for LUAD and LUSC were identified (Figure 2A).

It was found that the identified biomarkers have different molecular functions. However, the focus is on the functions of protein modification (400 protein biomarkers), metabolite conversion (363 protein biomarkers) and gene-specific transcriptional regulation (286 protein biomarkers) (Figure 2B).

In silico pre-clinical validation step 1: Lung cancer biomarkers cancer altering properties

It is clear that oncogenes and tumor suppressors play an important role in cancer development through loss of function or gain of function. If we compare a protein with oncogenic and tumor suppressor function with a protein without these functions, we can assume that the protein with oncogenic or tumor suppressor function is a more suitable candidate for a biomarker and can be used more efficiently as a cancer biomarker in clinics. Moreover, since the genes involved in these ten cancer hallmarks are drivers of tumor development, it is important to study these genes in different cancer types to understand the process of tumor development and use them as biomarkers.³² Accordingly, a total of 557 different proteins that have an oncoprotein function, a tumor suppressor function or both functions were extracted from the COSMIC database²² and a total of 374 different hallmark proteins were extracted from the CHG database²⁴ and these lists were integrated with our protein biomarker lists respectively.

In lung cancer, only 8.57% of the 3008 protein biomarkers, in LUAD only 11.85% of the 1189 protein biomarkers and in LUSC only 15.38% of the 182 protein biomarkers have at least oncoprotein or tumor suppressor functions. Of these protein biomarkers, 23 are common in all the 3 cancer types. In

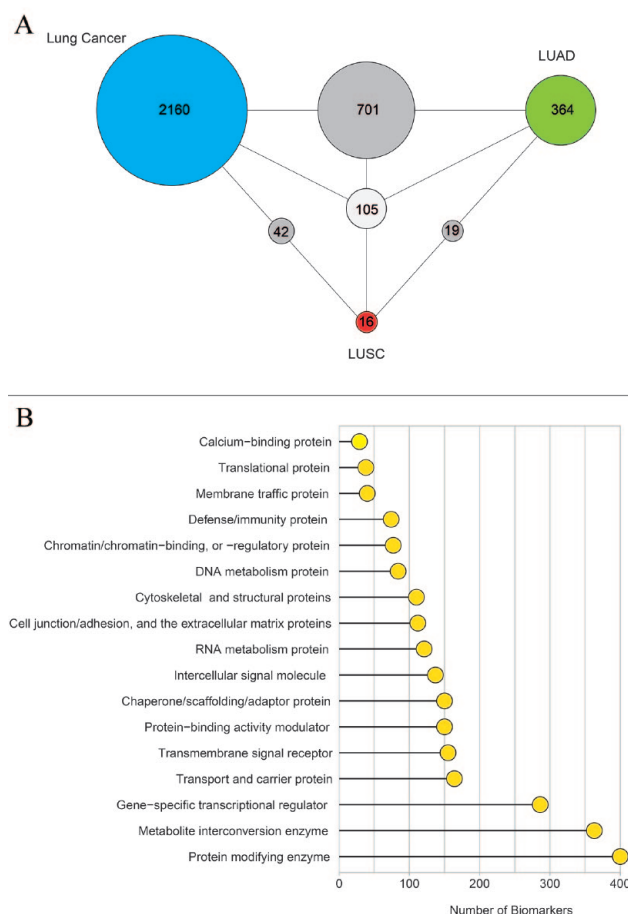


Figure 2. The number and molecular functions of the obtained protein biomarkers for lung cancer: (A) the Venn diagram representing the number of protein biomarkers collected for lung cancer, LUAD, and LUSC from the database and (B) the diagram representing the molecular functions of all protein biomarkers obtained. Abbreviations: LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

addition, 98 protein biomarkers common to lung cancer and LUAD and 5 protein biomarkers common to lung cancer and LUSC were identified that have at least oncoprotein or tumor suppressor functions (Figure 3A). Among these functions, the protein biomarkers with an oncoprotein function (46.90%, 49.65%, and 46.43% of protein biomarkers of lung cancer, LUAD, and LUSC, respectively), then the protein biomarkers with a tumor suppressor function (36.82%, 35.46%, and 35.71% of protein biomarkers of lung cancer, LUAD and LUSC, respectively) and finally protein biomarkers with both functions (16.28%, 14.89%, and 17.86% of protein biomarkers of lung cancer, LUAD, and LUSC, respectively) are the most represented in each of the 3 types (Figure 3B). It is assumed that a tumor suppressor gene is mutated, inactivated, or downregulated, while oncogenes are activated or upregulated in cancer cells. Since tumor suppressors are reduced in cancer cells under normal conditions, it is more difficult to detect tumor suppressors as biomarkers in cancer patients compared to oncoproteins.³³ Although many proteins are eliminated at this stage, it

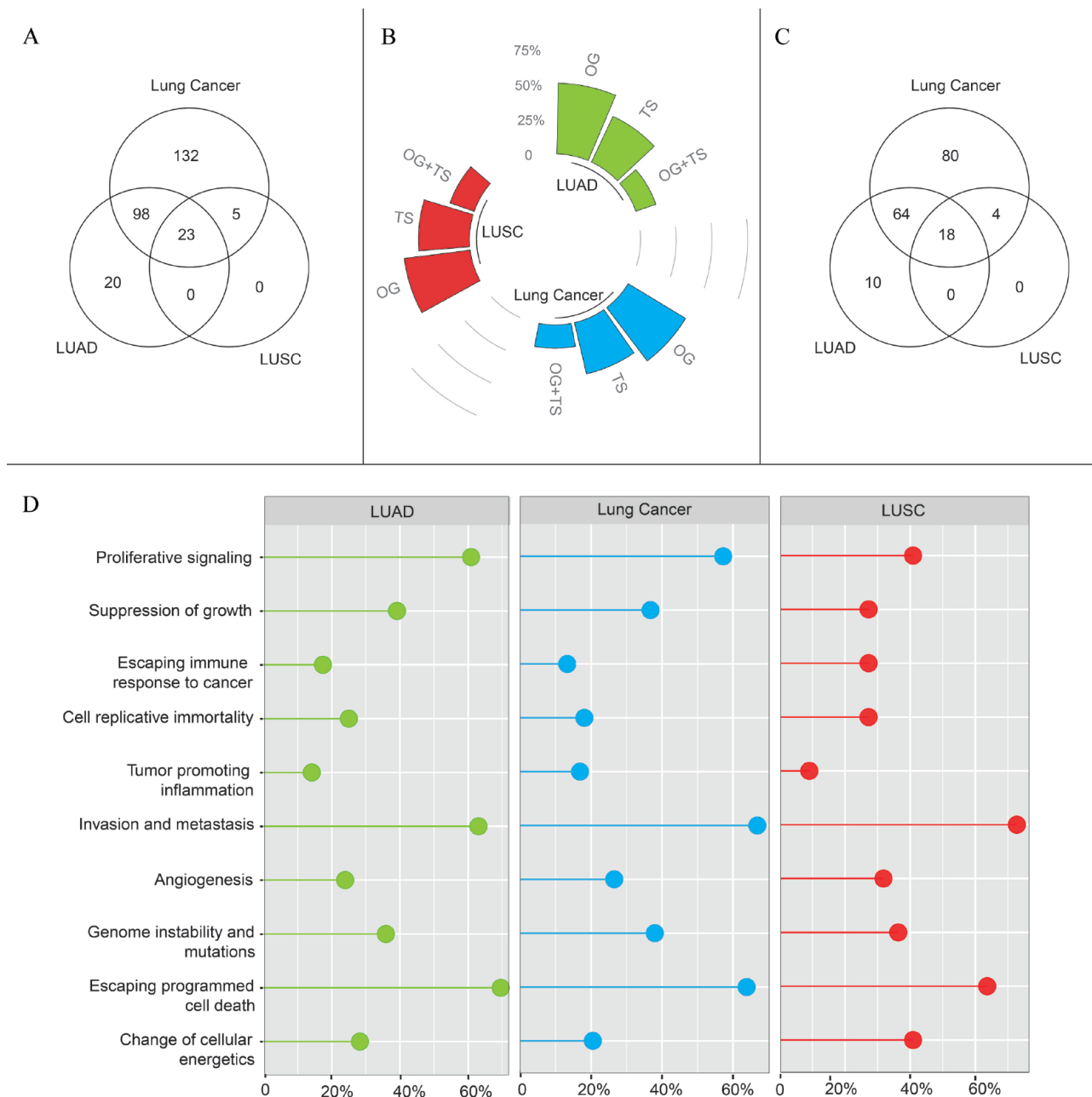


Figure 3. In silico pre-clinical validation Step 1: The investigation of the cancer altering properties of biomarkers for lung cancer: (A) the comparative plots of protein biomarkers for 3 lung cancer types (lung cancer, LUAD, and LUSC) showing at least one oncogenic or tumor suppressive feature, (B) the distribution of oncogenic and/or tumor suppressive protein biomarkers, (C) the comparative diagrams of protein biomarkers for 3 types showing both oncogenic or/and tumor suppressive and cancer hallmark features, and (D) the percentage distribution of protein biomarkers, which cancer hallmark feature they have among the 10 different cancer hallmarks.

Abbreviations: LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; OG, oncogene; TS, tumor suppressor.

is therefore desirable that biomarkers with oncoprotein functions become evident.

When these protein biomarkers were examined to determine whether they function as cancer hallmark proteins, 64.34%, 65.24%, and 78.57% of the protein biomarkers in lung cancer, LUAD, and LUSC, respectively, were found to function as hallmark proteins in addition to their oncoprotein and/or tumor suppressor function. Of these protein biomarkers, 18 are found in all 3 types. In addition, 64 protein biomarkers

common to lung cancer and LUAD and 4 protein biomarkers common to lung cancer and LUSC were identified that have at least an oncoprotein or tumor suppressor function and were reported as cancer hallmarks (Figure 3C). Of the ten cancer hallmarks for all 3 cancer types, we can report that protein biomarkers play a prominent role in the following 3 hallmarks: invasion and metastasis (66.87%, 63.04%, and 72.73% of protein biomarkers of lung cancer, LUAD, and LUSC, respectively), escape from programmed cell death (63.86%, 69.57%, and

63.64% of protein biomarkers of lung cancer, LUAD and LUSC, respectively), and proliferative signaling (57.23%, 60.87%, and 40.91% of protein biomarkers of lung cancer, LUAD, and LUSC, respectively) (Figure 3D). It is evident that one of the main targets of cancer cells in the early stages of malignancy development is profound proliferation. Moreover, targets that play an important role in promoting proliferation are still one of the best targets in cancer therapy.³⁴ Since we are studying a cancer, we believe that the biomarkers that have been or are being used in daily life for this disease should also have oncogenic features to accurately represent the disease.

We collected a total of 3407 different protein biomarkers for 3 cases, of which 176 (5.16%) have at least an oncogenic or tumor suppressive function and are presented as cancer hallmarks, that is, have cancer altering properties (S. Table 1).

In silico pre-clinical validation step 2: Detection of lung cancer biomarkers in body fluids

We can assume that one of the biggest challenges in transferring biomarkers to the clinic is clinical sampling. Non-invasive or minimally invasive collection and measurement of biomarkers is important for rapid, economical, and repeatable analyses.³⁵ Overall, the protein biomarkers identified for 3 types that have cancer altering properties (176 protein biomarkers) can also be obtained from at least 1 of the 4 biological fluids analyzed, with a value of about 72% (127 protein biomarkers) (S. Table 1). Accordingly, 58 of the protein biomarkers specific to lung cancer, 5 of the protein biomarkers specific to LUAD, 48 of the protein biomarkers common to lung cancer and LUAD, 3 of the protein biomarkers common to lung cancer and LUSC, and 13 of the protein biomarkers common to 3 cancer types were expressed in at least 1 of the body fluids investigated (Figure 4A). To determine more precisely which body fluid(s) (i.e., plasma/serum, saliva, urine and BAL) provide the best results, we comparatively analyzed the expression of the 127 protein biomarkers. This showed that for biological sampling, fluids such as urine (expressed 96 of the 127 protein biomarkers) and plasma/serum (expressed 92 of the 127 protein biomarkers) alone are more suitable for analyzing/measuring protein biomarkers. Of the protein biomarkers examined, 25.81% of the proteins in lung cancer, 28.36% of the proteins in LUAD and 50% of the proteins in LUSC were found to be expressed in both plasma and urine (Figure 4B). In contrast, 25 of the 127 protein biomarkers and 34 of the 127 protein biomarkers were expressed only in saliva and BAL, respectively.

In silico pre-clinical validation step 3: Druggability of lung cancer biomarkers

The fact that a biomarker has an approved target drug makes that biomarker a frontrunner, and this is a feature that is desired by oncologists. Accordingly, a total of 30 of the 127 protein

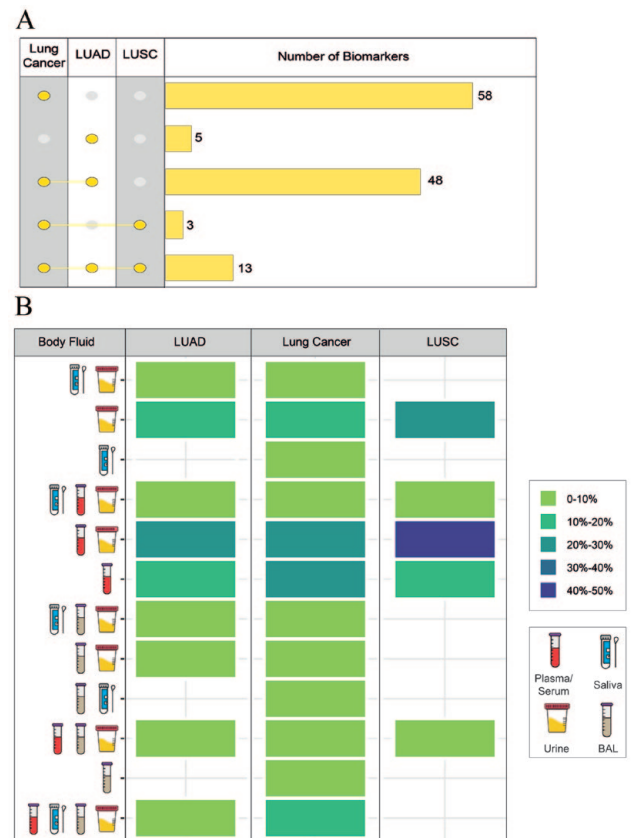


Figure 4. In silico pre-clinical validation Step 2: The analysis of lung cancer biomarkers in body fluids (A). The comparative visualization of protein biomarkers that exhibit oncogenic properties and are present in at least one of the four body fluids. (B) The distribution of protein biomarkers according to their occurrence in the body fluids. Abbreviations: BAL, bronchoalveolar lavage; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

biomarkers (23.62%) that have cancer altering properties and are expressed in at least 1 of the biological fluids analyzed also have FDA-approved drug target proteins. Of the 30 protein biomarkers, 2 biomarkers (FES and XPO1) were found to be specific for LUAD and 5 biomarkers (POLE, ATP1A1, PDGFRA, KDR, and ESR1) were found to be specific for lung cancer. In addition, 17 biomarkers were found to be present in both lung cancer and LUAD, and 6 biomarkers (EZH2, PIK3CA, MET, FGFR1, FGFR2, and EGFR) were found in all 3 cases studied (Figure 5). There are at most 27 FDA-approved drugs for the protein biomarker ESR1, followed by 26 FDA-approved drugs for EGFR and 17 FDA-approved drugs for KDR.

Discussion

New technologies are constantly being developed to identify molecular markers for complex diseases. Researchers have used various systems biology approaches, bioinformatics, and machine learning pipelines³⁶⁻⁴⁰ and experimental techniques^{41,42} to discover biomarkers for all types of cancer. Although these new technologies have enabled researchers to find biomarker candidates using simpler approaches, these biomarkers should

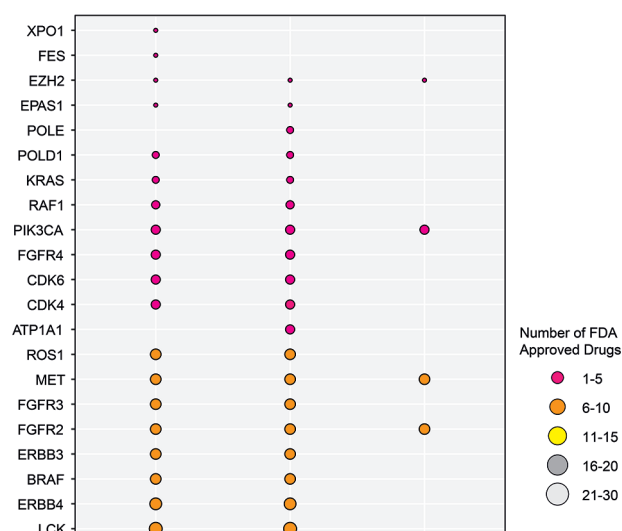


Figure 5. In silico pre-clinical validation Step 3: the investigation of druggability of lung cancer biomarkers. The bubble diagram shows the number of protein biomarkers targeted by FDA-approved drugs. Abbreviations: LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma.

be pre-validated and specialized for use in clinical practice and have specific properties to be identified as actionable biomarkers.

Strengthening cancer prevention and prognosis measures and creating a sustainable infrastructure for cancer treatment are crucial to the global fight against cancer. This health challenge can be reduced with the help of the potential biomarkers. As biomarkers can guide the optimization of decisions in clinical practice, these can meet the needs of the healthcare system.⁴³ However, the question is whether the researchers who discover new biomarkers can also apply them in clinical practice? Contrary to scientific belief, the translation of biomarkers into the clinic is low. For example, the *PD-L1* biomarker was approved by the FDA in 2018, the *EGFR* protein biomarker was approved by the FDA in 2012, the *ALK* biomarker was approved by the FDA in 2011, and the *CEA* biomarker was approved by the FDA in 1985 and is currently used in clinical practice for lung cancer.⁴⁴

To translate the potential biomarkers into clinical benefit, extensive standard requirements must be met, which can be efficiently assessed by in silico pre-clinical validation approaches, of which we have proposed 3 in silico pre-clinical steps in the study. We believe that when we talk about cancer biomarkers, these promising biomarkers should point to cancer altering properties. In cancer, most molecular targeted therapies are still

inhibitors of oncogenes or/and their associated signaling pathways.^{45,46} In addition, minimally invasive biomarkers are desirable in clinical application and the treatability of the biomarker is another important feature that may be coveted by oncologists. In the face of lung cancer types, in this study we obtained a total of 3008 protein biomarkers for lung cancer, 1189 protein biomarkers for LUAD and 182 protein biomarkers for LUSC. After the applying of the 3 in silico pre-clinical validation steps, these values decreased to 28, 25 and 6 protein biomarkers for lung cancer, LUAD and LUSC respectively (S. Figure 1), clearly demonstrating that several of the reported biomarkers have great potential for clinical testing or use in clinics. Overall, with these 3 sequential in silico pre-clinical validation approaches, molecular cancer biomarkers can be screened according to specific criteria and help researchers to identify robust biomarkers that need to be investigated experimentally. We therefore believe that this approach can make clinical trials easier (i.e., subject to budget, time and equipment requirements) and more affordable for researchers/experts.

In addition to the 3 in silico pre-clinical validation approaches that we have employed in this study to support the translation of biomarkers to the clinic, the following in silico approaches can also be investigated: (i) receiver operating characteristic (ROC) curve analysis, (ii) Kaplan-Meier analysis, and (iii) supervised and unsupervised machine learning approaches. The ROC analyses can be used to understand the diagnostic power of biomarkers by assessing both the sensitivity and specificity of a biomarker.⁴⁷ Kaplan-Meier analysis can be used to understand the prognostic power of biomarkers by assessing the specific time to event.⁴⁸ For biomarker assessment unsupervised machine learning algorithms that includes clustering (i.e., hierarchical and *K*-means clustering), association (Apriori algorithm) and dimensionality reduction (principle component analyses) approaches and supervised machine learning algorithms including regression (i.e., linear regression and decision tree) and classification (i.e., logistic regression, support vector machines, and naïve bayes) approaches can be used to assess patients and control groups discrimination capacity as well as to evaluate the disease prediction efficiency.^{49,50} Therefore, these in silico pre-clinical validation analyses could also reveal the biomarkers that are truly desirable for clinical investment.

Conclusion

While “omics” technologies and/or in silico approaches can accelerate biomarker discovery and increase the possibility of biomarker translation to the clinic, these technologies can also improve existing approaches and pave the way for easier and faster integration of discovered biomarkers into clinical applications. Translating a promising biomarker discovery into clinical practice requires large-scale clinical trials. However, the challenges posed by these large-scale studies can be reduced or overcome by in silico pre-clinical validation approaches which

we have proposed in this study. As protein biomarkers are becoming more prominent in the clinic alongside other molecular biomarkers, and lung cancer is the most common cause of cancer-related deaths, in this study we have used the protein biomarkers of lung cancer as an illustrative example of a *in silico* pre-clinical validation approach that can be applied to any cancer type. We recommend that these *in silico* pre-clinical validation steps should become a routine procedure for all clinical, biological, epidemiological, biostatistical and health research professionals.

Declarations

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Author Contributions

Medi Kori: Conceptualization; Data curation; Formal analysis; Writing—original draft; Writing—review & editing. Esra Gov: Writing—review & editing. Kazim Yalcin Arga: Writing—review & editing. Raghu Sinha: Conceptualization; Writing—review & editing.

Acknowledgements

Not applicable.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Competing Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Availability of Data and Materials

Publicly available databases were analyzed in this study.

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Krieghoff-Henning E, Folkerts J, Penzkofer A, Weg-Remers S. Cancer – an overview. *Med Monatsschr Pharm.* 2017;40:48–54.
- Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2024;74:229–263.
- Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024;74:12–49.
- Araghi M, Mannani R, Heidarnejad Maleki A, et al. Recent advances in non-small cell lung cancer targeted therapy; an update review. *Cancer Cell Int.* 2023;23:162.
- Amicizia D, Piazza MF, Marchini F, et al. Systematic Review of lung cancer screening: advancements and strategies for implementation. *Healthcare.* 2023;11:2085.
- Inamura K. Lung Cancer: Understanding its molecular pathology and the 2015 WHO classification. *Front Oncol.* 2017;7:193.
- Lahiri A, Maji A, Potdar PD, et al. Lung cancer immunotherapy: progress, pitfalls, and promises. *Mol Cancer.* 2023;22:40.
- Ahmad A, Imran M, Ahsan H. Biomarkers as biomedical bioindicators: approaches and techniques for the detection, analysis, and validation of novel biomarkers of diseases. *Pharmaceutics.* 2023;15:1630.
- Drucker E, Krapfenbauer K. Pitfalls and limitations in translation from biomarker discovery to clinical utility in predictive and personalised medicine. *EPMA J.* 2013;4:7.
- Antoniou M, Kolamunnage-Dona R, Wason J, et al. Biomarker-guided trials: challenges in practice. *Contemp Clin Trials Commun.* 2019;16:100493.
- Moradi M, Golmohammadi R, Najafi A, et al. A contemporary review on the important role of *in silico* approaches for managing different aspects of COVID-19 crisis. *Inform Med Unlocked.* 2022;28:100862.
- Chang Y, Hawkins BA, Du JJ, et al. A guide to *in silico* drug design. *Pharmaceutics.* 2022;15:49.
- Turewicz M, Frericks-Zipper A, Stepath M, et al. BIONDA: a free database for a fast information on published biomarkers. *Bioinform Adv.* 2021;1:vbab015.
- Xi X, Li T, Huang Y, et al. RNA biomarkers: frontier of precision medicine for cancer. *Noncoding RNA.* 2017;3:9.
- Su R, Zhang J, Liu X, Wei L. Identification of expression signatures for non-small-cell lung carcinoma subtype classification. *Bioinformatics.* 2020;36:339–346.
- Relli V, Trerotola M, Guerra E, Alberti S. Abandoning the notion of non-small cell lung cancer. *Trends Mol Med.* 2019;25:585–594.
- Takamochi K, Ohmiya H, Itoh M, et al. Novel biomarkers that assist in accurate discrimination of squamous cell carcinoma from adenocarcinoma of the lung. *BMC Cancer.* 2016;16:760.
- Chen JW, Dhahbi J. Lung adenocarcinoma and lung squamous cell carcinoma cancer classification, biomarker identification, and gene expression analysis using overlapping feature selection methods. *Sci Rep.* 2021;11:13323.
- Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* 2019;47:D419–D426.
- Novikov NM, Zolotaryova SY, Gautreau AM, Denisov EV. Mutational drivers of cancer cell migration and invasion. *Br J Cancer.* 2021;124:102–114.
- Osborne C, Wilson P, Tripathy D. Oncogenes and tumor suppressor genes in breast cancer: potential diagnostic and therapeutic applications. *Oncologist.* 2004;9:361–377.
- Tate JG, Bamford S, Jubb HC, et al. COSMIC: the catalogue of somatic mutations in cancer. *Nucleic Acids Res.* 2019;47:D941–D947.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144:646–674.
- Zhang D, Huo D, Xie H, et al. CHG: a systematically integrated database of cancer hallmark genes. *Front Genet.* 2020;11:29.
- Stelzer G, Rosen N, Plaschkes I, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Curr Protoc Bioinformatics.* 2016;54:1.30.1–1.30.33.
- Dayon L, Cominetti O, Affolter M. Proteomics of human biological fluids for biomarker discoveries: technical advances and recent applications. *Expert Rev Proteomics.* 2022;19:131–151.
- Huang L, Shao D, Wang Y, et al. Human body-fluid proteome: quantitative profiling and computational prediction. *Brief Bioinform.* 2021;22:315–333.
- Shao D, Huang L, Wang Y, et al. HBFP: a new repository for human body fluid proteome. *Database.* 2021;2021:baab065.
- Gromova M, Vaggelas A, Dallmann G, Seimet D. Biomarkers: opportunities and challenges for drug development in the current regulatory landscape. *Biomark Insights.* 2020;15:1177271920974652.
- Colburn WA. Biomarkers in drug discovery and development: from target identification through drug marketing. *J Clin Pharmacol.* 2003;43:329–341.
- Kelleher KJ, Sheils TK, Mathias SL, et al. Pharos 2023: an integrated resource for the understudied human proteome. *Nucleic Acids Res.* 2023;51:D1405–D1416.
- Kori M, Ozdemir GE, Arga KY, Sinha R. A pan-cancer atlas of differentially interacting hallmarks of cancer proteins. *J Pers Med.* 2022;12:1919.
- Kontomanolis EN, Koutras A, Syllaios A, et al. Role of oncogenes and tumor-suppressor genes in carcinogenesis: a review. *Anticancer Res.* 2020;40:6009–6015.
- Feitelson MA, Arzumanyan A, Kulathinal RJ, et al. Sustained proliferation in cancer: mechanisms and novel therapeutic targets. *Semin Cancer Biol.* 2015;35:S25–S54.

35. Su SB, Poon TC, Thongboonkerd V. Human body fluid. *Biomed Res Int.* 2013;2013:918793.
36. Kori M, Demirtas TY, Comertpay B, et al. A 19-gene signature of serous ovarian cancer identified by machine learning and systems biology: prospects for diagnostics and personalized medicine. *OMICS.* 2024;28:90-101.
37. Kori M, Arga KY. HPV16 status predicts potential protein biomarkers and therapeutics in head and neck squamous cell carcinoma. *Virology.* 2023;582:90-99.
38. Xiao Y, Bi M, Guo H, Li M. Multi-omics approaches for biomarker discovery in early ovarian cancer diagnosis. *EBioMedicine.* 2022;79:104001.
39. Kori M, Gov E. Bioinformatics Prediction and machine learning on gene expression data identifies novel gene candidates in gastric cancer. *Genes.* 2022;13:2233.
40. Zheng Y, Tang L, Liu Z. Multi-omics analysis of an immune-based prognostic predictor in non-small cell lung cancer. *BMC Cancer.* 2021;21:1322.
41. Davies MPA, Sato T, Ashoor H, et al. Plasma protein biomarkers for early prediction of lung cancer. *EBioMedicine.* 2023;93:104686.
42. Mukama T, Fortner RT, Katzke V, et al. Prospective evaluation of 92 serum protein biomarkers for early detection of ovarian cancer. *Br J Cancer.* 2022;126:1301-1309.
43. Sarhadi VK, Armengol G. Molecular biomarkers in cancer. *Biomolecules.* 2022;12:1021.
44. Pal M, Muinao T, Boruah HPD, Mahindroo N. Current advances in prognostic and diagnostic biomarkers for solid cancers: detection techniques and future challenges. *Biomed Pharmacother.* 2022;146:112488.
45. Morris LG, Chan TA. Therapeutic targeting of tumor suppressor genes. *Cancer.* 2015;121:1357-1368.
46. Bahar ME, Kim HJ, Kim DR. Targeting the RAS/RAF/MAPK pathway for cancer therapy: from mechanism to clinical studies. *Signal Transduct Target Ther.* 2023;8:455.
47. Bantis LE, Tsimikas JV, Chambers GR, et al. The length of the receiver operating characteristic curve and the two cutoff Youden index within a robust framework for discovery, evaluation, and cutoff estimation in biomarker studies involving improper receiver operating characteristic curves. *Stat Med.* 2021;40:1767-1789.
48. D'Arrigo G, Leonardi D, Abd ElHafeez S, et al. Methods to analyse time-to-event data: the Kaplan-Meier survival curve. *Oxid Med Cell Longev.* 2021;2021:2290120.
49. Bodaghi A, Fattahi N, Ramazani A. Biomarkers: promising and valuable tools towards diagnosis, prognosis and treatment of covid-19 and other diseases. *Heliyon.* 2023;9:e13323.
50. Ng S, Masarone S, Watson D, Barnes MR. The benefits and pitfalls of machine learning for biomarker discovery. *Cell Tissue Res.* 2023;394:17-31.