



# NAPH-Fluorescence Lifetime Imaging Informed Machine Learning Modelling Reliably Predicts Temozolomide Responsiveness in Glioblastoma

Aldo Pastore<sup>1,2\*</sup>, Elena Corradi<sup>1,2\*</sup>, Mariangela Morelli<sup>2</sup>, Chiara Maria Mazzanti<sup>2+</sup>, Paolo Aretini<sup>2+</sup>

## Abstract

Glioblastoma (GBM) is a highly deadly brain tumor. The chemotherapeutic treatment still lacks solid patient stratification, as temozolomide (TMZ) is administered to the majority of GBM patients. In this study, we explored the effectiveness of NAD(P)H-fluorescence lifetime imaging microscopy (NAD(P)H-FLIM) in furnishing clinically relevant insights into GBM responsiveness, a realm constrained by the absence of corresponding clinical outcome data. Using the information obtained by NAD(P)H-FLIM, we conducted a DE analysis on an RNA-seq private dataset, comparing TMZ responder and non-responder tumors. To validate the NAD(P)H-FLIM classification, we conducted a comparable DE analysis on the GBM TCGA (The Cancer Genome Atlas) RNA-seq data using the progression-free interval (PFI) as a responsiveness indicator.

We selected the most informative genes shared by both the DE analyses (BIRC3, CBLC, IL6, PTX3, SRD5A1, TNFAIP3) and employed them as transcriptomic signature. Using a different dataset (GBM TCGA Agilent-Microarray), we built a signature-based machine learning model capable of predicting the PFI. We also showed that the performance of our model is similar to that obtained with a well-established biomarker: the methylation status of the MGMT promoter. In conclusion, we assessed the reliability of the NAD(P)H-FLIM in providing clinically relevant drug response information in GBM and provided a new transcriptomic based model for determining patients' responsiveness to TMZ treatment.

**Keywords:** Glioblastoma; NAPH-FLIM; Temozolomide; Explants; TCGA; Machine-learning; Biomarkers

## Introduction

Glioblastoma (GBM) is the most common and aggressive primary brain tumor in adults. With a median survival of approximately 15 months, the prognosis for glioblastoma is poor [1, 2]. The standard of care for newly diagnosed glioblastoma is maximum safe surgical resection followed by radiotherapy with concurrent and adjuvant chemotherapy. Over the last two decades, temozolomide (TMZ) has been the main chemotherapeutic agent used in the clinical practice of glioblastoma [3], but it has not provided an effective cure over the long term. In fact, the majority of patients still experience recurrence, primarily due to the molecular heterogeneity of GBM tumors. Unfortunately, the known prognostic factors and predictive biomarkers do not explain the complexity and variability of GBM treatment response [4-6], and despite several models having been developed to assess the survival of an individual patient [7-9], none of them is currently used in clinical practice. In the last few years, the need to increase the personalized

## Affiliation:

<sup>1</sup>Laboratorio NEST, Scuola Normale Superiore, Pisa, Italy

<sup>2</sup>Fondazione Pisana per la Scienza, San Giuliano Terme, Pisa, Italy

\*co-first authors, +co-last authors

## \*Corresponding author:

Aldo Pastore, Laboratorio NEST, Scuola Normale Superiore, Pisa, Italy.

Elena Corradi, Laboratorio NEST, Scuola Normale Superiore, Pisa, Italy.

**Citation:** Aldo Pastore, Elena Corradi, Mariangela Morelli, Chiara Maria Mazzanti, Paolo Aretini. NAPH-Fluorescence Lifetime Imaging Informed Machine Learning Modelling Reliably Predicts Temozolomide Responsiveness in Glioblastoma. Journal of Biotechnology and Biomedicine. 7 (2024): 369-378.

**Received:** July 31, 2024

**Accepted:** August 05, 2024

**Published:** August 14, 2024

medicine applicability has led to a revival of interest in functional assays [10-12]. These tests basically consist of evaluating the drug response directly on biopsied tumor cells, and they can be used to monitor the course of the disease and adopt adaptive treatment regimens [13]. However, in the field of neuro-oncology, *in vitro* drug screening methods based on functional assays are rarely translated into clinical practice. This is partially due to the high inter- and intra-variability of brain tumors, which is difficult to represent in *in vitro* models. Moreover, traditional assays of drug response, mainly based on cellular proliferation and viability, require extended culturing times during which the samples diverge from the parental tumor *in vivo*. Despite these challenges, it is crucial to develop and validate new functional precision oncology approaches that can improve the success of clinical treatments. In the last decade, many researchers have addressed their efforts in validating an innovative functional strategy: the application of Fluorescence Lifetime Imaging Microscopy (FLIM) to *in vitro* drug testing [14-16]. FLIM overcomes the main limitations of traditional drug screening methods as it is a rapid, label-free, and non-destructive imaging technique that can be applied to living cells or tissues [17]. In this context, FLIM exploits the intrinsic auto-fluorescence molecular properties of NAD(P)H, a metabolic enzymatic cofactor associated with the metabolic state of cells/tissues. Briefly, the measured NAD(P)H fluorescence is used as an indicator of the relative balance between oxidative phosphorylation and glucose catabolism [18], which has been shown to be an early predictor of cellular drug response [19]. The use of NAD(P)H-FLIM as a tool to assess treatment response in glioblastoma was presented by Morelli et al. [20, 21]. They applied this technique to investigate the response to TMZ of GBM explants, which are an *ex vivo* pre-clinical model consisting of minimally handled fresh tumor tissues. However, as in most functional assays [22], the clinical course of the GBM cohort collected by Morelli et al. [21] is not yet available. This lack of information limits the translation of this approach into clinical practice because the response to TMZ predicted by NAD(P)H-FLIM cannot be directly confirmed with the clinical outcome.

To overcome this limitation and investigate the informative nature of the NAD(P)H-FLIM technique in predicting clinical outcomes, in this paper we propose a bioinformatics pipeline that utilizes external cohorts of GBM patients and leverages the progression-free interval (PFI) as an approximation of drug response. With this aim we found a panel of genes highlighted by the FLIM classification and in the same time playing a role in clinical outcomes of the external dataset.

Finally, to further assess the clinical relevance of the identified genes, and consequently of the NAD(P)H-FLIM functional assay, we attempted to use them as predictive biomarkers by building a machine learning model. The model

capability of assessing the clinical outcome on a different GBM cohort would, in turn, provide both a strong evidence of the NAD(P)H-FLIM reliability as a drug testing assay and a tool to effectively stratify GBM patients.

## Materials and Methods

### GBM explants cohort

**GBM explants cohort description:** In this study, we used the data presented by Morelli et al. [21]. These data are related to 33 samples of glioblastoma multiforme (GBM) that were cultured as explants, an *ex vivo* pre-clinical model consisting of minimally handled fresh tumor tissues. Out of the 33 GBM samples, 18 were sequenced as fresh-frozen samples, while the other 15 as GBM explants cultured *in vitro*. For the sake of simplicity, we will refer to this dataset as the “GBM explants”.

All GBM explants were exposed to temozolomide (TMZ) drug. The protocol for drug treatment, as well as the culture methods of GBM explants, are described in Morelli et al. [21].

To determine which GBM explants were responsive to the benchmark drug (temozolomide, TMZ), Morelli et al. labelled the samples as responder (RES) or non-responder (NRES) based on the NAD(P)H-fluorescence lifetime imaging microscopy (FLIM) technique [21]. In this context, FLIM exploits the intrinsic auto-fluorescence molecular properties of NAD(P)H, a metabolic enzymatic cofactor associated with the metabolic state of the cell/tissue.

**RNA-seq data, alignment and annotation:** We aligned the data using STAR (version 2.7.9a) against the GRCh38 version of the human genome. Read counts were generated during the alignment process (as described on the GDC guideline: [https://docs.gdc.cancer.gov/Data/Bioinformatics\\_Pipelines/Expression\\_mRNA\\_Pipeline/](https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/Expression_mRNA_Pipeline/)). We obtained the raw count matrix using an R script. All subsequent analyses were conducted using R version 4.2.0 within the RStudio 2022.07.02+576 suite or Python version 3.10.

**Batch correction:** Although recent works reported that fresh-frozen and cultured tissues have similar gene expression levels [23, 24], we investigated the presence of a batch effect in the RNA-seq data due to the different sample types. With this aim, we performed principal component analysis (PCA) using the PCATools R library. PCA revealed that the samples labelled by sample type clustered along the principal components 1 and 7, confirming the presence of a batch effect. To correct for this, we applied the ComBat-seq algorithm, which is specifically designed for RNA-seq data [25]. In addition, to preserve the biological variability of the samples, we set the ComBat-seq argument “group” as the NAD(P)H-FLIM classification (RES/NRES). Finally, we applied PCA again to verify that the corrected data were no longer clustered by sample type.

**Differential expression analysis:** After the RNA-seq data of the 33 GBM samples were batch-corrected, we normalized them and conducted a differential expression (DE) analysis using the DESeq2 algorithm [26]. The DE analysis compared the two groups of samples identified by the NAD(P)H-FLIM technique: RES (16 samples) and NRES (17 samples). To reduce the effect of genes with low counts, we corrected the estimated log-fold values using the normal shrinkage method [26]. We identified differentially expressed genes (DEGs) as the ones with a fold change greater than 1.5 in absolute value (log2-fold change greater than 0.58) and p value adjusted for multiple comparisons less than 0.1. A custom R script was used to perform the analysis.

### RNA-seq data of TCGA

**TCGA RNA-seq cohort description:** To validate the NAD(P)H-FLIM based classification, we conducted a comparable DE analysis on the GBM RNA-seq data sourced from The Cancer Genome Atlas (TCGA). We downloaded the RNA-seq dataset and the corresponding clinical data from the Genomic Data Commons (GDC) Data Portal. The raw counts table was obtained using the same alignment and annotation procedures as the GBM explants. From the whole set of patients, we only selected the ones treated for at least one cycle with TMZ (104 patients).

**Differential expression analysis:** Among the TMZ treated patients, we only selected the ones whose PFI value was non-censored. Then, we grouped them based on the progression-free interval (PFI): we defined two groups of patients with extreme PFI values, one group with PFI values within the 1<sup>st</sup> quartile (Short PFI, or S-PFI group, 22 patients) and the other with PFI values within the 4<sup>th</sup> quartile (Long PFI, or L-PFI group, 22 patients). Finally, the differential expression (DE) analysis was conducted on the selected RNA-seq data following the same approach and statistical thresholds used for the GBM explants dataset. Specifically, we used the DESeq2 algorithm to normalize raw counts data and then we conducted a DE analysis comparing the L-PFI with the S-PFI group [26]. Differentially expressed genes (DEGs) were identified as the ones with a log2-fold change (corrected using normal shrinkage) higher than 0.58 (fold change higher than 1.5) and a p-value adjusted lower than 0.1. Data selection was performed using a custom Python script, while the DE analysis was conducted using a custom R script.

### Functional-clinical differentially expressed genes

By only considering the DEGs whose expression pattern was consistent (preserved up- or down-regulation) among the two datasets, we found 19 DEGs shared by the two DE analyses. Subsequently, in order to select the most informative genes, we performed a supervised feature selection. With this aim we first standardized the RNA-seq expression values of

the 19 shared DEGs of both GBM explants and TCGA. Next, we aggregated the two standardized datasets into a unique dataset (77 samples) and the Long PFI (L-PFI) label was converted to TMZ responder (RES), as well as the Short PFI (S-PFI) label was converted to TMZ non-responder (NRES). Finally, we applied a linear discriminant analysis (LDA, [27]) and selected the genes having a discriminant coefficient (i.e. the loading) higher than the mean of the coefficients, as shown in Figure 3B. This process led to the selection of 6 genes among the initial 19. The described analyses were conducted in Python using custom-made scripts.

### Microarray data of TCGA Agilent

Here, we selected a different dataset: the TCGA AgilentG4502A\_07\_2 Microarray, downloaded from UCSC (University of California Santa Cruz) Xena ([https://tcga-xena-hub.s3.us-east-1.amazonaws.com/download/TCGA.GBM.sampleMap%2FagilentG4502A\\_07\\_2.gz](https://tcga-xena-hub.s3.us-east-1.amazonaws.com/download/TCGA.GBM.sampleMap%2FagilentG4502A_07_2.gz)). We decided to use a new dataset to avoid any information leakage. Additionally, we excluded the 36 patients who were also present in the RNA-seq dataset of TCGA. To be consistent with the previous analyses, we only selected patients who were treated with TMZ, resulting in a final dataset of 247 patients.

### Building a predictive model

To make the results as interpretable as possible, we implemented the Cox proportional hazards model using the lifelines library developed in Python [28, 29]. An important advantage of using a Cox regression is that a hazard ratio (HR) is computed for each of the model's covariates, which, in our case, are the 6 selected genes. Each HR gives insights on the effect of the corresponding covariate on the predicted clinical variable (progression-free in our case), i.e., whether it increases or decreases the event probability. Finally, the model can be used to divide subjects into two risk groups, i.e., high risk or low risk, by using the partial hazard (PH) assigned to each patient by the model.

### Leave-One-Subject-Out (LOO) training and testing procedure

To train and test the model we adopted a Leave-One-Subject-Out (LOO) approach, rather than a classical training and testing procedure. Practically, the dataset was split into train and test set, with the latter consisting of only one sample at a time. A Cox model was then trained using all the samples except for the one left out, which was used to test the model. This procedure was repeated as many times as the total number of patients, with each patient left out of the training set in turn and treated as a new observation by the model. It is important to remark that the train set related to two different iterations differs from just one observation, i.e., the left-out-patient.

Finally, for each patient in the dataset we obtained a risk prediction, which is called partial hazard when using the Cox regression. We split the patients into high and low risk groups by using as threshold the median of the computed partial hazards. By conducting a univariate Cox regression that uses as covariate the risk group, we can compute the corresponding hazard ratio (HR), which indicates the model ability to correctly categorize patients into high and low risk groups.

## Statistical analyses

Throughout this paper, we used the Kaplan-Meier plot to visualize time-to-event data. Log-rank test (<https://lifelines.readthedocs.io/en/latest/lifelines.statistics.html>) was applied to compare the progression-free interval (PFI) distributions of two sets of data. The Wilcoxon rank-sum test (<https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.wilcoxon.html>) was used to compare the risk scores distribution of MGMT-methylated and MGMT-unmethylated patients.

## Results

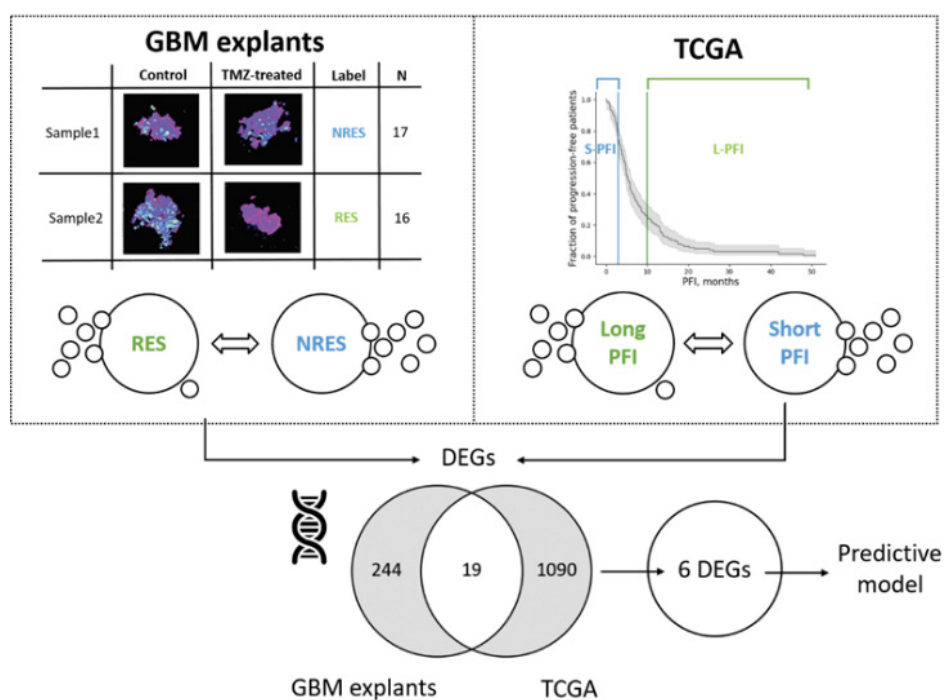
### Differential Expression Analyses

The workflow of the bioinformatics approach we propose to assess the NAD(P)H-FLIM technique is depicted in Figure 1. Firstly, we conducted a differential expression (DE) analysis on the RNA-seq data of the GBM explants by

comparing the RES and NRES group. Here we identified 244 differentially expressed genes (DEGs, defined as in 2.2.2), as shown in the volcano plot of Figure 2A. Out of the 244 DEGs, 172 of them were up-regulated in the NRES condition, while the remaining 72 down-regulated. Secondly, we conducted a comparable DE analysis on the GBM RNA-seq data of TCGA using the progression-free interval (PFI) to group patients, as described in section 2.2.2. As a result, 1090 genes were found to be differentially expressed when comparing the L-PFI (n = 22) and S-PFI (n = 22) groups, as shown in the volcano plot of Figure 2B. Out of the 1090 DEGs, 853 of them were up-regulated, while the remaining 237 down-regulated. Finally, we identified 19 genes that were differentially expressed and that exhibited consistent expression patterns (up- or down-regulation) in both the DE analyses.

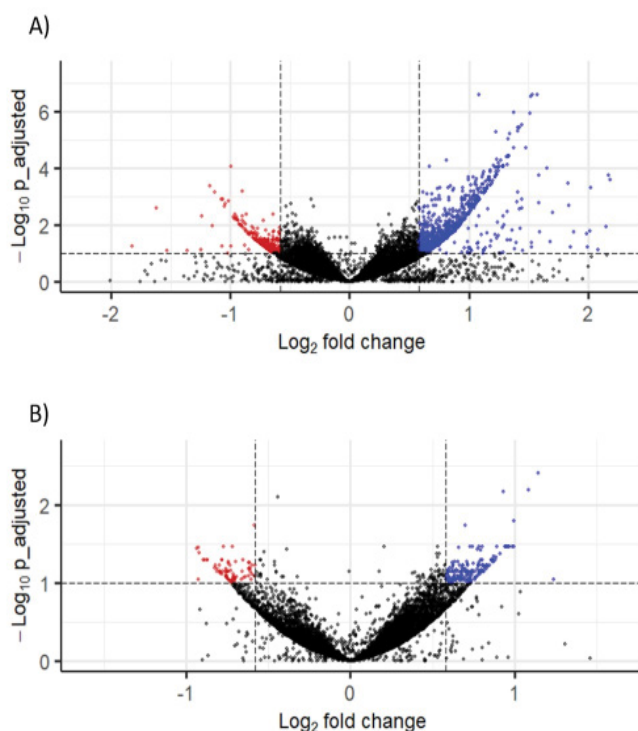
### A transcriptomic-based predictive model

Genes selection: From the original set of 19 shared DEGs, we identified a subset of the most informative ones. To select the genes, we applied a linear discriminant analysis (LDA) to the aggregated RNA-seq data of GBM explants and TCGA. LDA was used to identify the genes that contributed the most in separating RES and L-PFI samples from NRES and S-PFI samples. In figure 3A the LDA projected data are shown and are notably well separated according to their true label (RES+L-PFI/NRES+S-PFI), confirming that the discriminant algorithm nicely worked. In Figure 3B we show the LD scaling value assigned to each of the 19 DEGs. The higher the scaling

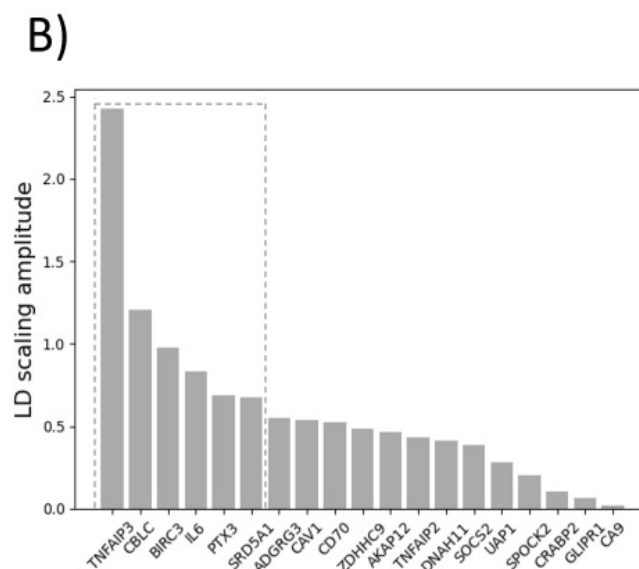
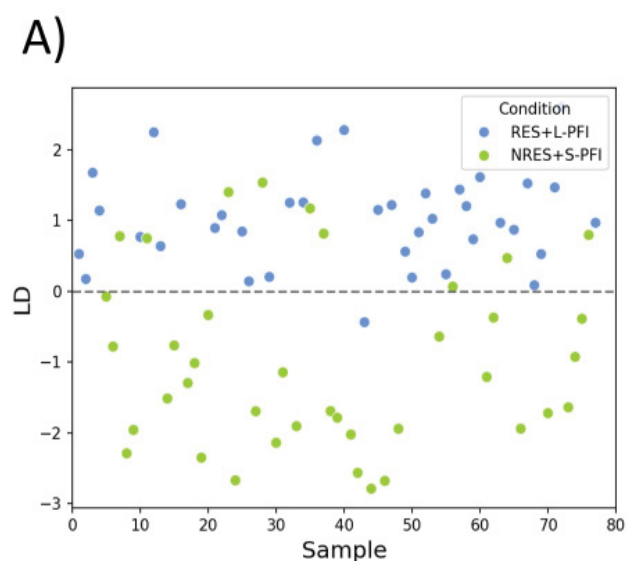


**Figure 1:** Qualitative workflow of the bioinformatics validation of NAD(P)H-FLIM technique. The left box shows the differential expression (DE) analysis conducted on the GBM explants classified as RES or NRES. The right box shows the DE analysis conducted on the RNA-seq data of TCGA by comparing patients with extreme progression-free interval (PFI) values. From the 19 shared DEGs, the 6 most informative ones were retained and used to build a predictive model.





**Figure 2:** Volcano plots resulting from the DE analyses of RNA-seq data. The log2 fold change between the two conditions is shown on the x-axis, while the minus log10 p adjusted is shown on the y-axis. The dotted lines indicate the threshold values (0.58 for the absolute value of the log2 fold change and 1 for the minus log10 p adjusted). The red dots show the down-regulated DEGs, while the blue dots show the up-regulated DEGs. The black dots are not DEGs. A) TCGA cohort. B) GBM explants cohort



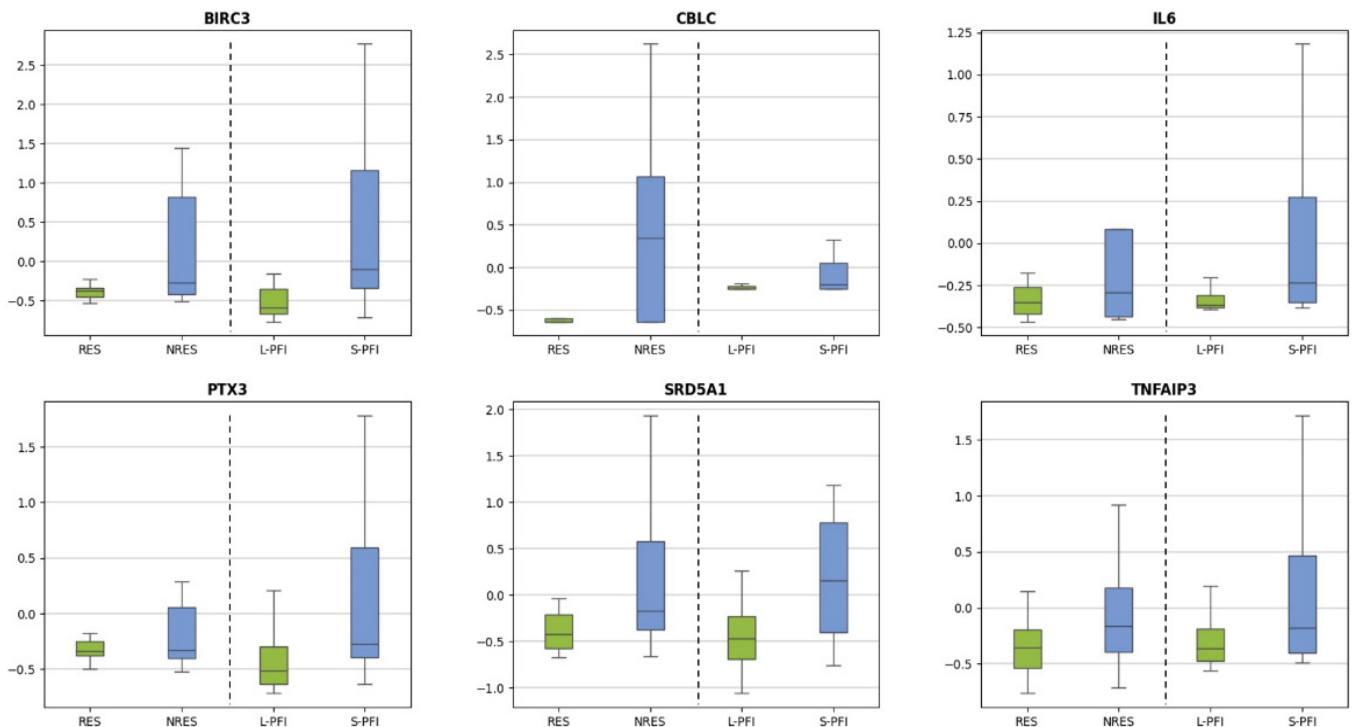
**Figure 3:** A) Linear Discriminant Analysis (LDA) projected data of the aggregated RNA-seq data of GBM explants and TCGA. The projected data are separated according to their label (RES+L-PFI and NRES+S-PFI), as highlighted by the dotted line. B) Linear Discriminant (LD) scaling value of the 19 genes. The selected most informative genes are the ones framed with the dotted line.

value, the greater the relevance of that gene in discriminating between the two labels. The selection process is described in details in section 2.3 and it led to the selection of 6 genes: BIRC3, CBLC, IL6, PTX3, SRD5A1, and TNFAIP3. The standardized expression levels of these genes are shown in Figure 4, where all of them show up-regulation in the NRES and S-PFI cases compared to the RES and L-PFI cases.

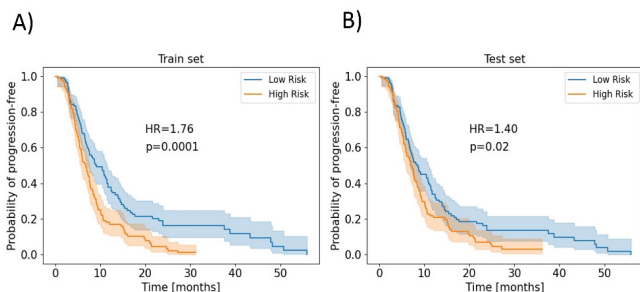
**Model design:** We employed the microarray data of TCGA Agilent, described in section 2.4, to build a model based on the Cox regression [28], described in sections 2.5 and 2.6. Our model fits a hazard function to progression-free interval (PFI) data using the expression values of the 6 identified genes and returns, for each patient, a label corresponding to the predicted risk class (High Risk/Low Risk). Figures 5A and 5B show the Kaplan-Meier curves of the two groups produced by the model on the train and test set, respectively. The hazard ratio (HR) computed on the High Risk/Low Risk variable is 1.76 for the train set (for one of the N splits of the full dataset) and 1.40 for the test set. The p values calculated with the LogRank test assessed that the PFI related to the two risk groups significantly differed both in the train and test set ( $p = 0.0001$  and  $p = 0.02$  for the train and test set, respectively).

### MGMT promoter methylation status

The MGMT promoter methylation status is a well-established clinical biomarker for GBM (4). Here we assess and compare its predictive capability with our model. In accordance with the previous analyses, we only selected the TCGA Agilent patients who were treated with TMZ and had information about their MGMT status (145 patients).

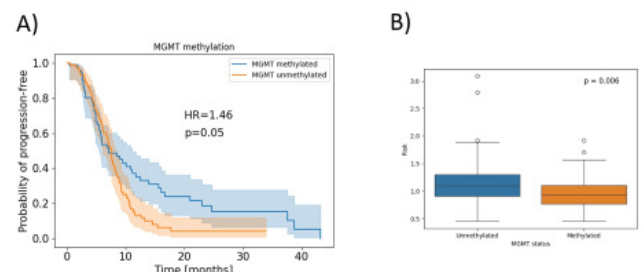


**Figure 4:** RNA-seq standardized expression level of the 6 selected DEGs: BIRC3, CBLC, IL6, PTX3, SRD5A1, and TNFAIP3. The direction (up- or down-regulation) of expression is coherent in the GBM explants and TCGA cohorts. All the 6 DEGs are up-regulated in the RES or L-PFI cases (green boxplot) compared to the NRES or S-PFI cases (light blue boxplot).



**Figure 5:** Kaplan-Meier curves of progression-free interval on the microarray data of TCGA Agilent. The High Risk and Low Risk groups were identified by the predictive model based on the 6-genes transcriptional signature. The hazard ratio (HR) between the two risk classes and p-value (LogRank test) are also displayed. A) Train set, B) Test set.

Then, we divided all 145 patients based on their methylation status. Figure 6A shows the Kaplan-Meier curves of the methylated (72 patients) and unmethylated (73 patients) groups. The hazard ratio (HR) obtained from the univariate Cox regression whose covariate is the MGMT status is 1.46, while the p value calculated with the LogRank test is at the significance threshold ( $p = 0.05$ ). Considering all the 145 patients grouped into methylated/unmethylated, we retrieved the matched risk scores (i.e., the partial hazard assigned by our model when testing). By using the Wilcoxon rank-sum test, we found that the methylated had significantly lower



**Figure 6:** A) Kaplan-Meier curves of progression-free interval of the MGMT methylated and unmethylated patients. The hazard ratio (HR) between the two groups and p-value (LogRank test) are also displayed. B) Risk scores assigned by our model to the MGMT unmethylated (blue) and methylated patients (orange). The unmethylated ones have significantly higher risk scores (Wilcoxon rank-sum test,  $p < 0.01$ ).

risk scores than unmethylated patients ( $p < 0.01$ , Figure 6B), pointing out the consistency of the MGMT status and model predictions. Finally, we focused on Agilent patients with extreme PFI values, grouping them into Agi S-PFI or Agi L-PFI groups using the criterion described in section 2.2.2. Among this subset of peculiar patients, we analyzed the degree of consistency between our model predictions and MGMT status. The congruent predictions resulted to be 60.8% of the total suggesting that the information provided by the two indicators did not fully overlap.

## Discussion

The molecular heterogeneity in glioblastoma (GBM) has led to limited treatment options and poor prognosis [1, 2]. The current lack of personalized chemotherapeutic therapies can be attributed to the limited predictive capability of known molecular biomarkers in determining drug response of an individual patient [4-6, 9]. In this context, there has been a growing interest in functional tests applied to precision oncology [12, 22]. These tests serve as drug screening assays by measuring changes induced by drug perturbations on tumor tissues derived from patients. One promising tool to improve the personalized treatment of malignancies is the functional assay based on NAD(P)H-FLIM imaging [14-16, 19]. Morelli et al. recently employed this technique to assess the response of GBM explants to temozolomide (TMZ) [20, 21]. Since TMZ is administered to the majority of GBM patients, it serves as a benchmark drug (Stupp et al., 2005). While Morelli et al. obtained biological and molecular confirmations of the validity of this approach, the clinical translation of NAD(P)H-FLIM, being a novel functional assay, is hindered by the lack of retrospective clinical data. In the present study, we solidify the clinical significance of the NAD(P)H-FLIM assay in GBM by addressing the challenge of the lack of matched clinical outcome data. Since direct confirmation of this functional assay is currently unfeasible due to the unavailability of clinical outcomes matched with the GBM explants we analysed, we have designed and implemented a novel bioinformatics approach.

### Functional-clinical differentially expressed genes

Starting from the cohort of GBM explants presented by Morelli et al. [21], we assessed the presence of differentially expressed genes (DEGs) between the groups of GBM samples classified as responsive (RES) or non-responsive (NRES) based on the NAD(P)H-FLIM investigation. The differential expression (DE) analysis unveiled 244 DEGs. This finding suggests that the classification derived from this functional approach reflects differences observed at the gene-expression level. We verified the treatment-related relevance of the identified DEGs by conducting a separate DE analysis using an external dataset, specifically the GBM RNA-seq data of TCGA. We ensured that the new analysis was strictly consistent, in terms of data acquisition and processing, with the previous one. Additionally, we only included TCGA patients who had received TMZ treatment, as the NAD(P)H-FLIM technique was utilized to evaluate the responsiveness to this drug. To perform the DE analysis on the RNA-seq data of TCGA, we categorized the selected patients based on the matched progression-free interval (PFI). We chose PFI as the variable for grouping because of its role in approximating the response to TMZ treatment. Indeed, the PFI is defined as “the length of time during and after the treatment of a disease, such as cancer, that a patient lives with the disease but it does not get worse” (<https://www.cancer.gov/publications/dictionaries/cancer-terms/def/pfs>). In the context of our study, PFI serves as a measure of treatment effectiveness or patient response to the therapeutic treatment, specifically TMZ.

To emphasize the gene expression differences associated to PFI, we defined two groups of patients with extreme PFI values: the short-PFI group, or S-PFI, and the long-PFI group, or L-PFI. The two groups we defined are similar to the non-responsive (NRES) and the responsive (RES) groups of the GBM explants cohort, respectively. Using the same differential expression algorithm and parameters as before, we identified 1090 DEGs when comparing the S-PFI and L-PFI groups. Having obtained these two sets of DEGs, related to both NAD(P)H-FLIM classification and the length of the PFI, we proceeded to intersect them. We found 19 DEGs that were shared by the two sets and exhibited consistent regulation (preserved up- or down-regulation). We remark that these candidate genes were identified by overlapping results obtained from entirely different cohorts. Crucially, the presence of a non-empty intersection indicates a candidate panel of genes that are involved together in the tumour metabolic alterations identified by the NAD(P)H-FLIM and in the clinical disease containment.

**Testing the genes predictive capability with machine learning**

To further assess the clinical relevance of the identified genes, and consequently of the NAD(P)H-FLIM functional assay, we attempted to use them as predictive biomarkers by building a machine learning model. Specifically, we evaluated their potential to predict the progression-free interval on a GBM cohort not used for previous analyses. With this aim, we performed a features selection to reduce the number of candidate biomarkers and retain the most informative ones. This step is crucial when training a machine learning model as it helps prevent overfitting and mitigates the risks associated with the “curse of dimensionality” [30]. To do so, we applied a Linear Discriminant Analysis (LDA) to the data, which successfully clustered the samples based on their predicted response to TMZ, as shown in Figure 3A. This result remarks that the genes identified through both DE analyses contain valuable information regarding treatment effectiveness. Finally, we selected the top 6 genes that had the most significant impact on the classification of samples responsiveness: BIRC3, CBLC, IL6, PTX3, SRD5A1, and TNFAIP3 (Figures 3B and 4).

### Biological function of the 6 candidate biomarkers:

According to the literature, the 6 selected genes are dysregulated in the tumor microenvironment and their overexpression is consistently associated to an unfavourable prognosis. Moreover, all the 6 selected genes have been associated with tumor severity [31-37], and 2 are involved in apoptosis resistance (BIRC3 and TNFAIP3). Additionally,

4 have already been recognized as key factors in brain tumors. Indeed, in GBM tumors high levels of BIRC3 and IL6 are associated with short-term survival [38, 39]; targeting TNFAIP3 leads to decreased growth and survival of brain tumor cells [40, 41] and PTX3 expression level is observed to correlate with the tumor grade and malignancy [36, 42]. In line with the literature, this panel of genes is overexpressed in the groups characterized by an unfavourable prognosis in both the GBM explants and TCGA cohorts (Figure 4). The consistency of our results with the literature provides further support for the reliability of the approach we have adopted. Crucially, 5 candidate biomarkers have been already identified as drug targets in cancer (BIRC3 [39], IL6 [31, 35], CBLC [43], TNFAIP3 [40, 41], and SRD5A1 [44]). This result highlights that the synergistic use of NAD(P)H-FLIM with omics techniques can unveil new treatment strategies based on targeting the genes that are involved in therapeutic response.

**Machine learning modelling:** We consequently exploited the selected set of 6 genes as a transcriptomic signature to develop a predictive model. Although machine learning is a powerful tool, its effectiveness is heavily affected by the number of available observations. Typically, a machine learning pipeline involves training the model on a train set, selecting the features and setting the model hyper-parameters to optimize the performance on the validation set, and testing the resulting best model on a test set. However, all these steps require splitting the entire dataset into subsets. This operation decreases the amount of data used to train the model, thus reducing the model capability to generalize to new data. Additionally, the performance of the model is evaluated on a small set of data, which can be highly variable and not representative of the full dataset. Due to these limitations, we have adopted a Leave-One-Subject-Out (LOO) procedure, described in section 1.6, instead of a typical one.

Additionally, to train and test the model, we used expression data obtained through a distinct technology (specifically microarray instead of RNA-seq) by patients not included in the previous analyses. By doing so, we avoided any potential information leakage and tested the technology independence of the identified genes. Hence, we designed a model based on the Cox regression and tested it both in the training set and in the test set, in order to verify the internal and the external validity. The hazard ratio (HR) computed between the two risk groups identified by the model (Low Risk and High Risk) proves the possibility to effectively stratify them. Indeed, when a new patient was tested, the model could assign a risk class using only the matched transcriptomic signature. Notably, the panel of 6 genes was obtained using RNA-seq data while the model was developed and validated using microarray-derived gene expression levels, implying the sequencing-technique independence of the candidate biomarkers.

**Comparing the predictive model performance with a clinical biomarker:** Finally, we wanted to evaluate the model's prediction performance using as benchmark a well-established clinical biomarker. With this aim, we compared the patient stratification performance of our model with the one obtained with the MGMT promoter methylation status [4, 45]. Hence, we first stratified patients respect to the MGMT status and, in accordance with the literature, we found that the methylated patients have a better prognosis than the unmethylated (Figure 6A). We then compared the stratification performance of MGMT status and our model. They reached strictly comparable results in terms of hazard ratio (Figure 5B and 6A), confirming the robustness of the NAPH-FLIM derived information.

Finally, we focused on Agilent patients with extreme PFI values (Agi Long-PFI/Agi Short-PFI), to assess the consistency of the model-based predictions and the MGMT methylation. The information provided by the two indicators matched for the 60%, thus did not fully overlap suggesting that the two variables could be synergistically used to generate a more accurate predictive model. We believe that the integration of this indicator with the 6 genes transcriptomic signature in a comprehensive model would remarkably improve patient stratification for GBM treatment.

Our results emphasize the valuable potential of the NAD(P)H-FLIM assay in assessing the drug response in GBM and show the possibility to use it in order to derive panels of predictive genes. The stratification performance achieved by the model, indeed, confirmed the possibility to use the FLIM derived information to make prognostic inferences.

We highlight that the bioinformatics approach we adopted can be tailored to validate other screening tests that cannot be directly investigated with a retrospective study.

## Author Contributions

AP and EC: study design; CMM and MM: design and execution of experiments; AP: design of analysis methodology; AP, EC and PA: curation and analysis of data; AP and EC: visualisation of results; AP and EC: writing of original paper; AP, CMM and PA: revision and editing of paper; AP and PA project supervision.

## Data Availability Statement

The data included into the GBM explants dataset is available upon request. The data included into the TCGA RNA-seq dataset is freely available at the Genomic Data Commons (GDC) Data Portal. The data included into the TCGA AgilentG4502A\_07\_2 Microarray dataset is freely available at UCSC (University of California Santa Cruz) Xena website. All the scripts to process and perform the analyses are available upon request.



## Funding

This work was funded by Fondazione Pisana per la Scienza ONLUS and Regione Toscana - Programma operativo regionale (POR) del Fondo sociale europeo (FSE) 2014-2020.

## Conflict of Interest

The authors declare no conflict of interest.

## References

- Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol* 131 (2016): 803-820.
- Ostrom QT, Gittleman H, Liao P, et al. CBTRUS Statistical Report: Primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. *Neuro Oncol* 19 (2017): v1-v88.
- Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med* 352 (2005): 987-996.
- Butler M, Pongor L, Su YT, et al. MGMT Status as a Clinical Biomarker in Glioblastoma. *Trends Cancer* 6 (2020): 380-391.
- Chen JR, Yao Y, Xu HZ, et al. Isocitrate Dehydrogenase (IDH)1/2 Mutations as Prognostic Markers in Patients With Glioblastomas. *Medicine (Baltimore)* 95 (2016): e2583.
- Szopa W, Burley TA, Kramer-Marek G, et al. Diagnostic and Therapeutic Biomarkers in Glioblastoma: Current Status and Future Perspectives. *Biomed Res Int* 2017 (2017): 8013575.
- Fathi Kazerooni A, Saxena S, Toorens E, et al. Clinical measures, radiomics, and genomics offer synergistic value in AI-based prediction of overall survival in patients with glioblastoma. *Sci Rep* 12 (2022): 8784.
- Senders JT, Staples P, Mehrtash A, et al. An Online Calculator for the Prediction of Survival in Glioblastoma Patients Using Classical Statistics and Machine Learning. *Neurosurgery* 86 (2020): E184-E192.
- Tewarie IA, Senders JT, Kremer S, et al. Survival prediction of glioblastoma patients-are we there yet? A systematic review of prognostic modeling for glioblastoma and its clinical potential. *Neurosurg Rev* 44 (2021): 2047-2057.
- Letai A. Functional precision cancer medicine-moving beyond pure genomics. *Nat Med* 23 (2017): 1028-1035.
- Letai A. Functional Precision Medicine: Putting Drugs on Patient Cancer Cells and Seeing What Happens. *Cancer Discov* 12 (2022): 290-292.
- Letai A, Bhola P, Welm AL. Functional precision oncology: Testing tumors with drugs to identify vulnerabilities and novel combinations. *Cancer Cell* 40 (2022): 26-35.
- Murphy SA. Optimal Dynamic Treatment Regimes. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 65 (2003): 331-355.
- Gillette AA, Babiarz CP, VanDommelen AR, et al. Autofluorescence Imaging of Treatment Response in Neuroendocrine Tumor Organoids. *Cancers (Basel)* 13 (2021).
- Lukina MM, Shimolina LE, Kiselev NM, et al. Interrogation of tumor metabolism in tissue samples ex vivo using fluorescence lifetime imaging of NAD(P)H. *Methods Appl Fluoresc* 8 (2019): 014002.
- Shirshin EA, Shirmanova MV, Gayer AV, et al. Label-free sensing of cells with fluorescence lifetime imaging: The quest for metabolic heterogeneity. *Proc Natl Acad Sci U S A* 119 (2022).
- Datta R, Heaster TM, Sharick JT, et al. Fluorescence lifetime imaging microscopy: fundamentals and advances in instrumentation, analysis, and applications. *J Biomed Opt* 25 (2020): 1-43.
- Kolenc OI, Quinn KP. Evaluating Cell Metabolism Through Autofluorescence Imaging of NAD(P)H and FAD. *Antioxid Redox Signal* 30 (2019): 875-889.
- Lukina MM, Dudenkova VV, Ignatova NI, et al. Metabolic cofactors NAD(P)H and FAD as potential indicators of cancer cell response to chemotherapy with paclitaxel. *Biochim Biophys Acta Gen Subj* 1862 (2018): 1693-1700.
- Morelli M, Lessi F, Barachini S, et al. Metabolic-imaging of human glioblastoma live tumors: A new precision-medicine approach to predict tumor treatment response early. *Front Oncol* 12 (2022): 969812.
- Morelli M, Franceschi S, Lessi F, et al. BIRC3: A Prognostic Predictor and Novel Therapeutic Target in TMZ-Resistant Glioblastoma Tumors. *bioRxiv* (2023): 554432.
- Stockslager MA, Malinowski S, Touat M, et al. Functional drug susceptibility testing using single-cell mass predicts treatment outcome in patient-derived cancer neurosphere models. *Cell Rep* 37 (2021): 109788.
- Abdullah KG, Bird CE, Buehler JD, et al. Establishment of patient-derived organoid models of lower-grade glioma. *Neuro Oncol* 24 (2022): 612-623.

24. Jacob F, Salinas RD, Zhang DY, et al. A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell* 180 (2020): 188-204 e22.
25. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2 (2020): lqaa078.
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15 (2014): 550.
27. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* 12 (2011): 2825-2830.
28. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society Series B (Methodological)* 34 (1972): 187-220.
29. Davidson-Pilon C. lifelines: survival analysis in Python. *Journal of Open Source Software* 4 (2019): 1317.
30. Berisha V, Krantsevich C, Hahn PR, et al. Digital medicine and the curse of dimensionality. *NPJ Digit Med* 4 (2021): 153.
31. Briukhovetska D, Dorr J, Endres S, et al. Interleukins in cancer: from biology to therapy. *Nat Rev Cancer* 21 (2021): 481-499.
32. Das K, Lorena PD, Ng LK, et al. Differential expression of steroid 5alpha-reductase isozymes and association with disease severity and angiogenic genes predict their biological role in prostate cancer. *Endocr Relat Cancer* 17 (2010): 757-770.
33. Frazzi R. BIRC3 and BIRC5: multi-faceted inhibitors in cancer. *Cell Biosci* 11 (2021): 8.
34. Hong SY, Kao YR, Lee TC, et al. Upregulation of E3 Ubiquitin Ligase CBLC Enhances EGFR Dysregulation and Signaling in Lung Adenocarcinoma. *Cancer Res* 78 (2018): 4984-4996.
35. Kumari N, Dwarakanath BS, Das A, et al. Role of interleukin-6 in cancer progression and therapeutic resistance. *Tumour Biol* 37 (2016): 11553-115572.
36. Locatelli M, Ferrero S, Martinelli Boneschi F, et al. The long pentraxin PTX3 as a correlate of cancer-related inflammation and prognosis of malignancy in gliomas. *J Neuroimmunol* 260 (2013): 99-106.
37. Verstrepen L, Verhelst K, van Loo G, et al. Expression, biological activities and mechanisms of action of A20 (TNFAIP3). *Biochem Pharmacol* 80 (2010): 2009-2020.
38. Tchirkov A, Khalil T, Chautard E, et al. Interleukin-6 gene amplification and shortened survival in glioblastoma patients. *Br J Cancer* 96 (2007): 474-476.
39. Wang D, Berglund A, Kenchappa RS, et al. BIRC3 is a novel driver of therapeutic resistance in Glioblastoma. *Sci Rep* 6 (2016): 21710.
40. Guo Q, Dong H, Liu X, et al. A20 is overexpressed in glioma cells and may serve as a potential therapeutic target. *Expert Opin Ther Targets* 13 (2009): 733-741.
41. Hjelmeland AB, Wu Q, Wickman S, et al. Targeting A20 decreases glioma stem cell survival and tumor growth. *PLoS Biol* 8 (2010): e1000319.
42. Tafani M, Di Vito M, Frati A, et al. Pro-inflammatory gene expression in solid glioblastoma microenvironment and in hypoxic stem cells from human glioblastoma. *J Neuroinflammation* 8 (2011): 32.
43. Kim B, Lee HJ, Choi HY, et al. Clinical validity of the lung cancer biomarkers identified by bioinformatics analysis of public expression data. *Cancer Res* 67 (2007): 7431-7438.
44. Sinreih M, Anko M, Zukunft S, et al. Important roles of the AKR1C2 and SRD5A1 enzymes in progesterone metabolism in endometrial cancer model cell lines. *Chem Biol Interact* 234 (2015): 297-308.
45. Szyłberg M, Sokal P, Sledzinska P, et al. MGMT Promoter Methylation as a Prognostic Factor in Primary Glioblastoma: A Single-Institution Observational Study. *Biomedicines* 10 (2022).