International Journal of Biology and Life Sciences
ISSN: 2957-9511 | Vol. 5, No. 1, 2024

A New Polyamine Metabolism-related Gene Signature Predicts Breast Cancer Prognosis

Ruiqi Liu, Xiaoqian Huang, Xiaozhou Chen *

School of Mathematics and Computer Science, Yunnan Minzu University, Kunming, Yunnan, 650504, China

* **Corresponding author**: Xiaozhou Chen (Email: ch_xiaozhou@163.com)

Abstract: Polyamine metabolism is involved in several cellular processes in organisms and is closely associated with tumorigenesis, progression, and metastasis. In this study, using breast cancer data from The Cancer Genome Atlas (TCGA) database as well as polyamine metabolism-related genes obtained from previous studies, key genes were screened using LASSO regression, and a prognostic label was created, which was validated using an independent cohort, Metabric. The breast cancer cohort of this prognostic label TCGA was divided into high-risk and low-risk groups, and the prognosis of high-risk and low-risk groups was significantly different. The biological indicators of high-risk and low-risk groups were significantly different, and the prediction results of this model were validated in the Metabric cohort. Our model can effectively predict the prognosis of breast cancer patients based on the characteristics of polyamine metabolism-related genes, which are also related to immune cell infiltration and immunotherapy response, and have some potential for clinical application.

Keywords: Polyamine Metabolism; Breast Cancer; Tumor Microenvironment; Prognosis.

1. Introduction

With the continuous advancement of medical research, society is increasingly concerned about breast cancer (BC). Breast cancer is not only a killer that jeopardizes women's health, but also brings a heavy burden to many governments and families. In recent years, the incidence of breast cancer has shown an increasing trend year by year [1], and according to the newly released Annual Cancer Report 2023, breast cancer has become the first cancer with the highest incidence rate among women [2,3]. In this context, it is important to find more effective methods to predict the prognosis of breast cancer patients.

such as gene regulation, cell proliferation, differentiation, and death. Various enzymes and genes are involved in polyamine metabolic pathways to ensure that polyamine synthesis and degradation are in equilibrium, thus maintaining polyamine homeostasis requires stringent cellular regulatory processes [4-6]. Different biological processes and disease states may affect the regulation of polyamine metabolism, and it has been demonstrated that the homeostasis of polyamine metabolism is dysregulated in many tumors, which is closely related to cancer development [7,8]. Therefore, it would be highly likely to establish a new prognostic signature linking tumor genotype and phenotype through polyamine metabolism-related genes.

In this study, we obtained 59 polyamine metabolism-related genes from previous studies and screened 11 key genes closely related to the prognosis of breast cancer patients by LASSO regression. These genes play important roles in cellular metabolism, protein degradation, and other biological processes. We also created a novel prognostic label that not only verified excellent performance in predicting patient prognosis in the training set but also obtained satisfactory results in the test set. Based on the results of the prognostic model analysis, we demonstrated the heterogeneity of polyamine metabolism in the tumor microenvironment of breast cancer as well as its comprehensive clinical application, which not only has important theoretical value but also promises to provide patients with more precise therapeutic options.

2. Results

2.1. Association between the Genetic Characterization of Polyamine Metabolism and Clinical Prognosis in Breast Cancer

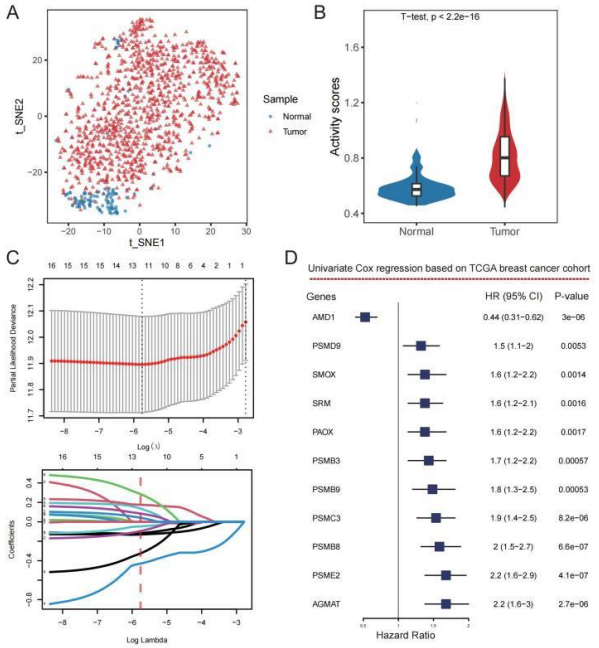

Figure 1. Polyamine metabolism-related genes affect the prognosis of breast cancer patients

(A) Tsne plot showing the difference in the distribution of polyamine metabolism-related gene expression in breast cancer normal and cancer samples. (B) Violin plot showing the difference in polyamine metabolism-related gene activity scores in normal and cancer samples of breast cancer. P values were obtained by t-test. (C) LASSO regression was used for

polyamine metabolism-related genes and 11 key genes were screened. (D) Forest plot showing the prognostic impact of the 11 key genes for polyamine metabolism on the TCGA breast cancer cohort.

To investigate the relationship between polyamine metabolism and the prognosis of breast cancer patients, we downloaded the transcriptomic data and clinical data of the breast cancer cohort from the TCGA database and collected 59 polyamine metabolism-related genes based on the existing studies on polyamine metabolism. A significant difference in the distribution of polyamine metabolism-related gene expression levels between tumor samples and normal samples of breast cancer could be seen (Figure 1A). By calculating the activity scores of polyamine metabolism-related genes it can be seen that the activity scores of tumor samples were significantly higher than those of normal samples (Figure 1B). To make the results of the subsequent analysis more valid and reliable, we screened 59 genes using LASSO regression and obtained 11 key genes (Figure 1C). These genes play important roles in cellular metabolism, protein degradation, and other biological processes. The one-way COX regression analysis showed that all these 11 key genes were significantly associated with the prognosis of breast cancer patients, and most of the genes were shown to be risk factors (Figure 1C). The above analysis indicates that polyamine metabolism shows significant heterogeneity in tumor and normal samples of breast cancer and can significantly affect the prognosis of patients.

2.2. Constructing Prognostic Labels for Breast Cancer Associated with Polyamine Metabolism

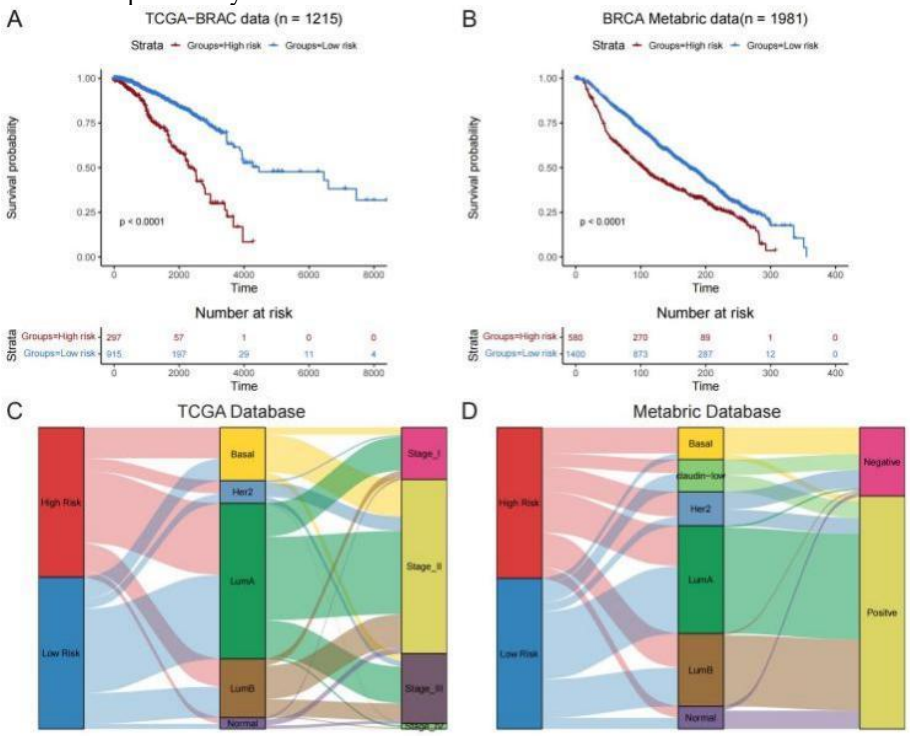

Figure 2. Prognostic models based on polyamine metabolism-related genes has good predictive performance

(A) KM curves showing the prognostic difference between the high-risk and low-risk groups in the TCGA cohort. High-risk and low-risk groups were divided by the median risk score. p-values were assessed by log-rank test. (B) KM curves showing prognostic differences between high- and low-risk groups in the Metabric dataset. High-risk and low-risk groups were divided by the median risk score. p-values were assessed by log-rank test. (C) Sankey plot showing the association between high and low-risk groups and tumor status in the TCGA cohort. (D) Sankey plot showing the association between high and low-risk groups and tumor status in the Metabric dataset.

We used the 11 polyamine metabolisms obtained from the LASSO regression, weighted and summed their expression levels with their regression coefficients in the multivariate Cox regression analysis to build a new prognostic label. The resulting value was called the risk score, which was used to classify the breast cancer cohort of the TCGA into high-risk and low-risk groups according to the median of the risk scores, and we downloaded the Metabric dataset as a test set. It can be seen that both in the training set (TCGA cohort) and the test set (Metabric dataset), the polyamine metabolism-based prognostic labeling demonstrated a good ability to predict prognosis (Figure 2A, B). By comparing the associations of the high-risk and low-risk groups with other tumor characteristics, we found that both Basal and Her2 subtypes were more predominant in the high-risk group, both in the TCGA and Metabric cohorts (Figure 2C, D). All these results are significant in that the higher the risk score, the worse the prognosis of breast cancer patients.

2.3. Biological Significance of Breast Cancer Risk Score

(A) Bubble plots showing the biological processes significantly involved in the TCGA cohort and the high- and low-risk groups in the Metabric dataset, respectively. (B) Bubble plots showing pathways significantly involved in the TCGA cohort and the high-risk and low-risk groups in the Metabric dataset, respectively. (C) Violin plots showing the difference in the distribution of immune scores between the high-risk and low-risk groups in the TCGA cohort and the Metabric dataset. p-values were obtained by the Wilcoxon

rank-sum test. (D) Violin plot showing the difference in the distribution of tumor extent between the high-risk and low-risk groups in the TCGA cohort and the Metabric dataset. p-values were obtained by the Wilcoxon rank sum test. (E) Violin plots showing the difference in the distribution of EMT scores between the high-risk and low-risk groups in the TCGA cohort and Metabric dataset. p-values were obtained by the Wilcoxon rank sum test. (F) Heatmaps showing the cancer signature pathway activity scores for the high and low-risk groups in the TCGA cohort and Metabric datasets, respectively. These pathways were obtained from the MSigDB database. (G) Heatmap showing the immune cell gene set activity scores for the high and low-risk groups in the TCGA cohort and Metabric dataset, respectively.

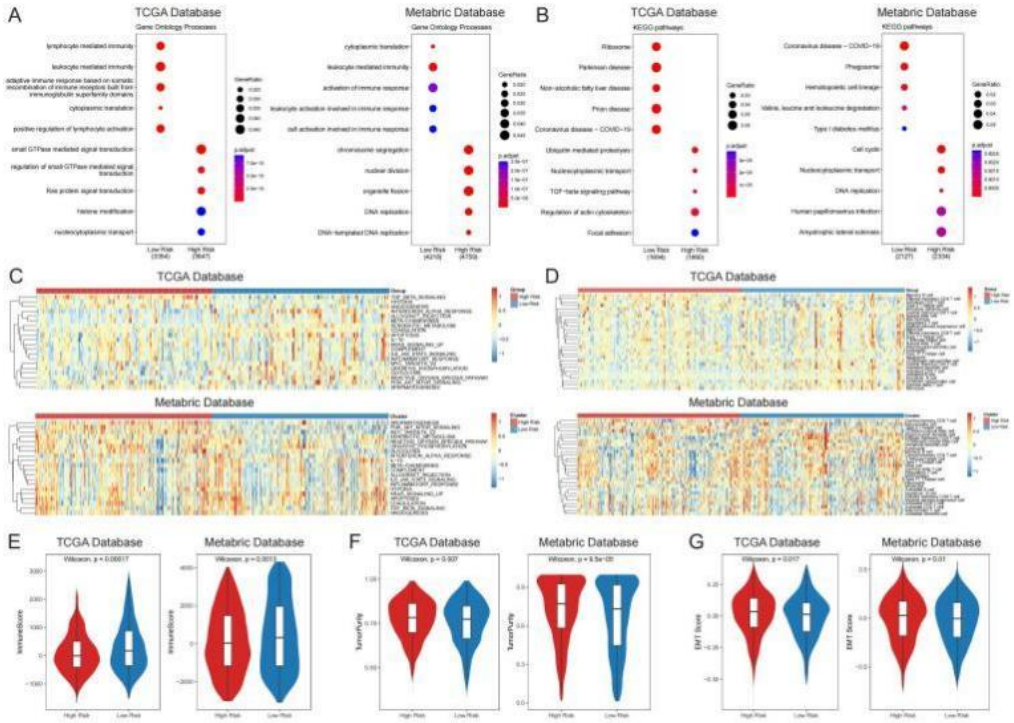

Figure 3. Biological differences between high- and low-risk groups

To explore the biological functions associated with the high-risk and low-risk groups, we performed GO and KEGG analyses. GO analyses showed that in both the TCGA cohort and the Metabric dataset, the low-risk group was closely associated with leukocyte-mediated immunity, and cytoplasmic translation, and the high-risk group did not show the same biological terms (Figure 3A). KEGG analysis showed that both low-risk groups were closely associated with Coronavirus disease-COVID-19, and both high-risk groups were closely associated with Nucleocytoplasmic transport (Figure 3B). We also assessed the cancer signature pathway activity scores as well as the degree of immune infiltration in the high- and low-risk groups [9], and both the TCGA cohort and the Metabric dataset showed the same trend (Figure 3C, D). In addition, we introduced the ESTIMATE tool to calculate the immune score, tumor purity, and EMT score of the high- and low-risk groups, and it can be seen that the distributions of the three types of biological indexes in the high- and low-risk groups showed significantly identical differences, with the high-risk group having a significantly lower immune score, and significantly higher tumor purity and EMT scores (Fig. 3E, F, G), and all of the above results verified that the high-risk group had a poor prognosis.

2.4. Constructing a Prognostic Column-line Diagram Model based on Risk Scores

(A) Nomogram showing the performance of multivariate Cox regression analysis by combining multiple predictors, including risk score, age, and tumor status. (B) Calibration curves show the agreement between the 3-year DFS probability predicted by the nomograms in the TCGA cohort and Metabric dataset and the actual 3-year DFS. (C) Calibration curves showing the agreement between the 5-year DFS probability predicted by the nomogram in the TCGA cohort and Metabric dataset and the actual 5-year DFS. (D) Calibration curves showing the agreement between the 10-year DFS probabilities predicted by the nomograms in the TCGA cohort and Metabric datasets and the actual 10-year DFS.

Finally, we constructed a nomogram model to predict the survival of breast cancer patients. In the TCGA cohort, the nomogram incorporated three features: risk score, patient age, and tumor stage (Figure 4A). In the Metabric dataset, the model incorporated three features: risk score, patient age, and immunohistochemistry (Figure 4B). And, we plotted the calibration curves of 3-, 5-, and 10-year disease-free survival (DFS) of the TCGA cohort and the Metabric dataset with the standard curves, respectively, and the two curves roughly overlapped, suggesting that the survival probability predicted by the nomogram was very close to the actual survival probability (Figure 4C, D). In summary, the prediction model based on the polyamine metabolic risk score is likely to provide effective guidance in the clinic.

3. Materials and Methods

3.1. Data Sources

Transcriptomic data and clinical data for this study were downloaded from the TCGA website (https:// portal. gdc. cancer. gov/) and the cbio portal (https://www.cbioportal.org).

The TCGA cohort had a total of 1,215 breast cancer patient samples and the Metabric dataset had a total of one breast cancer patient sample.

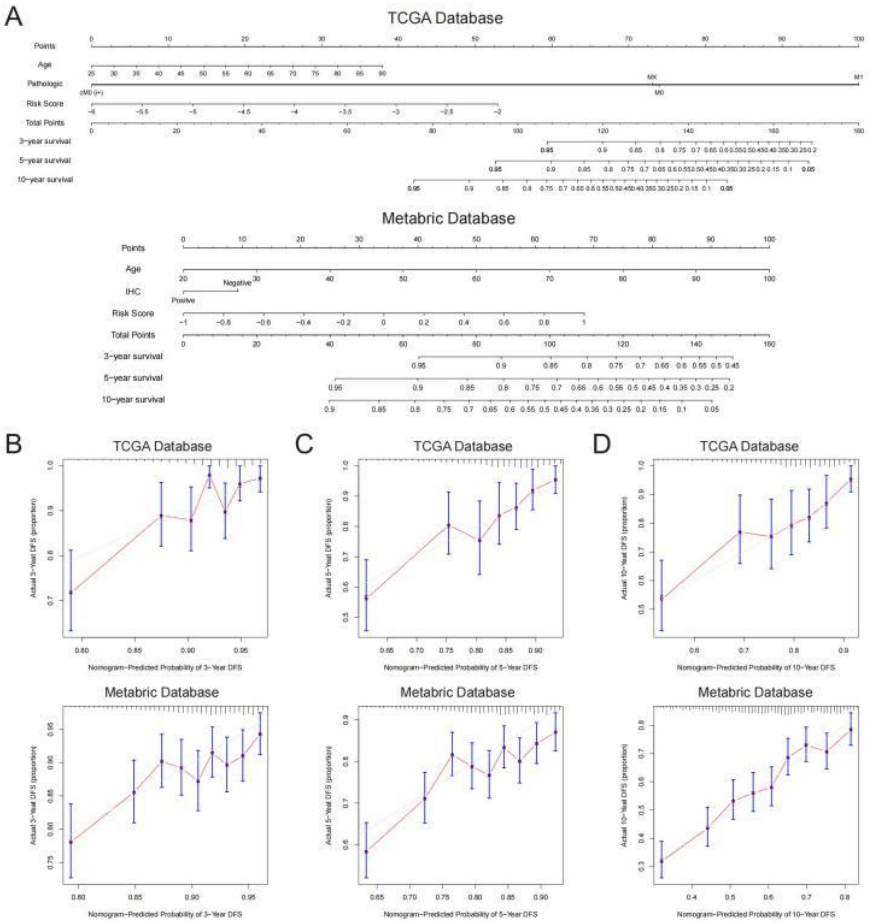

Figure 4. Comprehensive nomogram model based on risk scores

3.2. Calculation of Enrichment Scores

We used "method = ssgsea" in the GSVA software package (version 1.48.0) [10] to detect the activity scores of polyamine metabolism-related genes in the TCGA cohort and the Metabric dataset for breast cancer patients.

3.3. Development of a Prognostic Model based on Polyamine Metabolism

To create a valid prognostic label, we used the TCGA dataset as a training set, processed the genes associated with polyamine metabolism in the training set using the LASSO algorithm, screened out 11 genes to ensure simplicity of the model, and then assessed the correlation between the expression of the 11 genes in the polyamine metabolism-related traits and the overall survival of the patients in the training set using the multivariate Cox proportional risk regression model. The following equation obtained the risk score for each patient:

$$\text{risk score (patient)} = \sum_{i=1}^{11} \text{coef}(g_i) \times \text{expr}(gi)$$

where 11 is the number of key genes, g_i represents the normalized expression value of gene i, and coef(g_i) represents the corresponding risk regression coefficient.

3.4. Prognostic Analysis

We categorized patients in both the TCGA cohort and the Metabric dataset into high-risk and low-risk groups based on the best cutoffs inferred using the surv_cutpoint function in the "survminer" software package. Kaplan-Meyer (KM) survival analysis and the log-rank test were used to determine whether there was a significant difference in overall survival between each high- and low-risk group.

3.5. Functional Enrichment Analysis

We used the "clusterProfiler" software package (version 4.0.5) [11] to convert gene symbols to Entrez IDs using the bitr function, followed by enrichment analyses of the Gene Ontology (GO) and KEGG pathways using the enrichGO function with the "ont = BP" parameter and the enrichKEGG function, respectively. Statistical significance was determined as a p-value less than 0.05.

3.6. Immune Characterization

We used the "ESTIMATE" software package (version 1.0.13) [12] to estimate the immune scores, tumor purity, and EMT scores of the high-risk and low-risk groups.

3.7. Statistical Analysis

Standard statistical tests such as Student's t-test, Wilcoxon rank sum test, and log-rank test were used to analyze the clinical and expression data in this study. p-value less than 0.05 was considered statistically significant. All statistical analyses were performed in R4.3.0.

4. Conclusion

In this study, we investigated the expression patterns of polyamine metabolism genes in breast cancer and screened 11 key genes using LASSO regression, which were found to be very closely associated with the prognosis of breast cancer patients in the TCGA cohort based on the multivariate Cox risk regression model. We developed a prognostic label, based on which we assigned a risk score to each patient in the TCGA cohort, and classified the patients into high-risk and low-risk groups, and the results showed that our model could predict the survival of patients very well, and the prognosis of the high-risk group was significantly worse than that of the low-risk group, and such a result was validated in another independent cohort and the Metabric dataset as well. There were also significant differences between high and low-risk groups about breast cancer pathologic characteristics.

To explore the biological significance of the model, we performed GO and KEGG analyses and calculated the cancer signature pathway activity scores as well as the degree of immune infiltration in the high, low-risk groups, in addition, we introduced the ESTIMATE tool to calculate the immune scores, tumor purity as well as the EMT scores in the high and low-risk groups. All results demonstrated significant biological differences between the high-risk and low-windup groups, and these biological differences may lead to a worse prognosis in the high-risk group. Finally, we constructed a column-line graph model based on the polyamine metabolic risk score, and the ability of the column-line graph to predict patient prognosis was similarly validated in the TCGA cohort and the Metabric dataset.

In summary, we collected gene signatures related to polyamine metabolism, investigated their relationship with the tumor microenvironment of breast cancer as well as patient prognosis, and constructed a prognostic model for breast cancer based on polyamine metabolism, which effectively predicts the prognosis of patients, and has a certain potential for application, and may provide meaningful guidance in the clinic.

Acknowledgments

XZC designed the study. RQL did data analysis and wrote the manuscript. XQH revised the manuscript. We would like to express our heartfelt gratitude to those who have contributed to the creation and maintenance of the databases utilized in this study.

References

[1] Fan, L.; Strasser-Weippl, K.; Li, J.J.; St Louis, J.; Finkelstein, D.M.; Yu, K.D.; Chen, W.Q.; Shao, Z.M.; Goss, P.E. Breast cancer in China. The Lancet. Oncology 2014, 15, e279-289, doi:10.1016/s1470-2045(13)70567-9.

[2] Jemal, A.; Siegel, R.; Ward, E.; Murray, T.; Xu, J.; Thun, M.J. Cancer statistics, 2007. CA: a cancer journal for clinicians 2007, 57, 43-66, doi:10.3322/canjclin.57.1.43.

[3] Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA: a cancer journal for clinicians 2021, 71, 209-249, doi:10. 3322/ caac. 21660.

[4] Hesterberg, R.S.; Cleveland, J.L.; Epling-Burnette, P.K. Role of Polyamines in Immune Cell Functions. Medical sciences (Basel, Switzerland) 2018, 6, doi:10.3390/medsci6010022.

[5] Pegg, A.E. Functions of Polyamines in Mammals. The Journal of biological chemistry 2016, 291, 14904-14912, doi:10. 1074/ jbc. R116.731661.

[6] Dever, T.E.; Ivanov, I.P. Roles of polyamines in translation. The Journal of biological chemistry 2018, 293, 18719-18729, doi:10.1074/jbc.TM118.003338.

[7] Casero, R.A., Jr.; Murray Stewart, T.; Pegg, A.E. Polyamine metabolism and cancer: treatments, challenges and opportunities. Nature reviews. Cancer 2018, 18, 681-695, doi: 10. 1038/s41568-018-0050-3.

[8] Guo, Y.; Ye, Q.; Deng, P.; Cao, Y.; He, D.; Zhou, Z.; Wang, C.; Zaytseva, Y.Y.; Schwartz, C.E.; Lee, E.Y.; et al. Spermine synthase and MYC cooperate to maintain colorectal cancer cell survival by repressing Bim expression. Nature communications 2020, 11, 3243, doi:10.1038/s41467-020-17067-x.

[9] Charoentong, P.; Finotello, F.; Angelova, M.; Mayer, C.; Efremova, M.; Rieder, D.; Hackl, H.; Trajanoski, Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. Cell reports 2017, 18, 248-262, doi: 10. 1016/j.celrep.2016.12.019.

[10] Hänzelmann, S.; Castelo, R.; Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC bioinformatics 2013, 14, 7, doi:10.1186/1471-2105-14-7.

[11] Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics: a journal of integrative biology 2012, 16, 284-287, doi:10.1089/omi.2011.0118.

[12] Yoshihara, K.; Shahmoradgoli, M.; Martínez, E.; Vegesna, R.; Kim, H.; Torres-Garcia, W.; Treviño, V.; Shen, H.; Laird, P.W.; Levine, D.A.; et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nature communications 2013, 4, 2612, doi:10.1038/ncomms3612.