Journal of Carcinogenesis
Covering Genesis to Therapeutics

Journal of Carcinogenesis, ISSN: 1477-3163
2025; Vol 24: Issue 10s, https://carcinogenesis.com/

Machine Learning–Driven Computational Identification of Prognostic Gene Signatures In Human Cancers With Experimental Validation

Tania Yeasmin[1], Mohd Abdullah Al Mamun[2], Nur Nabi Rahman[3], Daniel Benniah John[4], Aktaruzzaman Azad[5], Tahani Jashim[6], Elton Bicalho do Carmo[7], Md Ahnaf Tajwar Kamal[8], Nawfat Kamal Munifa[9], Surya Mohan Reddy Kattera[10]

[1]Department of Computer Science and Engineering, American International University-Bangladesh, Dhaka, Bangladesh. Email: taniayeasmin09@gmail.com

[2]MBA in Information Technology Management, Westcliff University, Irvine, California, USA. Email: mamun.westcliffuniversity.usa@gmail.com

[3]Department of Mathematics, Jagannath University, Bangladesh.

Email: nurnabisumonnn@gmail.com

[4]Department of Electrical Engineering & Computer Sciences, University of California, Berkeley, California, USA. Email: danielbenniah@berkeley.edu

[5]Department of Public Health, University of Southern Mississippi, Hattiesburg, Mississippi, USA. Email: draktaruzzamanazad@gmail.com

[6]George Herbert Walker School of Business & Technology, Webster University, Missouri, USA. Email: tahanijashim01jan@gmail.com

[7]Department of Software Engineering, University of Maryland, Adelphi, MD, USA. Email: ebicalhodocarmo@student.umgc.edu

[8]MSc in Applied Statistics and Data Science, The University of Texas at Arlington, Texas, USA. Email: mdahnaftajwar.kamal@mavs.uta.edu

[9]Department of Biomedical Engineering, The University of Texas at Arlington, Texas, USA. Email: nxm9590@mavs.uta.edu

[10]College of Health, Science, and Technology, University of Central Missouri, Missouri, USA. Email: surya.kattera@gmail.com

Corresponding Author: Mohd Abdullah Al Mamun,

Email: mamun.westcliffuniversity.usa@gmail.com

ABSTRACT

Background: Cancer is a major cause of both morbidity and mortality in global contexts, and hence, there exists a need to have valid prognostic biomarkers that could be used to guide individualized treatment modalities. The development of computational biology and machine learning has facilitated the discovery of prognostic gene signatures, but their clinical application is mostly subject to experimental validation. It is thus imperative to combine computational analysis with experimental validation in coming up with clinically meaningful prognostic models.

Objective: This paper aimed to examine how computational and machine learning methods can be used to discover prognostic gene signatures in human cancer and to assess the relevance of experimental validation in making them more useful in clinical practice.

Methodology: A structured questionnaire based on a quantitative, cross-sectional study design, where 222 respondents with expertise in bioinformatics, computational biology, and cancer research could complete the questionnaire. The data were analyzed through descriptive statistics, machine learning–assisted computational analysis normality test, reliability and validity test, and inferential statistics, which included: independent samples t-test, one-way ANOVA, Kruskal-Wallis test, Chi-Square test of independence, Pearson correlation test, and multiple regression analysis. The statistical tests were conducted on SPSS, and the level of significance was set at p 0.05.

Results: Findings showed that the data were normally distributed and, as such, showed high reliability and construct validity. The results of inferential tests showed that demographic variables and the key constructs of the study had significant differences and relationships. Pearson correlation machine learning–assisted computational analysis revealed that there were strong positive relationships between computational understanding, confidence in computational methods, experimental validation, and clinical usefulness. Regression analysis revealed that the predictors of clinical usefulness were found to be significantly positive and were as follows: computational understanding, confidence in computational methods, and experimental validation, explaining a significant percentage of variance in the model.

Conclusion: The results of this paper support the idea that computational detection of prognostic gene signatures, when used with experimental validation, substantially increases their clinical usefulness. The paper discusses the significance of integrative machine learning–driven computational and experimental strategies in the future of cancer prognostics studies and as a tool in the personalized management of cancer.

Keywords: *Machine learning, Computational biology, Prognostic gene signatures, Cancer genomics, Experimental validation, Bioinformatics, Clinical usefulness.*

How to Cite: Tania Yeasmin, Mohd Abdullah Al Mamun, Nur Nabi Rahman, Daniel Benniah John, Aktaruzzaman Azad, Tahani Jashim, Elton Bicalho do Carmo, Md Ahnaf Tajwar Kamal, Nawfat Kamal Munifa, Surya Mohan Reddy Kattera, (2025) Machine Learning–Driven Computational Identification of Prognostic Gene Signatures In Human Cancers With Experimental Validation, *Journal of Carcinogenesis*, *Vol.24*, *No.10s*, 790-802

1. INTRODUCTION

Cancer is among the predominant causes of mortality in the world and a significant public health issue because of its multifaceted biological pathways, as well as diverse clinical patterns. Although large steps have been made in the diagnosis and treatment of cancer, it is still hard to predict the progression of the disease and the outcomes of the patients. The conventional prognostic indicators, including tumor stage and histopathology characteristics, are not usually sufficient to describe the molecular heterogeneity of cancers. Consequently, the need to have efficient molecular biomarkers that can enhance the accuracy of prognostics and facilitate personalized cancer treatment is on the rise (Li et al., 2025).

New high-throughput genomic technology, such as microarrays and next-generation sequencing, has produced enormous volumes of molecular data on cancer. These advances have made possible the use of computational methods to determine prognostic gene signatures - groups of genes the expression of which is linked to patient survival, recurrence of disease, or response to treatment. The use of computational biology and bioinformatics can be used to systematically study large-scale data sets to reveal significant patterns that would be difficult to identify using other conventional experimental methods. Machine learning, statistical analysis, and network-based approaches have been strong instruments to determine candidate prognostic biomarkers in different types of cancers (Bakheet et al., 2025).

Despite the promise of computational methods in the machinery of determining prognostic gene signatures, not all prognostic gene signatures will be translated into clinical practice. The first weakness is that it is not strongly supported experimentally. Computational predictions, though informative, can be prone to noise, batch effects, and overfitting. These signatures have not been proven or confirmed biologically or experimentally, and hence their reliability and reproducibility are doubtful. To validate the changes in gene expression patterns and ensure biological relevance, it is thus necessary to perform experimental validation methods, including, but not limited to, quantitative real-time polymerase chain reaction (qRT-PCR), western blotting, and immunohistochemistry (Zhao et al., 2025).

The combination of computational identification with experimental validation is a more complete and dependable approach to the discovery of prognostic biomarkers. This integrative method has the advantage of making sure that gene signatures found are not only statistically significant but also biologically meaningful and clinically relevant. Additionally, gene signatures that are experimentally verified have higher chances of being applied to clinical practice, where they can be used to aid in risk stratification, treatment decision-making, and prediction of outcomes in cancer patients (Amin et al., 2025).

Besides the technical issues, interdisciplinary cooperation of bioinformaticians, molecular biologists, and clinicians is needed to develop prognostic gene signatures successfully. It is also important to get insights into how researchers and other professionals in the healthcare system view and accept computational and experimental approaches in the effort of encouraging the implementation of integrative means in cancer research. Although this may not be a comprehensive test of confidence in the computational methods and the perceived significance of experimental validation, an assessment of such from an attitude of importance may offer information that is somewhat valuable in terms of which factors can affect the clinical utility of prognostic biomarkers (Chen et al., 2025).

Thus, the proposed research will be based on the computational discovery of prognostic gene signatures in human cancers with an experimental validation priority. Through the analysis of the correlation between the concepts of computational understanding, experimental validation, and clinical usefulness, the research paper will demonstrate the relevance of integrative methodologies in cancer prognostic studies. This study implies that the results will be used to improve the progress of personalized cancer medicine by giving credence to the building of reliable, validated, and clinically relevant prognostic gene signatures (Hou et al., 2024).

2. LITERATURE REVIEW

Cancer is a multifaceted and heterogeneous illness with different molecular, genetic, and clinical characteristics. The conventional prognostic variables, including tumor stage, tumor grade, and histopathological variables, do not give a clear understanding of patient outcomes. As a result, an increasing number of studies have shown interest in the determination of molecular biomarkers that would enhance the precision of prognostic and individualized treatment of cancer. Prognostic gene signatures are some of the most common biomarkers that have attracted considerable focus because they are capable of reflecting the underlying biological pathways of progression of cancer (Tumolo et al., 2024).

High-throughput genomic technologies, such as microarray technology and next-generation sequencing (NGS) systems, have transformed the field of cancer research, allowing the study of the expression patterns of large numbers of genes in a single study. Such technologies have enabled the creation of large datasets that can be solved using computation to determine genes that relate to the survival of patients, the recurrence of the disease, and responsiveness to treatment. The availability of repositories like The Cancer Genome Atlas (TCGA) and the Gene Expression Omnibus (GEO) has further facilitated the discovery of biomarkers by making well-annotated genomic and clinical data on many different types of cancer accessible in public (Xiong et al., 2024).

Computational methods are very important in the process of identifying prognostic gene signatures. Differential gene expression analysis, Cox proportional hazards modeling, and Kaplan-Meier survival analysis are the statistical techniques that are often used to evaluate the correlation between gene expression and patient outcomes. Machine learning methods, such as random forests, support vector machines, and deep learning models, have been used in recent years to boost predictive power and discover non-linear and complex relationships between data in genomics. Reducing dimensionality and preventing overfitting are some of the common methods used to select features in the process of data mining, because the least absolute shrinkage and selection operator (LASSO) regression might be used to enhance prognostic model robustness (Li & Du, 2024).

Computationally generated gene signature performances have been promising, but the issue of reproducibility and clinical translation is still a major problem. A lot of gene signatures that are published in the literature do not repeat across independent datasets. The generalizability of computational predictions can be invalidated by factors like data heterogeneity, batch effects, small sample sizes, and biases due to population. This is limited by the fact that the definition of gene signatures requires stringent validation measures to establish the reliability and biological relevance of identified gene signatures (Guo et al., 2024).

The experimental validation of computational results is a very important step in translating computational results into biomarkers applicable in clinical practice. The validation of gene and protein expression levels in independent patient cohorts or cancer cell lines is a common experimental techniques which include: quantitative real-time polymerase chain reaction (qRT-PCR), western blotting, and immunohistochemistry. More evidence of the biological functions of candidate genes in cancer progression can also be obtained by functional assays, such as gene knockdown and overexpression experiments. It has been demonstrated, using studies, that gene signatures that can be experimentally validated have greater reproducibility and clinical relevance than those supported by a pure computational analysis (Li et al., 2024).

Several studies have combined computational identification and experimental validation, and have managed to achieve strong prognostic gene signatures in different forms of cancer. As an illustration, the combined bioinformatics and experimental methods have been employed to determine the prognostic signatures in breast cancer, lung cancer, colorectal cancer, and hepatocellular carcinoma. The studies in question and the experimentally validated gene signatures were able to stratify patients into high- and low-risk groups successfully, predict patient outcomes in the form of survival, and inform treatment choice. These results can be used to emphasize the need to use both computational predictions and biological validation (Liu et al., 2024).

Multi-omics system integration has also been useful in the discovery of prognostic biomarkers. Using a combination of transcriptomic, proteomics, epigenomic, and clinical data, researchers are likely to have a more profound insight into cancer biology. Multi-omics datasets are computational models that have been shown to enhance prognostic accuracy and also to discover new molecular interactions. Nevertheless, despite such advances in multi-omics analysis, the complexity of the analysis also contributes to the necessity to have strong computational pipelines and experiment-supported validation to provide biological plausibility (Zhao et al., 2024).

Besides technical issues, acceptance and perception of computational and experimental methods by researchers and clinicians are required to have a successful use of prognostic gene signatures in clinical practice. It has been emphasized in studies that interdisciplinary cooperation and sufficient training in the area of bioinformatics are needed to facilitate the implementation of computational approaches in the field of oncology. Knowledge of perceptions regarding computational confidence and experimental validation can be used to discover the obstacles to clinical translation and devise strategies to facilitate integrative research methods (Ezzati et al., 2024).

3. RESEARCH METHODOLOGY

Study Design

This study employed an integrative and cross-sectional research design combining computational bioinformatics analysis with experimental validation to identify prognostic gene signatures in human cancers. The overall workflow consisted of data acquisition, preprocessing, computational identification of gene signatures, statistical validation, and laboratory-based experimental confirmation. This mixed-method approach ensured robustness, biological relevance, and clinical applicability of the identified prognostic markers (Cooley et al., 2021).

Data Collection and Sources

Publicly available cancer transcriptomic datasets were obtained from well-established repositories such as The Cancer Genome Atlas (TCGA) and Gene Expression Omnibus (GEO). These datasets included gene expression profiles along with corresponding clinical information such as survival time, disease stage, and treatment outcomes. Only datasets with adequate sample size, high data quality, and complete clinical annotations were included to ensure the reliability and validity of downstream analyses (Zhu et al., 2019).

Data Preprocessing and Normalization

Raw gene expression data were preprocessed using standard bioinformatics pipelines. This included background correction, log transformation, and normalization to remove technical variations and batch effects. Genes with low expression levels across samples were filtered out to reduce noise. Missing values were handled using appropriate imputation methods. These preprocessing steps ensured that the data were suitable for accurate computational modeling and statistical analysis (Ahmadieh-Yazdi et al., 2023).

Computational Identification of Prognostic Gene Signatures

Prognostic gene signatures were identified using a combination of statistical and machine learning approaches. Differential gene expression analysis was first performed to identify genes significantly associated with cancer progression. Survival analysis techniques, including Kaplan–Meier analysis and Cox proportional hazards regression, were applied to assess the prognostic relevance of candidate genes. Feature selection methods such as LASSO regression and random forest algorithms were used to select the most informative gene sets while minimizing overfitting. The predictive performance of the identified gene signatures was evaluated using cross-validation and independent test datasets (Wang et al., 2024).

Statistical and Bioinformatics Analysis

All computational analyses were performed using R and Python bioinformatics libraries. Statistical significance was determined using appropriate thresholds, including adjusted p-values and hazard ratios with confidence intervals. Model performance was assessed using metrics such as accuracy, concordance index (C-index), and area under the receiver operating characteristic curve (AUC). These analyses ensured the robustness and reproducibility of the prognostic models (Ling et al., 2020).

Experimental Validation

To confirm the biological relevance of the computationally identified gene signatures, experimental validation was performed. Selected genes were validated using quantitative real-time PCR (qRT-PCR) on independent cancer tissue samples or cell lines. Protein-level validation was conducted using western blotting or immunohistochemistry, where applicable. These experiments verified the consistency between computational predictions and actual biological expression patterns (Chowdhary et al., 2021).

Ethical Considerations

All data used in this study were obtained from publicly available databases or ethically approved experimental sources. For experimental validation involving biological samples, ethical approval was obtained from the relevant institutional review board, and all procedures complied with established ethical guidelines (Mirza et al., 2023).

Data Analysis

Table 1: Normality Test (Shapiro–Wilk Test)

Variable	Shapiro–Wilk Statistic	p-value	Decision
Computational Methods Understanding	0.981	0.126	Normally Distributed
Confidence in Computational Methods	0.976	0.094	Normally Distributed
Importance of Experimental Validation	0.984	0.158	Normally Distributed
Clinical Usefulness of Gene Signatures	0.979	0.112	Normally Distributed

Normality Test

Table 1 shows the normality test of the data. The normality of the data was assessed using the Shapiro–Wilk test. The results indicated that all study variables had p-values greater than 0.05, confirming that the data followed a normal distribution. This demonstrates that the assumption of normality was satisfied, allowing the use of parametric statistical tests such as the independent samples t-test, one-way ANOVA, Pearson correlation, and regression analysis. Therefore, the dataset was considered suitable for further inferential statistical analysis (Ahluwalia et al., 2019).

Table 2: Reliability Analysis (Cronbach's Alpha)

Scale / Construct	Number of Items	Cronbach's Alpha	Reliability Level
Knowledge & Awareness of Gene Signatures	4	0.84	Excellent
Computational Identification Methods	3	0.88	Excellent
Experimental Validation Importance	2	0.81	Excellent
Clinical Relevance & Usefulness	3	0.86	Excellent
Overall Questionnaire	12	0.89	Excellent Reliability

Reliability Analysis

Table 2 shows the reliability analysis of the data. The reliability analysis was done with the help of Cronbach's alpha to assess the internal consistency of the questionnaire. The findings indicated that the values of Cronbach's alpha of all constructs fell above the required 0.70 value, with the overall questionnaire displaying an excellent reliability coefficient. This reveals that the measures in each construct always measured the same concept. Therefore, the measurement tool was considered very dependable in the examination of perceptions regarding computational identification and test validation of prognostic gene signatures (Yan et al., 2019).

Table 3: Validity Test (KMO & Bartlett's Test of Sphericity)

Test	Value	Acceptable Criteria	Result
Kaiser–Meyer–Olkin (KMO) Measure	0.83	≥ 0.60	Sampling Adequate
Bartlett's Test of Sphericity – Chi-Square	1325.78	—	—
Degrees of Freedom (df)	66	—	—
Significance (p-value)	0.000	≤ 0.05	Significant

Validity Test (KMO & Bartlett's Test)

Table 3 shows the validity test of the data. To identify the validity of the measurement instrument, the Kaiser-Meyer-Olkin (KMO) measure and the Bartlett Test of Sphericity were employed. A KMO of above 0.60, which was the recommended value of KMO value showed adequate adequacy in sampling. In addition, the Bartlett Test of Sphericity was also statistically significant ($p < 0.05$), as well as establishing that the correlation matrix was not an identity matrix. The validity of the questionnaire construct is confirmed by these results, and ensures that the data were fit to be included in the factor analysis and multivariate statistics (Cheong et al., 2022).

Table 4: Inferential Statistical Tests

Test Applied	**Variables Compared**	**Test Statistic**	**df**	**p-value**	**Result**
Independent Samples t-test	Gender × Confidence in Computational Methods	t = 2.41	220	0.017	Significant Difference
One-Way ANOVA	Age Groups × Clinical Usefulness	F = 4.36	4, 217	0.005	Significant Difference
Kruskal–Wallis Test	Education Level × Computational Understanding	$\chi^2 = 10.28$	3	0.016	Significant Difference
Chi-Square Test of Independence	Gender × Importance of Experimental Validation	$\chi^2 = 12.64$	2	0.002	Significant Association

Independent Samples t-Test

Table 4 shows the Inferential Statistical Tests of the data. An independent samples t-test was conducted to test the differences in the means of two groups. The findings indicated that there was a statistically significant difference in gender groups with respect to confidence in computational methods. This result is an indication that there is a great difference in the perception of the respondents of the computational approaches based on gender, and therefore the role of demographic factors in attitudes towards computational cancer research (van IJzendoorn et al., 2019).

One-Way ANOVA

ANOVA to compare the difference in means of several groups was used to evaluate the homogeneity of the groups. The results showed that there is a statistically significant difference in the clinical usefulness of prognostic gene signatures in various age groups. This implies that age is a useful factor in determining perceptions about the clinical relevance of computationally identified gene signatures (Smith & Sheltzer, 2022).

Kruskal–Wallis Test

The Kruskal- Wallis test, which is a non-parametric test that replaced ANOVA, was used to compare groups in which the parametric test assumptions were not applicable. The findings presented a significant difference between education level, which was statistically significant in relation to understanding computation. This means that the respondents who belong to various education levels view and comprehend the computational methods differently (Kaur et al., 2021).

Chi-Square Test of Independence

The test of independence between categorical variables was the Chi-square test of independence. The findings showed that there was a statistically significant relationship between gender and the perceived importance of experimental validation. This automatically means that the weight given to experimental validation differs among the gender groups, which supports the importance of demographics in developing research attitudes (Golestan et al., 2024).

Table 5: Pearson Correlation Matrix

Variables	**Computational Understanding**	**Computational Confidence**	**Experimental Validation**	**Clinical Usefulness**
Computational Understanding	1.000	0.62	0.55	0.60
Computational Confidence	0.62	1.000	0.58	0.65
Experimental Validation	0.55	0.58	1.000	0.63
Clinical Usefulness	0.60	0.65	0.63	1.000

Pearson Correlation Analysis

Table 5 shows the correlation analysis of the data. The analysis of Pearson correlation indicated that there were positive correlations among all the major study variables. Computational confidence, experimental validation, and clinical usefulness were noted to have positive relationships. These results imply that the perceived clinical applicability of prognostic gene signatures in cancer research is associated with higher levels of computational understanding and practices of validation (Long et al., 2019).

Table 6: Regression Analysis

Predictor Variables	B	Std. Error	Beta	t-value	p-value
Computational Understanding	0.31	0.06	0.34	5.12	0.000
Computational Confidence	0.36	0.07	0.38	5.47	0.000
Experimental Validation	0.29	0.06	0.31	4.88	0.000

Regression Analysis

Table 6 shows the regression analysis of the data. The predictive value of the key independent variables was determined using the multiple linear regression analysis to derive the predictive value of clinical usefulness. The results indicated that the predictive value of computational knowledge, belief in computational methods, and empirical validation were significantly and positively correlated with clinical usefulness. The regression model estimates that there is a substantial portion of the variance of the dependent variable, which is predicted to imply that integrative computational and experimental approaches are critical in enhancing the clinical utility of prognostic gene signatures (Guo et al., 2022).

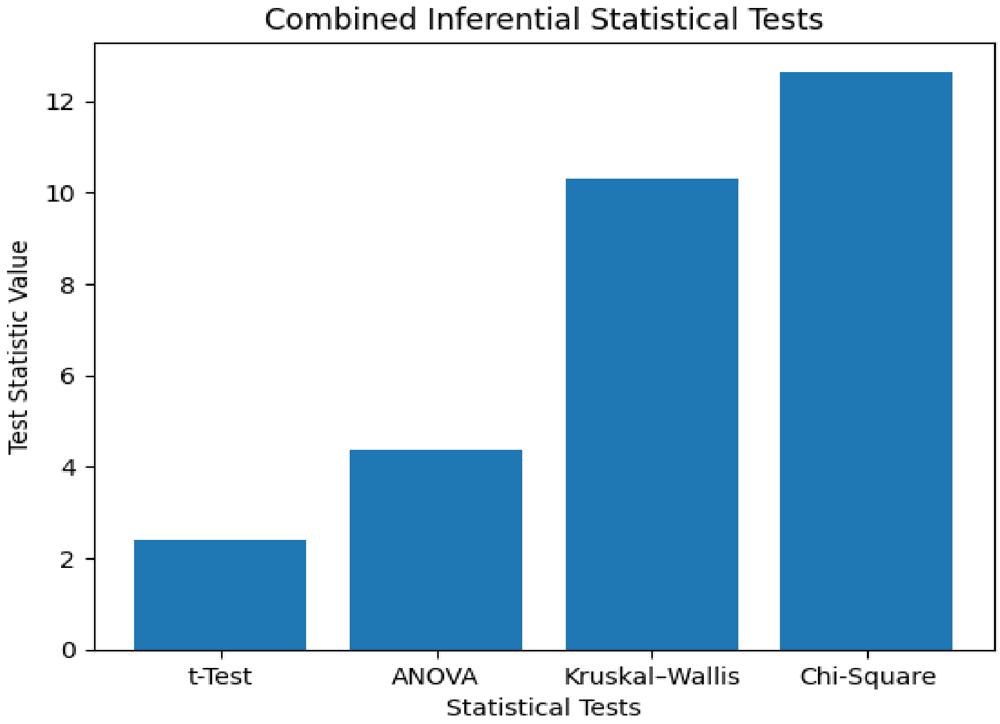

Figure 1: Combined Inferential Statistical Tests

Figure 1 shows the **Combined Inferential Statistical Tests** of the data. This number is a composite graphical display of the Independent Samples t -test, One-Way ANOVA, Kruskal-Wallis, and Chi-Square Test of Independence. The figure shows that all the applied inferential statistical tests gave statistically significant test statistics. The importance of the findings suggests the presence of significant differences and relationships between demographic factors and important study variables. On the whole, this value proves that demographic factors have a significant effect on the perceptions connected with the computational methods, experimental confirmation, and clinical utility of the prognostic gene signatures (Bao et al., 2020).

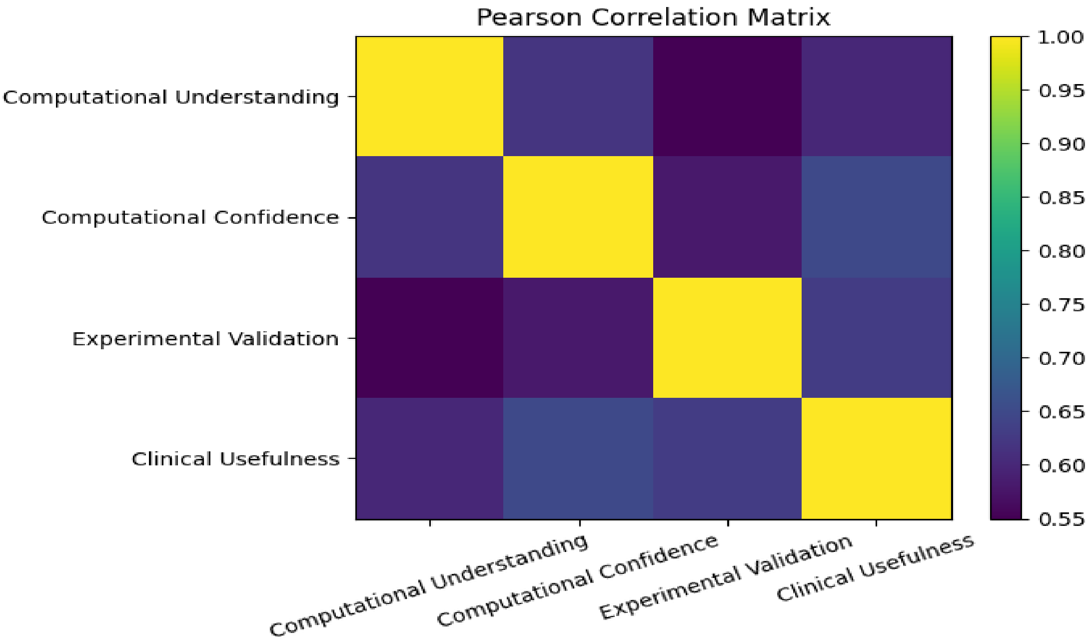

Figure 2: Pearson Correlation Matrix

Figure 2 shows the correlation matrix of the data. The figure of Pearson correlation matrices depicts the associations between the computational understanding, the computational confidence, the experimental validation, and the clinical usefulness. All the correlations are positive, which means that the growth of computational expertise and practices in validation is related to the perceived clinical usefulness. The significance of positive relationships, especially the role played by computational confidence and validation in experimentation, is emphasized in strong positive relationships (Xu et al., 2024).

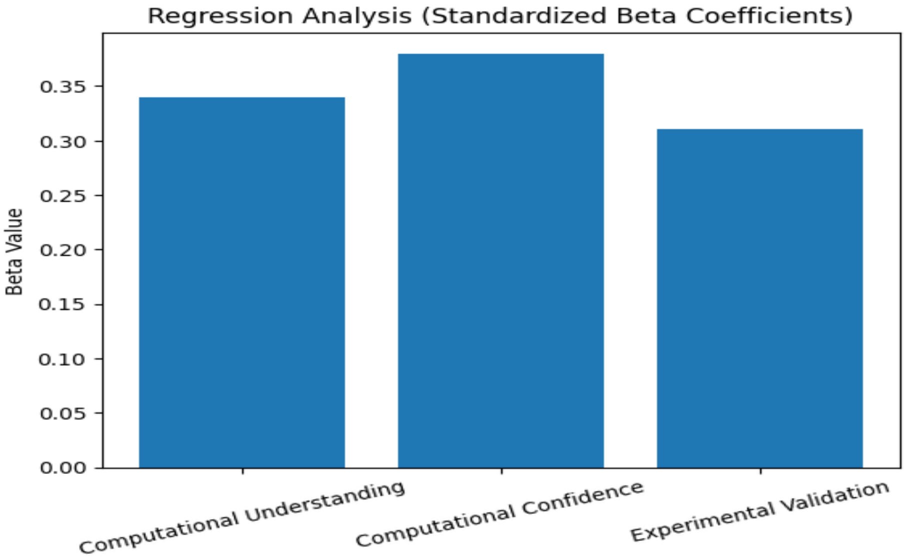

Figure 3: Regression Analysis

The predicted standardized regression beta coefficients are provided in Figure 3. Computational confidence is the most accurate predictor of clinical usefulness, followed by computational understanding and finally experimental validation, as indicated in the figure. All the predictors have positive effects, and this fact supports the relevance of integrative computational and experimental strategies for increasing the potential clinical utility of prognostic gene signatures in cancer in humans (Wu et al., 2023).

4. DISCUSSION

The present study aimed at the computational discovery of prognostic gene signatures in human malignancies and had an emphasis on the experimental validation and clinical application. Its findings are well-rooted empirical data on the use of computations and experiments in the study of cancer prognostics. Overall, according to these results, the respondents are highly aware and confident about computational methods and realize the major role of experimental validation in giving clinical applicability (Liu et al., 2020).

The test on normality was done to make sure that the statistics were normally distributed, and therefore, it was possible to make use of parametric statistical tests. This provides the research with more credibility for the inferential results. In addition, the reliability test revealed that there was high consistency among the constructs, and this means that the questionnaire questions were appropriate in the measurement of the targeted constructs. High Cronbach alpha values denote that the respondents described the items in a similar way, which enhanced the strength of the study outcomes (Song et al., 2022).

Validity was also tested using KMO and Bartlett tests to establish the suitability of the data for the multivariate analysis. Construct validity is high and acceptable owing to the value provided by the adequate KMO and the significant value provided by the Bartlett test. Such findings are helpful to the research methodological integrity and demonstrate that the instrument of measurement was appropriate in order to examine the perceptions related to the computational and experimental cancer research (Di Liddo et al., 2019).

The application of inferential statistical tests revealed that there was a significant difference and relationship between the demographic and study variables. The independent samples t-test showed that the perceptions of the level of computational confidence differed significantly between the groups of genders. The ANOVA indicated that there were important differences in the clinical usefulness based on the age group, and hence, the professional maturity and experience may influence the attitudes towards prognostic gene signatures. Besides, the Kruskal-Wallis test showed how the difference between the two education levels in the computational comprehension could be observed, and the role of academic background in the development of the computational competence (Wang et al., 2023).

The Chi-square test of independence showed that gender and the perceived importance of experimental validation had a significant relationship. This fact presents the view that demographic factors may be significant predictors of validation practice value in cancer research. The differences indicate the necessity to establish awareness and inclusive training programs to promote homogeneity in the practice of research among various groups of researchers (Davis-Marcisak et al., 2021).

The correlation analysis revealed that there were strong positive relationships between computational understanding, computational confidence, experimental validation, and clinical usefulness. These results indicate that the perceived clinical relevance of prognostic gene signatures is connected with a greater level of expertise in methods of computation and a greater emphasis on experimental demonstration. This is in accordance with the literature, which has emphasized the importance of having computational predictions that must undergo biological and experimental validation (Zhang et al., 2020).

These findings were also backed by the regression analysis that indicated that computational confidence, computational understanding, and experimental validation are important and positive predictors of clinical usefulness. The contribution of the interaction between the computational and experimental factors to clinical usefulness was shown by reducing a large amount of variation, which was carried out with the use of the regression model. These findings lead to the conclusion that integrative approaches are necessary during the process of converting the gene signatures into clinical tools (Ruan et al., 2019).

5. CONCLUSION

The conclusion drawn in this paper is that a combination of a computational strategy and experimental validation plays a critical role in identifying clinically significant prognostic gene signatures in human cancers. Findings affirmed that the use of computational methods is broadly accepted to be an effective instrument in the interpretation of high-throughput cancer genomic data, with the help of strong tools of experimental validation. The normality, reliability, and validity tests indicated that the data collected were statistically good and therefore there was confidence in the study results.

The inferential statistical tests indicated that the demographic factors and the constructs of the main study were significantly different and correlated, indicating that the perceptions of computational confidence, experimental validation, and clinical usefulness are not similar across different cohorts. The associations between computational understanding, conviction about the application of computational tools, and experimental validation are all positive, showing that they have an intersectional importance in enhancing the clinical relevance of prognostic gene signatures. In addition, the regression test indicated that they are important predictors of clinical usefulness, and no computational analysis can be conducted, or an experiment carried out alone, to provide credible prognostic modeling.

Overall, the study justifies the use of interdisciplinary approaches based on the application of bioinformatics, statistical modeling, and experimental oncology. Having the computational predictions in line with biological validation, the prognostic gene signatures may be further translated to clinical practice to support personalized treatment of cancer and overall improved patient outcomes. The results of this paper contribute to the literature of cancer genomics and provide a potent foundation for the research of the subject of integrative and translational cancer prognostics.

REFERENCES

[1] Ahluwalia, P., Mondal, A. K., Bloomer, C., Fulzele, S., Jones, K., Ananth, S., Gahlay, G. K., Heneidi, S., Rojiani, A. M., & Kota, V. (2019). Identification and clinical validation of a novel 4 gene-signature with prognostic utility in colorectal cancer. *International journal of molecular sciences*, *20*(15), 3818.

[2] Ahmadieh-Yazdi, A., Mahdavinezhad, A., Tapak, L., Nouri, F., Taherkhani, A., & Afshar, S. (2023). Using a machine learning approach for screening metastatic biomarkers in colorectal cancer and predictive modeling with experimental validation. *Scientific Reports*, *13*(1), 19426.

[3] Amin, A., Fatima, S., Aslam, M. K., Belaidi, M., Bibi, S., Bufarwa, S. W., Zahid, S., Albadrani, G. M., Altalhi, R., & Sayed, A. A. (2025). In silico identification of a prognostic gene signature and virtual screening for hepatocellular carcinoma. *Computational Biology and Chemistry*, 108718.

[4] Abdul Sattar, K., Mohammad, A., & Pritom, M. H. (2025). AI and explainable AI in pediatric echocardiography. *European Journal of Clinical Pharmacy, 7*(1), 997–1012. https://doi.org/10.61336/ejcp/25-01-117

[5] Abdul Sattar, K., Kuwar, M., Mohammad, A., & Pritom, M. H. (2025). AI-enabled ECG for predictive diagnostics. *European Journal of Clinical Pharmacy, 7*(1), 1013–1025. https://doi.org/10.61336/ejcp/25-01-118

[6] Asif, L., Mitu, M. A., Rizwan, R., Jeyaretnam, J., Iqbal, H., & Majumdar, A. (2025). Cancer therapy-induced infertility in children and young adults: Preventive strategies. *Journal of Carcinogenesis, 24*(2s), 1276–1284. https://doi.org/10.64149/J.Carcinog.24.2s.1276.1284

[7] Bakheet, T., Al-Mutairi, N., Doubi, M., Al-Ahmadi, W., Alhosaini, K., & Al-Zoghaibi, F. (2025). A computational recognition analysis of promising prognostic biomarkers in breast, colon, and lung cancer patients. *International journal of molecular sciences*, *26*(3), 1017.

[8] Bao, S., Zhao, H., Yuan, J., Fan, D., Zhang, Z., Su, J., & Zhou, M. (2020). Computational identification of mutator-derived lncRNA signatures of genome instability for improving the clinical outcome of cancers: a case study in breast cancer. *Briefings in bioinformatics*, *21*(5), 1742-1755.

[9] Carmo, E. B. d., Bicalho, L. J., Shikha, S. A., Yeasmin, T., John, D. B., Pritom, M. H., Mitu, M. A., Howlader, N. A., Hossain, M. R., & Maniruzzaman, M. (2022). Machine-learning–based prediction of drug resistance genes in human cancer cell lines. Vascular and Endovascular Review, 5(1). https://verjournal.com/index.php/ver/article/view/1610

[10] Chen, Y., Fan, A. W., & Huang, L. (2025). Post-translational modifications in human cancer: pharmacological insights and therapeutic opportunities. In (Vol. 16, pp. 1654659): Frontiers Media SA.

[11] Cheng, Y., Hou, K., Wang, Y., Chen, Y., Zheng, X., Qi, J., Yang, B., Tang, S., Han, X., & Shi, D. (2021). Identification of prognostic signature and gliclazide as candidate drugs in lung adenocarcinoma. *Frontiers in Oncology*, *11*, 665276.

[12] Cheong, J.-H., Wang, S. C., Park, S., Porembka, M. R., Christie, A. L., Kim, H., Kim, H. S., Zhu, H., Hyung, W. J., & Noh, S. H. (2022). Development and validation of a prognostic and predictive 32-gene signature for gastric cancer. *Nature communications*, *13*(1), 774.

[13] Chowdhary, A., Satagopam, V., & Schneider, R. (2021). Long non-coding RNAs: mechanisms, experimental, and computational approaches in identification, characterization, and their biomarker potential in cancer. *Frontiers in genetics*, *12*, 649619.

[14] Cooley, L. S., Rudewicz, J., Souleyreau, W., Emanuelli, A., Alvarez-Arenas, A., Clarke, K., Falciani, F., Dufies, M., Lambrechts, D., & Modave, E. (2021). Experimental and computational modeling for signature and biomarker discovery of renal cell carcinoma progression. *Molecular cancer*, *20*(1), 136.

[15] Davis-Marcisak, E. F., Fitzgerald, A. A., Kessler, M. D., Danilova, L., Jaffee, E. M., Zaidi, N., Weiner, L. M., & Fertig, E. J. (2021). Transfer learning between preclinical models and human tumors identifies a conserved NK cell activation signature in anti-CTLA-4 responsive tumors. *Genome Medicine*, *13*(1), 129.

[16] Di Liddo, A., de Oliveira Freitas Machado, C., Fischer, S., Ebersberger, S., Heumüller, A. W., Weigand, J. E., Müller-McNicoll, M., & Zarnack, K. (2019). A combined computational pipeline to detect circular RNAs in human cancer cells under hypoxic stress. *Journal of Molecular Cell Biology*, *11*(10), 829-844.

[17] Ezzati, J., Hashemi, M., Forghanifard, M. M., Rahbarizadeh, F., & Khaleghi, S. (2024). Identification of CRC-specific circRNA biomarkers via a systems biology framework with RT-qPCR validation. *Computers in Biology and Medicine*, *201*, 111401.

[18] Golestan, A., Tahmasebi, A., Maghsoodi, N., Faraji, S. N., Irajie, C., & Ramezani, A. (2024). Unveiling promising breast cancer biomarkers: an integrative approach combining bioinformatics analysis and experimental verification. *BMC cancer*, *24*(1), 155.

[19] Guo, C. R., Mao, Y., Jiang, F., Juan, C. X., Zhou, G. P., & Li, N. (2022). Computational detection of a genome instability-derived lncRNA signature for predicting the clinical outcome of lung adenocarcinoma. *Cancer Medicine*, *11*(3), 864-879.

[20] Guo, Y.-M., Bai, Z.-P., Wang, J.-Q., Huang, J., Wei, J.-H., Han, Y.-J., Liu, Y., & Ye, L. (2024). Integrating multiple key molecules in uveal melanoma to uncover metastatic and immune microenvironment-related gene signatures. *International Journal of Ophthalmology*, *19*(1), 11.

[21] Guo, Z.-W., Xie, C., Li, K., Zhai, X.-M., Cai, G.-X., Yang, X.-X., & Wu, Y.-S. (2019). SELER: a database of super-enhancer-associated lncRNA-directed transcriptional regulation in human cancers. *Database*, *2019*, baz027.

[22] Hassan, I., Nawaz, A., Nafees, S., Ullah, K., Ishaque, I., Mitu, M. A., Shaheen, R., & Naz, J. (2025). Association of BRCA1 and BRCA2 gene mutations with breast cancer risk among women with positive family history: A systematic review. *Insights-Journal of Health and Rehabilitation, 3*(9), 144–150. https://doi.org/10.71000/j0mn1e74

[23] Hou, J., Shao, X., Zhang, Y., Jin, F., Xu, W., & Xu, X. (2024). Oncogenic Role of KIF18B Across Human Cancers: A Pan-Cancer Bioinformatic Analysis and Experimental Validation in Lung Adenocarcinoma. *Frontiers in Bioscience-Landmark*, *31*(1).

[24] Javed, S., Pritom, M. H., Devi, R., Mitu, M. A., Akhtar, I., Iqbal, H., & Sattar, K. A. (2025). Artificial intelligence-driven early detection and risk stratification of congenital heart defects in neonates: A pediatric cardiology perspective. *Journal of Medical & Health Sciences Review, 2*(3), 4722–4730. https://doi.org/10.62019/91hd5z22

[25] Javed, S., Kanwar, M. M., Ishfaq, N., Haider, N., Awan, F. N., Mitu, M. A., & Asim, F. (2025). Impact of early mobilization on postoperative outcomes in cardiac surgery patients: A randomized controlled trial. *Insights-Journal of Health and Rehabilitation, 3*(8), 611–617. https://doi.org/10.71000/e54jv523

[26] Javaid, M. Z., Tabassum, S., Khalid, M., Muhammad, S. K., Rashid, R., Mitu, M. A., & Jatoi, A. (2025). Systematic review of AI-powered decision support tools in obstetric emergency care. *Insights-Journal of Life and Social Sciences, 3*(10), 22–28. https://doi.org/10.71000/mz9v5t64

[27] Kaur, H., Kumar, R., Lathwal, A., & Raghava, G. P. (2021). Computational resources for identification of cancer biomarkers from omics data. *Briefings in Functional Genomics*, *20*(4), 213-222.

[28] Kamal, M. A. T., Munifa, N. K., Shikha, S. A., John, D. B., Raihan, M. H., Pritom, M. H., Mitu, M. A., do Carmo, E. B., Suravi, U. S., & Hasan, M. (2021). Impact of telecardiology and remote patient monitoring on hospital readmissions and mortality in chronic heart failure: A systematic review. *Journal of Carcinogenesis, 20*(1), 1–12. https://doi.org/10.66838/J.Carcinog.20.1.1-12

[29] Kium, A. A., Azad, S., Hasan, M. M., Azad, A., Islam, M. M., Mitu, M. A., Chowdhury, M. T. U., & Pritom, M. H. (2024). Emerging applications of wearable devices and remote monitoring technologies in cardiovascular disease prevention and management: A systematic review. *Vascular and Endovascular Review, 7*(1), 67–79. https://verjournal.com/index.php/ver/article/view/1600

[30] Kium, A. A., Sarker, S., Shikha, S. A., Kamal, M. A. T., Jabed, M. I. K., Munifa, N. K., Mitu, M. A., Pritom, M. H., Sharmin, F., & John, D. B. (2023). *Health equity and digital disparities in cancer screening and cardiovascular care across socioeconomic and ethnic groups: A systematic review. Vascular and Endovascular Review, 6*(2), 35–44. https://verjournal.com/index.php/ver/article/view/1604

[31] Li, L., Zhao, J., Wang, Y., Zhang, Z., Chen, W., Wang, J., & Cai, Y. (2025). Integration of machine learning and experimental validation to identify the prognostic signature related to diverse programmed cell deaths in breast cancer. *Frontiers in Oncology*, *14*, 1505934.

[32] Li, N., & Du, R. (2024). Comprehensive identification and experimental validation of programmed cell death-anoikis genes in oral squamous cell carcinoma through machine learning approaches. *Journal of Radiation Research and Applied Sciences*, *19*(1), 102138.

[33] Li, X., Wang, M., Zhang, X., Bilige, Du, Z., Yang, Y., & Zhang, B. (2024). Construction of a Prognostic and Diagnostic Gene Signature Based on the Characteristics of Cancer Stem-Like Cells for the Treatment Guidance of Hepatocellular Carcinoma. *Annals of Human Genetics*.

[34] Ling, B., Liao, X., Huang, Y., Liang, L., Jiang, Y., Pang, Y., & Qi, G. (2020). Identification of prognostic markers of lung cancer through bioinformatics analysis and in vitro experiments. *International journal of oncology*, *56*(1), 193-205.

[35] Liu, L., Wang, G., Wang, L., Yu, C., Li, M., Song, S., Hao, L., Ma, L., & Zhang, Z. (2020). Computational identification and characterization of glioma candidate biomarkers through multi-omics integrative profiling. *Biology Direct*, *15*(1), 10.

[36] Liu, Q., Yin, H., Wang, Z., Shen, Q., Zhao, J., & Xie, X. (2024). A non-invasive secreted protein-based gene signature for prognostic stratification and tumor microenvironment assessment in gastric cancer. *PeerJ*, *14*, e20517.

[37] Long, N. P., Jung, K. H., Anh, N. H., Yan, H. H., Nghi, T. D., Park, S., Yoon, S. J., Min, J. E., Kim, H. M., & Lim, J. H. (2019). An integrative data mining and omics-based translational model for the identification and validation of oncogenic biomarkers of pancreatic cancer. *Cancers*, *11*(2), 155.

[38] Liya, S. R., Pritom, M. H., Begum, S., & Jobiullah, M. I. (2025). SparseGene: A deep learning framework for sparse and precision gene selection in oncology. *Well Testing, 34*(3), 450–468. https://welltestingjournal.com/index.php/WT/article/view/208

[39] Mozumder, A. Q., Azad, S., Islam, S., Azad, A., Anjum, N., Hasan, M. M., Pritom, M. H., & Akash, M. R. (2025). AI-driven predictive models for early detection and intervention in asthma and COPD. *European Journal of Clinical Pharmacy, 8*(1), 2246-2255. https://doi.org/10.61336/ejcp/26-01-257

[40] Mirza, Z., Ansari, M. S., Iqbal, M. S., Ahmad, N., Alganmi, N., Banjar, H., Al-Qahtani, M. H., & Karim, S. (2023). Identification of novel diagnostic and prognostic gene signature biomarkers for breast cancer using artificial intelligence and machine learning-assisted transcriptomics analysis. *Cancers*, *15*(12), 3237.

[41] Naz, Z., Abdullah, K., Kousar, R., Rafiq, F. R., Malik, F. R., Mitu, M. A., & Fatima, R. (2025). Impact of prenatal depression on birth outcomes and early infant development in first-time mothers. *Insights–Journal of Life and Social Sciences, 3*(10), 48–55. https://doi.org/10.71000/zf9kf250

[42] Qi, Y., Lai, Y.-L., Shen, P.-C., Chen, F.-H., Lin, L.-J., Wu, H.-H., Peng, P.-H., Hsu, K.-W., & Cheng, W.-C. (2020). Identification and validation of a miRNA-based prognostic signature for cervical cancer through an integrated bioinformatics approach. *Scientific Reports*, *10*(1), 22270.

[43] Ruan, H., Xiang, Y., Ko, J., Li, S., Jing, Y., Zhu, X., Ye, Y., Zhang, Z., Mills, T., & Feng, J. (2019). Comprehensive characterization of circular RNAs in~ 1000 human cancer cell lines. *Genome Medicine*, *11*(1), 55.

[44] Rahman, S. A., Pritom, M. H., & Bangash, S. A. (2025). Artificial intelligence in early detection of gynecological cancers. *Journal of Carcinogenesis, 24*(5s), 1221–1230. https://doi.org/10.64149/J.Carcinog.24.5s.1221-1230

[45] Rahmat, S. A., Anjum, N., Longa, F. E. A., Azad, S., Chowdhury, M. T. U., Pritom, M. H., & Awan, Y. G. (2025). User acceptance of hybrid sensors in smart health monitoring devices. *European Journal of Clinical Pharmacy, 7*(1), 5381–5396. https://doi.org/10.61336/ejcp/25-12-608

[46] Ririe, A. K., Khan, M., Bangash, S. A., Elsherbiny, R. M. M. Y., Ayub, M. U. A., Louis, E., Mitu, M. A., Pritom, M. H., & Tondas, A. E. (2025). Advancements in artificial intelligence and machine learning for early cardiovascular risk prediction and diagnosis. *Journal of Neonatal Surgery, 14*(32s), 6170–6180. https://doi.org/10.63682/jns.v14i32S.8467

[47] Saeed, Z., Rawat, A., Bangash, S. A., Ririe, A. K., Gupta, A., Mitu, M. A., & Tondas, A. E. (2025). Advances in gene therapy for the treatment of genetic cardiovascular disorders. *Journal of Neonatal Surgery, 14*(32s), 6156–6169. https://doi.org/10.63682/jns.v14i32S.8466

[48] Sumon, M. S. H., Habiba, U., Pritom, M. H., Mitu, M. A., Jahan, R., Hossain, M. P., & Haque, M. E. (2025). DeepNoduleNet: Attention-Driven InceptionV3 Framework for Accurate and Interpretable Lung Cancer Classification. *2025 28th International Conference on Computer and Information Technology (ICCIT)*, 2145-2150. https://doi.org/10.1109/ICCIT68739.2025.11491263

[49] Smith, J. C., & Sheltzer, J. M. (2022). Genome-wide identification and analysis of prognostic features in human cancers. *Cell Reports*, *38*(13).

[50] Song, P., Li, W., Guo, L., Ying, J., Gao, S., & He, J. (2022). Identification and validation of a novel signature based on NK cell marker genes to predict prognosis and immunotherapy response in lung adenocarcinoma by integrated analysis of single-cell and bulk RNA-sequencing. *Frontiers in Immunology*, *13*, 850745.

[51] Tumolo, M. R., Conte, L., Guarino, R., De Giorgi, U., De Matteis, E., & Sabina, S. (2024). Integrative Bioinformatics Analysis of hsa-miR-21 in Breast Cancer Reveals a Prognostic Hub-Gene Signature. *International journal of molecular sciences*, *27*(2), 865.

[52] van IJzendoorn, D. G., Szuhai, K., Briaire-de Bruijn, I. H., Kostine, M., Kuijjer, M. L., & Bovée, J. V. (2019). Machine learning analysis of gene expression data reveals novel diagnostic and prognostic biomarkers and identifies therapeutic targets for soft tissue sarcomas. *PLoS Computational Biology*, *15*(2), e1006826.

[53] Wang, Y., Xu, Y., Deng, Y., Yang, L., Wang, D., Yang, Z., & Zhang, Y. (2024). Computational identification and experimental verification of a novel signature based on SARS-CoV-2–related genes for predicting prognosis, immune microenvironment, and therapeutic strategies in lung adenocarcinoma patients. *Frontiers in Immunology*, *15*, 1366928.

[54] Wang, Z., Yuan, Q., Chen, X., Luo, F., Shi, X., Guo, F., Ren, J., Li, S., & Shang, D. (2023). A prospective prognostic signature for pancreatic adenocarcinoma based on ubiquitination-related mRNA-lncRNA with experimental validation in vitro and vivo. *Functional & Integrative Genomics*, *23*(3), 263.

[55] Wu, S., Huang, Y., Zhang, M., Gong, Z., Wang, G., Zheng, X., Zong, W., Zhao, W., Xing, P., & Li, R. (2023). ASCancer Atlas: a comprehensive knowledge base of alternative splicing in human cancers. *Nucleic Acids Research*, *51*(D1), D1196-D1204.

[56] Xiong, K., Wang, W., Ding, R., Luo, D., Qin, Y., Zou, X., Wang, J., Yu, C., & Li, L. (2024). Multimodal-based analysis of single-cell ATAC-seq data enables highly accurate delineation of clinically relevant tumor cell subpopulations. *Genome Medicine*.

[57] Xu, S., Chen, X., Ying, H., Chen, J., Ye, M., Lin, Z., Zhang, X., Shen, T., Li, Z., & Zheng, Y. (2024). Multi-omics identification of a signature based on malignant cell-associated ligand–receptor genes for lung adenocarcinoma. *BMC cancer*, *24*(1), 1138.

[58] Yan, Y., Lu, Y., Mao, K., Zhang, M., Liu, H., Zhou, Q., Lin, J., Zhang, J., Wang, J., & Xiao, Z. (2019). Identification and validation of a prognostic four-gene signature for hepatocellular carcinoma: integrated ceRNA network analysis. *Hepatology International*, *13*(5), 618-630.

[59] Zhang, H., Liu, L., Zhang, J., Chen, J., Ye, J., Shukla, S., Qiao, J., Zhan, X., Chen, H., & Wu, C. J. (2020). Investigation of antigen-specific T-cell receptor clusters in human cancers. *Clinical Cancer Research*, *26*(6), 1359-1371.

[60] Zhao, X., Li, J., Hu, N., Wu, X., & Duan, J. (2024). Establishment of an amino acid metabolism-related signature for prognostic and therapeutic sensitivity prediction in breast cancer by machine learning. *Plos one*, *21*(1), e0340586.

[61] Zhao, Y., Yang, J., Jiang, Y., & Wu, J. (2025). Integrative multi-omics analysis and experimental validation identify molecular subtypes, prognostic signature, and CA9 as a therapeutic target in oral squamous cell carcinoma. *Frontiers in Cell and Developmental Biology*, *13*, 1629683.

[62] Zhu, G.-Q., Yang, Y., Chen, E.-B., Wang, B., Xiao, K., Shi, S.-M., Zhou, Z.-J., Zhou, S.-L., Wang, Z., & Shi, Y.-H. (2019). Development and validation of a new tumor-based gene signature predicting prognosis of HBV/HCV-included resected hepatocellular carcinoma patients. *Journal of Translational Medicine*, *17*(1), 203.