The Slovenian translational platform GlioBank for brain tumour research: identification of molecular signatures of glioblastoma progression

Metka Novak[1]*, Bernarda Majc[1], Marta Malavolta[2], Andrej Porčnik[3], Jernej Mlakar[4], Matjaž Hren[5], Anamarija Habič[1,6], Mateja Mlinar[1], Ivana Jovčevska[7], Neja Šamec[7], Alja Zottel[7], Marija Skoblar Vidmar[8], Tina Vipotnik Vesnaver[9], Andrej Zupan[4], Alenka Matjašič[4], Saša Trkov Bobnar[5], Dejan Georgiev[2], Aleksander Sadikov[2], Roman Bošnjak[3], Borut Prestor[3], Radovan Komel[7], Tamara Lah Turnšek[1], Barbara Breznik[1]*

[1]*Department of Genetic Toxicology and Cancer Biology, National Institute of Biology, 121 Večna pot, 1000, Ljubljana, Slovenia*

[2]*Faculty of Computer and Information Science, University of Ljubljana, 1000 Ljubljana, Slovenia*

[3]*Department of Neurosurgery, University Medical Centre Ljubljana, 1000 Ljubljana, Slovenia*

[4]*Institute of Pathology, Faculty of Medicine, University of Ljubljana, 1000 Ljubljana, Slovenia*

[5]*Biosistemika, d.o.o., 1000 Ljubljana, Slovenia*

[6]*Jožef Stefan International Postgraduate School, 1000 Ljubljana, Slovenia*

[7]*Medical Centre for Molecular Biology, Institute of Biochemistry and Molecular Genetics, Faculty of Medicine, University of Ljubljana, 1000 Ljubljana, Slovenia*

[8]*Institute of Oncology Ljubljana, 1000 Ljubljana, Slovenia*

[9]*Institute of Radiology, University Medical Centre Ljubljana, 1000 Ljubljana, Slovenia*

* Correspondence to: barbara.breznik@nib.si, metka.novak@nib.si

121 Vecna pot, 1000 Ljubljana, Slovenia, Phone + 386 59 232 881

Abstract

Background: Glioblastoma (GB) is one of the most lethal solid tumours in humans, with an average patient life expectancy of 15 months and 5-year survival rate of 5–10%. GB is still uncurable due to tumour heterogeneity and invasive nature as well as therapy-resistant cancer cells. Centralised biobanks with clinical data and corresponding biological material of GB patients facilitate the development of new treatment approaches and the search for clinically relevant biomarkers, with the goal of improving outcomes for GB patients. The aim of this study was firstly to establish a Sloveniantranslation platform, GlioBank, and secondly to demonstrate its utility through the identification of molecular signatures associated with GB progression and patient survival.

Methods: GlioBank contains tissue samples and corresponding tumour models as well as clinical data from patients diagnosed with glioma, with a focus on GB. Primary GB cells, glioblastoma stem cells (GSCs) and organoids have been established from fresh tumour biopsies. We performed expression analyses of genes associated with GB progression and bioinformatics analyses of available clinical and research data obtained from a subset of 91 GB patients. qPCR was performed to determine the expression of genes associated with therapy- resistance and cancer cell invasion, including markers of different GB subtypes, GSCs, epithelial-to-mesenchymal transition, and immunomodulation/chemokine signalling in tumour tissues and corresponding cellular models.

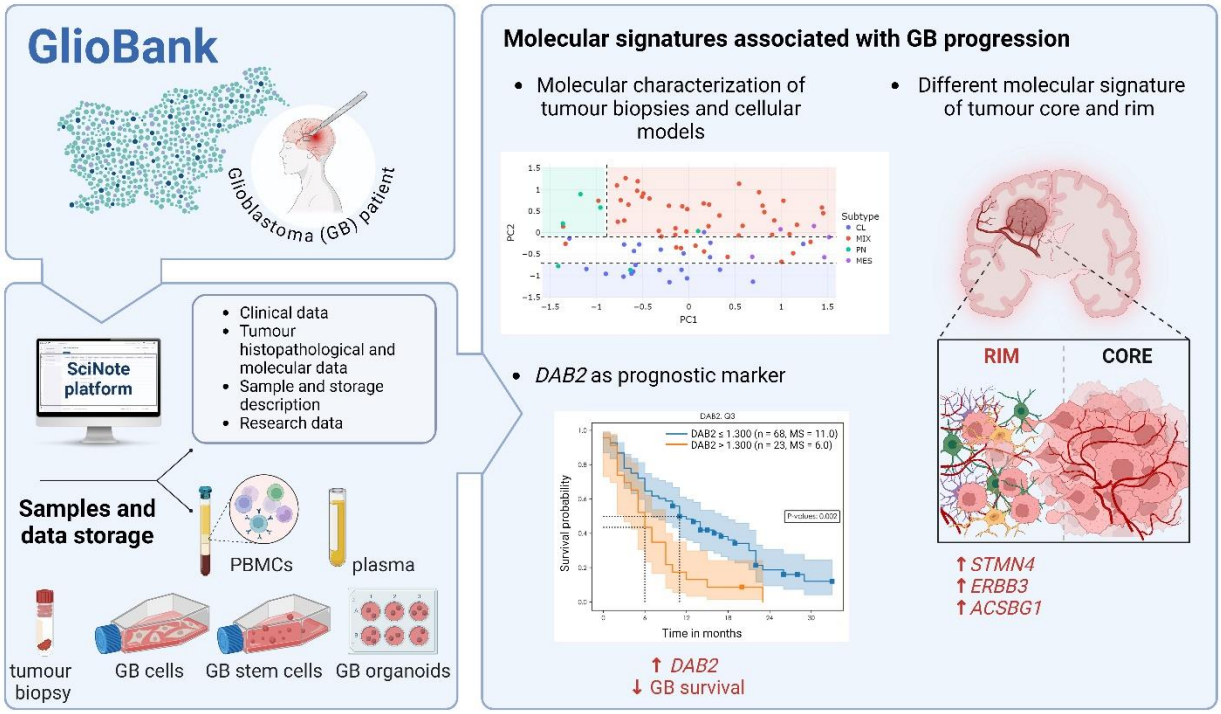
Results: GlioBank contains biological material and research, and clinical data collected in the SciNote electronic laboratory notebook. To date, more than 240 glioma tissue samples have been collected and stored in GlioBank, most of which are GB tissues (205) and were further processed to establish primary GB cells (n=64), GSCs (n=14), and GB organoids (n=17). Corresponding blood plasma (n=103) and peripheral blood mononuclear cells (n=101) are also stored. GB tumours were classified into four different subtypes that differed regarding patient survival; the mixed subtype exhibited the longest patient survival. High *DAB2, S100A4*, and *STAT3* expression was associated with poor overall patient survival, and *DAB2* was found to be an independent prognostic marker for GB survival. We analysed the molecular signatures between different tumour regions (core vs. rim). *STMN4, ERBB3*, and *ACSBG1* were upregulated in the tumour rim, suggesting that these genes are associated with the invasive nature of GB.

Conclusions: GlioBank is a centralised biobank that has been built by a multidisciplinary network with the aim to facilitate disease-oriented basic and clinical research. The advantages of GlioBank include the molecular characterization of GB based on targeted gene expression, the availability of diverse cellular

models (e.g. GB cells and organoids), and a large number of patient-matched tumour core and rim samples, all with accompanying molecular and clinical data. We report here for the first time an association between *DAB2* expression and low overall survival in GB patients, indicative of a prognostic value of *DAB2*.

Keywords: biobank, biomarker, glioblastoma, tumour models

Graphical abstract



Keypoints:

- A Slovenian glioblastoma biobank has been established, called GlioBank.
- *DAB2* is an independent prognostic marker for glioblastoma patient survival.
- The tumour core and rim exhibit different molecular signatures.

Importance of the study:

This study presents the establishment of the Slovenian translational platform GlioBank and its utility through the identification of molecular signatures associated with glioblastoma progression and patient survival. GlioBank relies on a multidisciplinary network and collects biological material and data from glioma patients with a focus on glioblastoma, an incurable cancer and the most aggressive primary brain tumour in adults. We report here for the first time the prognostic value of *DAB2*, as its expression was associated with low overall survival in GB patients. We analysed differences in molecular signatures between different tumour regions (core vs. rim) in 27 patient-matched samples. Our findings suggest that differences between these tumour regions must be considered in the search for therapeutic modalities targeting invading tumour cells. Our centralised GlioBank will facilitate clinical research and the development of new treatment approaches, with the ultimate goal of improving outcomes for GB patients.

Introduction

Glioblastoma (GB) is one of the most lethal solid tumours in humans, with an average patient life expectancy of 15 months[1] and 5-year survival rate of 5–10%[2]. It is considered a rare disease, with an incidence of 0.59–3.69 per 100,000 person-years[3–5]. However, its incidence increases with age and peaks between the ages of 75 and 84 years, with an incidence of 15.24 per 100,000 person-years[6,7]. The survival rate of GB patients is low, and thus new diagnostic tools and treatment approaches are urgently needed to prolong survival.

The current standard of care follows the Stupp protocol and consists of maximal safe surgical resection combined with radiotherapy and temozolomide chemotherapy, followed by adjuvant chemotherapy[8,9]. The poor response to therapy has been attributed to cellular and molecular heterogeneity, the activity of the DNA repair protein O6-methylguanine-DNA methyltransferase, and the existence of divergent cell subpopulations (e.g. GB stem cells (GSCs)) in GB[10,11]. GSCs represent a subset of cancer cells with stem cell-like features and driver mutations that survive radio- and chemotherapy[12]. GSCs are plastic and characterized by dynamic cell states that are affected by the microenvironment and therapeutic signals and therefore contribute to therapy resistance, angiogenesis, invasion, and recurrence[13,14].

GB exhibits extensive inter- and intra-tumour heterogeneity, including different cell populations and a continuum of transcriptional states, including proneural (PN), mesenchymal (MES), and classical (CL) subtypes, based on the well-established Behnan's classification system[15] of bulk tumours. Among these subtypes, the MES subtype is considered the most aggressive and resistant to therapeutic interventions, such as radiotherapy[14–17]. Moreover, the epithelial-to-mesenchymal transition (EMT) is a known cancer-related process that converts cancer cells into migratory and invasive cells[18]. The tumour microenvironment contains a plethora of different non-cancerous cells (e.g. endothelial and immune cells), which contributes to heterogeneity and creates an immunosuppressive environment, thereby limiting the efficiency of immunotherapeutic approaches. Many known immunoregulatory molecules, including CCL5/CCR3/5 and IL-6, are involved in shaping the GB microenvironment[14,19].

By contrast, less is known about the characteristics of invasive GB cells or a variety of other neuronal, glial, immune, and vascular cell types located at the edge (rim) of the GB[20]. An important feature of GB is its invasive nature, which renders standard therapies ineffective and leads to tumour recurrence. The microenvironments of the GB core and rim differ dramatically: the core contains hypoxic, necrotic, and angiogenic regions, whereas the rim contains adjacent non-tumour brain tissue[21] and invasive GB cells. As such, the GB rim region is an important topic for research.

Currently, GB is considered incurable, and progress in identifying clinically relevant biomarkers and effective therapeutic options remains limited[22]. However, with the rise and availability of omics and spatial biology technologies, new findings regarding the pathobiology and progression of GB are constantly emerging. Such studies require a large number of biological samples with associated clinical and molecular data. Data generated by large omics projects and smaller research groups should be stored in public repositories for future use. Biobanks with high-quality and well-annotated clinical samples are crucial for neuro-oncological research in the era of "precision" medicine. Biobanks of patient-derived tumour biopsies and cellular models, together with corresponding histopathological, molecular, and clinical data, have become an indispensable tool in translational cancer research[23].

In the first part, the purpose of this study was to establish an academic disease-specific repository, named GlioBank, that will help accelerate the field of glioma research with an emphasis on GB and the subsequent translation into clinical practice. GlioBank is a database of different clinical and molecular tumour characteristics, together with corresponding tumour tissue samples and cellular models from glioma patients. GlioBank thus provides a disease-oriented biobank in Slovenia that can facilitate the development of new treatment approaches and the search for new biomarkers, with the goal of improving outcomes for GB patients. The advantages of GlioBank include the molecular characterization of GB based on targeted gene expression, the availability of diverse cellular models (e.g. GB cells and organoids), and a large number of patient-matched tumour core and rim samples, all with accompanying molecular and clinical data. Conversely, its disadvantages include the lack of omics data that are envisioned for future studies focusing on spatial biology data of tumour tissues and cellular models to understand the GB microenvironment. In the second part of the study, the GlioBank datasets were used to determine the molecular characteristics of GB and identify biomarkers that are associated with poor GB survival.

Materials and Methods

1. Ethics statement

Approval by the National Medical Ethics Committee of the Republic of Slovenia was obtained (numbers 92/06/12, 0120-190/2018-4, 0120-190/2018-26, 0120-190/2018-32, and 0120-190/2018-35) for collecting and processing tumour tissue material, performing research on patient's material, and using the corresponding personal data. Patients or their authorized representatives signed the informed consent form in accordance with the Declaration of Helsinki.

2. The establishment of a physical GlioBank of biological specimens and a GlioBank database

2.1. Sample and clinical data collection

The patient inclusion criteria were as follows: (1) male and female patients aged ≥ 18 years old with preoperatively suspected and subsequently histologically and molecularly confirmed low- or high-grade glioma, (2) available representative tumour tissue from surgery or biopsy, and (3) either *de novo* or recurrent gliomas. Resected tumour tissues have been obtained from the Department of Neurosurgery at the University Medical Centre Ljubljana, Slovenia, from January 2012 - 2024. Since the samples are continuously collected the Gliobank is growing. Low-grade gliomas include pilocytic astrocytoma, astrocytoma, and oligodendroglioma of WHO grade 1 and 2; grade 3 gliomas include anaplastic astrocytomas and anaplastic oligodendrogliomas; and grade 4 gliomas include astrocytoma isocitrate dehydrogenase mutant (IDHmut) and GB IDH wild-type (IDHwt).

Immediately after surgical resection, the tumour tissues were transported to the Institute of Pathology at the Faculty of Medicine, University of Ljubljana, Slovenia (for histopathological and molecular examination) and National Institute of Biology, Slovenia, for storage and further analysis. Part of the tumour tissue was snap frozen and stored in liquid nitrogen containers. Fresh tumour tissue was used to establish patient-derived organoids, primary GB cells, and GSCs. Procedures for collecting and storing biological material were developed in accordance with best practice guidelines.

All research and clinical data, including neuroclinical, histopathological, and oncological data, were collected using the software SciNote, a cloud-based ELN software with lab inventory, compliance, and team management tools. Neuroclinical data were provided by the Department of Neurosurgery, Department of Neurology, and Institute of Radiology, all at the University Medical Centre Ljubljana, Slovenia. For tumours, confirmed diagnoses and histopathological and molecular data were provided by the Institute of Pathology at the Faculty of Medicine, University of Ljubljana, Slovenia. The oncological data were provided by the Institute of Oncology, Ljubljana, Slovenia. The description of specimen storage and research data were provided by the National Institute of Biology and Faculty of Medicine, Ljubljana, Slovenia. The obtained data are listed in Supplementary Table S 1. Tables with complete clinical and research data for each patient are not available publicly due to regulatory and ethical restrictions. The data is available on joint research projects with GlioBank consortium.

2.2. The development of the GlioBank database with the SciNote electronic laboratory notebook (ELN)

Our database was established with SciNote ELN, which is also available under an open-source licence. SciNote ELN provided several baseline functions for our database, including an established data model, user and access management, and baseline inventory capabilities.

We identified the data model and functional needs to store anonymized patient- and specimen-related data in the database through a gap analysis. The analysis was performed with GlioBank project stakeholders that were involved in the management, diagnostics, and treatment of GB patients as well as in related research in Slovenia. Identified data model and functional needs were converted to requirements and implemented in SciNote ELN and presented to the GlioBank project stakeholders. Data collected from stakeholders was cleaned up, mapped to the GlioBank database data model, and imported to the database.

2.3. Intraoperative collection of tumour core and rim samples

Samples of tumour core and rim regions were obtained during craniotomy and tumour resection according to Smith et al. and Porcnik et al.[24,25] and provided separately. Each sample was first taken from the core region of the tumour (named core), according to the enhancement area on the MRI guidance navigation system. The second sample was taken from the invasive edge or margin (named rim) and was defined by the 5-aminolevulinicacid-fluorescence-positive area beyond the enhancement, according to the MRI guidance navigation system. Patient-matched tumour core and rim samples from 27 GB patients were analysed.

2.4. The establishment of primary GB cells and GSCs

GB cells and GSCs from fresh GB tissues were established and cultured as described before[24,26–28]. For details, see the Supplementary Materials and Methods.

2.5. The establishment of GB organoids

GB organoids from resected patient tumour tissues were generated, passaged, and stored according to the protocols of Jacob et al.[29,30] and Majc et al.[31]. For details, see the Supplementary Materials and Methods.

2.6. Isolation of plasma and peripheral blood mononuclear cells

Peripheral blood samples (10 ml) were collected in K_2-EDTA tubes (Thermo Fischer Scientific, MA, USA) and transported to the National Institute of Biology, Slovenia. Plasma and peripheral blood mononuclear cells were isolated from whole blood using a density gradient medium (Lympholyte-H Cell Separation Media, Cedarlane, Canada). For details, see the Supplementary Materials and Methods.

3. Real-time quantitative PCR (RT-qPCR)

Gene expression was analysed according to previously described protocols[24,27,28]. Briefly, tissue and cell samples were snap frozen and stored in liquid nitrogen for further analysis. Total RNA from GB tissues was isolated using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, MD, USA) according to the manufacturer's instructions, and cDNA was prepared from 1 µg of total RNA using the High-Capacity cDNA Reverse Transcription Kit (Thermo Fischer Scientific, MA, USA). RT-qPCR was performed to determine the mRNA levels in our samples using FAM™/MGB probes (Thermo Fischer Scientific, MA; USA), the Fluidigm BioMark HD System RT-PCR (Fluidigm Corporation, San Francisco, CA, USA), and the 48.48 Dynamic Arrays IFC (Fluidigm Corporation, San Francisco, CA, USA). The RT-qPCR results were visualised and analysed using Biomark Data Collection software, Fluidigm RT-qPCR analysis software (both from Fluidigm Corporation), and quantGenius software[32]. Relative mRNA copy numbers were normalized to the housekeeping genes *HPRT1* and *GAPDH*. The assays and genes selected for analysis are listed in Supplementary Tables S 2 and S 3, respectively.

4. Data analyses

4.1. Classification of GB subtypes: data processing and algorithmic modelling

Based on the work of Porčnik et al.[24], the 90 GB samples were divided into the following subtypes: CL (n = 24 samples), MES (n = 5 samples), PN (n = 7 samples), and MIX (i.e. a mixture of all subtypes; n = 54 samples). Our modelling pipeline consisted of three main steps: i) capping gene expression data, ii) feature space reduction, and iii) model generation.

The gene expression level likely contains useful information, and thus we did not want to binarize it. However, expression data is not bounded upwards, and the outliers can skew this information. Therefore, capping was applied as per Equation 1, where x represents the expression level for a given gene. The 75th percentile (Q_3) was selected as an adequate value.

$$f(x) = \ f(x) = \begin{cases} \frac{x_i}{Q_3}, & x_i \leq Q_3 \\ 1, & x_i > Q_3 \end{cases} \quad (1)$$

We employed a decision tree (DT) algorithm to distinguish among the four tumour subtypes because of its clear interpretability. However, DT algorithms can overfit with complex data, and thus we restricted the depth to two levels, maintaining interpretability without losing crucial information. To address data complexity, we reduced dimensionality using principal component analysis, which condenses the dataset while retaining maximum information and minimizing noise. Therefore, we created a new dataset by selecting the two principal components as features from the initial 15, and we used entropy as the criteria for separating DTs. To evaluate DT performance, we used stratified 10-fold cross-validation.

4.2. Survival analysis and GB subtype comparisons

Overall survival was estimated using the Kaplan-Meier methodology, which calculated the duration in months from the date of diagnosis to either the date of death (event) or the last recorded follow-up (censored data). This analysis aimed to assess and compare survival outcomes between the four GB subtypes, namely CL, MIX, MES, and PN. The log-rank test was used for the comparisons. We calculated the median survival for each group, representing the point in time when the survivor function reached or dropped below 0.5. We also generated Kaplan-Meier curves to assess overall survival in relation to the expression of 40 genes. Patients were categorized into groups based on quartiles of gene expression in GB tissues, enabling the evaluation of the prognostic importance of these specific markers. Moreover, to explore survival determinants within our GB cohort, we employed a Cox proportional hazard model. The multivariate Cox analysis combined critical clinical variables (e.g. gender, age at surgery, and pre-operative and post-operative Karnofsky performance scale, the extent of surgical removal, postoperative treatment, and the three most significant gene expressions identified in the above-mentioned Kaplan-Meier analysis (*S100A4*, *STAT3*, and *DAB2*). Variables with a significance level (p-value) below 0.01, assessed using the chi-squared test, were considered relevant. All survival analyses, including Kaplan-Meier and Cox proportional hazard models, were performed using the Lifelines Python package.

4.3. Correlations between gene expression

We investigated the statistical correlations among the expression of different genes within distinct groups, including genes related to GSCs, EMT, and immune responses (Supplementary Table S 3). To quantify the correlations between these variables, we computed the Pearson correlation coefficient, which assesses the strength and direction of linear relationships between pairs of variables, alongside their respective p-values. To address multiple comparisons and reduce false positives, we corrected the obtained p-values using the Benjamini-Hochberg method. All statistical analyses were conducted in Python using Statsmodels and Scipy packages.

Results

1. Datasets and the establishment of GlioBank with SciNote ELN software

GlioBank contains biological material and research and clinical (neuroclinical, histopathological, and oncological) data collected in SciNote ELN, which is also available under an open-source licence (Figure 1)[33]. In this way, the data is accessible in one centralized location and can be expanded in the future with additional clinical, diagnostic, and research data. This was made possible by establishing highly efficient communication executed by contributors from academic and medical collaborative networks. GlioBank was established by the National Institute of Biology and Faculty of Medicine, University of Ljubljana, in collaboration with the Department of Neurosurgery, Department of Neurology, and Institute of Radiology, all at the University Medical Centre Ljubljana, and the Institute of Oncology, Ljubljana. Currently, frozen tumour tissues, plasma, peripheral blood mononuclear cells, GB cells, GSCs, and organoids from patients are stored in GlioBank (see Table 1 for details). The success rates for the establishment of GB cells, GSCs, and organoids are 49%, 39%, and 69%, respectively.

Collection of samples is in progress and currently more than 200 samples are collected (see Table 1) in GlioBank. For this research study a cohort of 101 GB patients was used for subsequent analysis since complete data and biological material are available. Supplementary Tables S 4–6 present a comprehensive summary of key clinical data, patient characteristics, and molecular features of GB patients included in our study. In total, 66% of the patients were men (67/101) with a mean age at the time of diagnosis of 62.8 years and median survival of 9 months after the diagnosis. Moreover, 60% of the patients underwent subtotal tumour resection (61/101), and 73% of the patients underwent complete treatment with irradiation and temozolomide chemotherapy (74/101). The median pre- and post-operative Karnofsky performance status was 70. Detailed molecular characterization identified mutations in the following genes: *EGFR* (in 27 out of 54 tested patients), *TP53* (in 18 out of 79 tested patients), *TERT* (in 8 out of 54 tested patients), *PDGFRA* (in 6 out of 54 tested patients), and *ATRX* (in 5 out of 60 tested patients). Furthermore, 18 (out of 32 tested patients) carried O6-methylguanine-DNA methyltransferase methylation. Other gene mutations were found in three or less patients. Chromosomal aberrations were detected in 10 patients out of 101 tested patients. MRI images are also available for 93 patients (Supplementary Table S 7).

To comprehensively investigate the determinants of survival within our GB cohort, we conducted a multivariate Cox analysis, considering several critical clinical variables: gender, age at the time of surgery,

pre-operative and post-operative Karnofsky performance status, the extent of surgical removal, and treatment after surgery (Supplementary Table S 8). Treatment after surgery showed the most substantial positive impact, with a hazard ratio of 2.35, highlighting a strong correlation between complete treatment (irradiation and temozolomide) and improved survival outcomes.

2. GB tumour subtyping, computational model for subtype categorisation and association with survival

The dataset of the GB tissues contains expression data for 15 selected genes, which were clustered into four GB subtypes based on Behnan's classification[15] (Supplementary Table S 9). The CL subtype exhibited high expression of *NF-KB*, *ACSBG1*, *S100A4*, and *KCNF1*. The MES subtype exhibited high expression of *DAB2*, *TGFB1*, *THBS1*, *COL1A2*, and *COL1A1*. The PN subtype exhibited high expression of *P2RX7*, *STMN4*, *SOX10*, *NOTCH*, *ERBB3*, and *OLIG2*. The MIX subtype exhibited a mixed expression of the genes of the other subtypes.

The frequency distribution of the GB subtypes within a subset of the patient cohort is presented in Supplementary Table S 10. Representative images of cellular models from tumour tissues of different molecular subtypes, PN, CL and MIX, respectively, are shown in Supplementary Figure S 2. GB cells from GB tissues of different molecular subtypes possess different morphology. That is not the case for GSCs and organoids, for which the morphology of cultures is similar between different GB subtypes. In general, we cannot detect specific morphological patterns of GB subtype based on morphology-based evaluation.

We aimed to develop machine learning tool to categorize GB subtypes based on gene expression signatures. To distinguish between the four subtypes, we employed a DT algorithm. Additionally, we addressed DT instability by reducing data dimensionality using principal component analysis. We were able to separate three out of four subtypes using only three rules. This can be observed in the right part of Figure 2a, which plots the first two principal components. The tree-like structure showing the decision paths for each subtype highlights a clear separation between subtypes MIX, CL, and PN. Unfortunately, the model did not separate the MES subtype, most likely because of the limited number of MES subtype samples.

The loadings of the first two principal components, which represent the contribution of each characteristic to determining the variance of the dataset, revealed that the loadings of the 15 genes in the first principal component are divided into three subtypes. In fact, all genes belonging to the CL subtype decrease the expression of this component, whereas genes belonging to the MES subtype (*THBS1*, *COL1A2*, and

COL1A2) have a positive relationship with PC1, increasing its expression (Figure 2b). The distributions of features in each subtype are presented in Supplementary Figures S 3–6.

We performed the Kaplan-Meier survival analysis to gain insight into the survival outcomes of GB patients among the four subtypes. We assessed the respective survival patterns of the four subtypes (Figure 2c) and compared the two most frequent subtypes, CL and MIX, which collectively represent a substantial portion of the patient cohort (Figure 2d). Our analysis revealed notable differences in survival curves between the CL and MIX subtypes. The CL subtype exhibited a median survival of 7 months, whereas the MIX subtype exhibited a median survival of 11 months, highlighting a substantial difference in prognosis between these two frequently observed molecular profiles.

3. High *S100A4*, *STAT3*, and *DAB2* expression correlated with poor overall survival of GB patients

Next, the expression of several genes associated with the GB subtypes, EMT, GSCs, and immunomodulation (Supplementary Tables S 11 and 12) was analysed in GB tissues and correlated with overall survival. GB patients were categorized into distinct groups based on the quartiles of gene expression in GB tissues, enabling the evaluation of the prognostic significance of these selected markers. Our analysis revealed significant disparities in overall survival among GB patients with varying *S100A4, STAT3,* and *DAB2* expression (Figure 3). Specifically, patients with high intra-tumour expression of these genes had more unfavourable overall survivals compared with those with lower expression. For instance, the Kaplan-Meier estimates for median patient survival were 12 and 6 months for individuals with low and high *S100A4* expression, respectively. Likewise, patients with low intra-tumour STAT3 and DAB2 expression had a median survival of 10.5 and 11 months, respectively, and patients with high intra-tumour STAT3 and DAB2 expression had a median survival of 7 and 6 months, respectively. Further details regarding the expression of the remaining selected genes and correlation with survival are provided in Supplementary Tables S 11 and 12. The expression of other genes did not correlate with survival. The multivariate Cox analysis (Figure 3d) revealed a positive impact of *DAB2* expression and survival variation, with a hazard ratio of 1.49. This indicates that higher *DAB2* expression is associated with lower survival. The other clinical factors included in this analysis exhibited the same trends as shown before.

S100A4, *DAB2*, and *STAT3* expression was then analysed in glioma tissues, non-tumour brain tissues, patient-derived GB cells, and non-cancerous astrocytes. All three genes were upregulated in GB tissues compared with non-tumour brain tissues. *STAT3* expression was higher in grade 3 astrocytoma compared with non-tumour brain tissue. *S100A4* expression was significantly higher in recurrent GB tissues compared with tissues of *de novo* GB, grade 2 and 3 astrocytomas, grade 2 oligodendrogliomas, and non-

tumour brain tissues. All three genes were expressed in primary GB cells. *DAB2* and *STAT3* were highly expressed in non-cancerous astrocytes, and *STAT3* was expressed in GSCs (Supplementary Figure S 7).

4. Correlations between the gene expression of markers related to the immune response, EMT, and GSCs

Next, we assessed whether EMT, immunomodulation, and mesenchymal and stem-like features are interconnected in GB tissues by analysing correlations between the expression of genes related to GSCs, EMT, and immunomodulation/chemokine signalling (Figure 4). Two groups of positively correlated genes were identified among GSC markers. One group contained *FUT4, ID1*, and *SOX2*, and the other group contained *CD9, TRIM28, PROM1, OLIG2, NOTCH*, and *ALYREF.* Positive correlations were also identified among EMT markers. The GSC markers *CD44, CD9,* and *NOTCH* positively correlated with EMT markers (Figure 4a). EMT markers positively correlated with *CCL5* and *CCR5* (Figure 4b). Moreover, the GSC markers *SOX2, CD44, FUT4,* and *ID1* positively correlated with the immune-related markers *CCL5* and *CCR3* (Figure 4c).

5. Differential expression of several genes in patient-matched tumour rim and core samples

To detect molecular differences between GB rim and core regions, we analysed mRNA expression of several genes associated with GB subtypes, GSCs, EMT, immunosuppression, and immunomodulation/cytokine signalling in 27 patient-matched tumour rim and core samples. The expression of *ID1, STMN4, P2RX7, ERBB3*, and *ACSBG1* was higher in the tumour rim compared with the tumour core. Conversely, the expression of *SNAI1, THBS1*, and *IL-6* was lower in the tumour rim compared with the tumour core (Figure 5). The mRNA expression of other analysed genes was not significantly different between the tumour core and rim (Supplementary Figure S 8).

To compare our results obtained with GlioBank, the Ivy Glioblastoma Atlas Project (IvyGAP) cohort[34] with 10 GB patients was analysed. IvyGAP, which includes gene expression data from morphologically distinct regions within tumours, provided similar results. The expression of *STMN4, ERBB3*, and *ACSBG1* was higher in the leading edge of tumours compared with the core of tumours, which includes the cellular tumour, perinecrotic zone, pseudopalisading cells around necrosis, hyperplastic blood vessels, and the microvascular area. The expression of *SNAI1, THBS1*, and *IL-6* was lower in the leading edge of tumours compared with the core of tumours. In the IvyGAP cohort[34], *P2RX7* and *ID1* expression differed, as it was lower in the leading edge of tumours compared with the core of tumours (Supplementary Figure S 9). These findings indicate that the PN subtype genes *ERBB3* and *STMN4* and the CL subtype gene *ACSBG1* are enriched in the invasive tumour area.

Discussion

Understanding the genetic, molecular, and metabolic characteristics of GB is essential for optimizing precision medicine. Such a comprehensive understanding of the disease shall enable better patient stratification and an effective search for new therapeutic options. Biobanks provide access to high-quality biological samples. However, the availability of samples and associated data can be limited due to the fragmentation of data, as is the case in the field of neuro-oncology in Slovenia.

In the first part of the study, the GlioBank was established to enable centralised research and active collaboration with key biomedical research stakeholders, e.g. research institutes, hospitals, universities, and companies. Our goals are to enhance research capabilities, promote multidisciplinary studies, and develop targeted therapies for personalised medicine, ultimately benefiting GB patients and reducing their burden on the healthcare system[35]. GlioBank incorporates multiple data sources, including clinical patient data and research data, to avoid data fragmentation. It has become part of the international biobanking consortium The European Research Infrastructure for Biobanking and Biomolecular Resources in Health and Life Sciences[36], which adheres to strict guidelines on ethical, legal, social, and research framework issues[37]. Thus, it is a valuable resource of high-quality patient samples and corresponding cellular models with associated clinical and molecular data.

To date, more than 200 glioma tissue samples have been collected and stored in GlioBank, most of which were further processed to establish primary GB cells, GSCs, and GB organoids with success rates of 49%, 34% and 69%, respectively. The success rate of differentiated GB cells is consistent with previous studies, although slightly higher compared to our results. For example, one study reported a success rate of around 60% for the establishment of differentiated GB cells from fresh surgical specimens[38]. The establishment and cultivation of GSCs is generally more difficult and less efficient compared to differentiated GB cells due to their specific growth requirements. Nevertheless, the success of GSC establishment was described in a study in which the authors correlated neurosphere formation with clinical outcome in malignant gliomas. Neurosphere formation was successful in 31 of the 43 GB samples examined, which corresponds to a success rate of 70[39]. In comparison, Jacob et al.[29] reported a success rate of 91.4% for GBOs formation. Our success rate of 69% is lower, but still remarkable as it underlines the reproducibility of our GBO establishment methods. Corresponding blood plasma has been collected to enable the analysis of potential biomarkers in liquid biopsies for the development of non-invasive diagnostic tools. Corresponding peripheral blood mononuclear cells have been collected to establish immunocompetent cellular models *in vitro*. Peripheral blood mononuclear cells can be used as a source of patient-matched

immune cells (e.g. macrophages, T cells, and natural killer cells) for co-cultures with GB cells or organoids of the same patient. Such *in vitro* models with immune cells enable investigations of cellular interactions and the role of the immune cell compartment in GB progression[26]. MRI images have been collected to study their association with biological and molecular features of tumours, as radiomics with artificial intelligence approaches may be implemented for GB stratification and more comprehensive diagnoses.

To demonstrate the utility of GlioBank for basic and translational research, we performed a set of further studies in the second part of the study. We molecularly characterised the samples, analysed molecular signatures of GB tissues, and performed correlation analyses between molecular and clinical data. We focused on potential new targets and biomarkers that are already used in clinical settings and/or are associated with poor GB therapeutic responses, e.g. genes related to stem-like features, EMT, mesenchymal features, and immunomodulation[14,16,19,24,28,40–44]. Overall, genes related to GSCs and EMT are expressed in GB tissues, and their expression is correlated, implying interconnected processes (e.g. EMT, mesenchymal, and stem-like features) in GB pathobiology[41]. We found correlations between poor overall GB patient survival and *STAT3* and *S100A4* expression. Accordingly, these genes were related to GB progression and poor survival in other cohorts and studies[45,46]. Correlations between poor survival of glioma patients and high *STAT3* and *S100A4* expression in tumour tissues have been confirmed by the TCGA database – GlioVis data portal[47] (Supplementary Figures S 10 and 11).

S100A4 is a small calcium-binding protein involved in multiple biological processes in cancer and inflammatory diseases[42,45,46,48,49]. A recent study on methylated differentially expressed genes and associated signalling pathways in GB revealed that *S100A4* is one of the most overexpressed genes and is associated with poor overall survival in GB patients[50]. Increased S100A4 protein expression is associated with poor survival in glioma mouse models, and its expression in GB-associated T cells and macrophages plays a critical role in promoting immunosuppression and glioma growth[48]. S100A4 was also identified as a novel marker and critical regulator of GSCs[43].

STAT3 is a transcription factor that regulates diverse cellular processes, such as proliferation, differentiation, and apoptosis[51]. Similar to our survival analysis of GB tumour samples, increased *STAT3* expression correlates with poor prognostic outcomes of GB patients[49], which is associated with a transition of the molecular subtype to a more aggressive mesenchymal profile[42]. It has also been suggested that STAT3 regulates the growth and self-renewal of GSCs[52]. In our study, *STAT3* was expressed in patient-derived GSCs.

DAB2 is an adaptor protein in clathrin-mediated endocytosis with a role in regulating signalling pathways involved in homeostasis, cell positioning, and EMT. It acts as both a tumour promotor and tumour suppressor via regulation of EMT, Wnt/β-catenin, TGF-β and ERK/MAPK signalling, angiogenesis, metastasis, and polarisation and function of tumour-associated macrophages[53–56]. Our current study identified *DAB2* as an independent prognostic marker, which is in line with another study that found *DAB2* expression to be associated with shorter survival of cancer patients[56]. To the best of our knowledge, present study is the first study to show that *DAB2* is a prognostic marker for GB patients. The results were confirmed by the TCGA database – GlioVis data portal[47] (Supplementary Figure S 12). The exact role of *DAB2* in GB progression remains unknown.

The simplified Behnan's classification[15] of GB tumours that is based on the expression of 15 genes enabled the stratification of GB into four subtypes. Our findings demonstrate that the CL subtype is indicative of a worse prognosis than the MIX subtype. We decided to include Behnan's classification because it is more clinically accessible than those with large number of targets, such as Verhaak's[57] and Patel's[58]. To discriminate between GB subtypes an explainable model based on robust machine learning approach has been developed. This machine learning tool can based on gene expression signatures discriminate between 3 GB subtypes - MIX, CL, and PN. Unfortunately, the model does not separate the MES subtype, most likely because of the limited number of MES subtype samples used in study.

We also compared the GB subtype Kaplan-Meier survival data between the GlioBank and TCGA cohorts. As shown in the Supplementary Figure S 13, significant differences in median survival months are evident between the two datasets. In the GlioBank cohort, the median survival for the CL, MES, and PN subtypes is 7, 15, and 4 months, respectively. In contrast, in the TCGA dataset, these subtypes show median survival times of 12, 10, and 13 months, respectively. In particular, the PN subtype exhibits a much longer survival in TCGA (13 months) compared to GlioBank (4 months), and the MES subtype has a shorter survival in TCGA (10 months) compared to GlioBank (15 months). We think these survival differences between the two cohorts can be attributed to two factors. First, there is a notable difference in gene expression distributions between the two datasets due to the distinct methodologies used for the extraction of the expression and, consequently the division into subtypes. As a matter of fact, the TCGA dataset utilizes RNA-seq, while the GlioBank dataset is based on qPCR, which focuses on specific genes with more targeted sensitivity, and their behaviour is visible in the boxplots comparing gene expression levels between the two datasets. Second, the difference in sample sizes between the two cohorts also plays a significant role. In GlioBank, the PN subtype is represented by only 7 samples and the MES subtype is represented by 5

samples. By contrast, TCGA includes 45 PN and 51 MES samples. The limited sample size in GlioBank affects the statistical power of the survival estimates, where the small cohort size may exaggerate survival differences between the datasets. Nevertheless, a larger cohort of GB patients is needed to confirm our results and develop explainable computational model to discriminate between GB subtypes and determine the prognostic value of the subtypes.

Furthermore, GlioBank also contains patient-matched samples from *de novo* and recurrent tumours (which enable us to track changes in tumours over time) and from different tumour regions (the core and invasive edge, i.e. rim). One of the main causes of GB recurrence are invasive cells at the rim of the tumour that escape surgical removal and lead to tumour regrowth[17]. Despite its critical role in tumour recurrence, the invasive GB rim remains poorly understood, largely due to the lack of available patient material.

We showed that genes belonging to the PN subtype, such as *ERBB3, P2RX7*, and *STMN4*, were upregulated in the tumour rim compared with the tumour core. Furthermore, we observed higher expression of *THBS1,* a MES gene, and *ACSBG1*, a CL gene, in the tumour core. In addition, we found significant upregulation of *SNAI1* in the tumour core, which is an important regulator of EMT[18] and is associated with the MES subtype. Using transcriptional subtype markers, Jin et al.[59] demonstrated that PN genes were expressed in the peripheral region of patient tumour samples, whereas MES genes were expressed in the central core, which confirms our findings. In the IvyGAP cohort, which includes gene expression data from morphologically distinct regions within tumours[34], the PN genes *STMN4* and *ERBB3* are enriched in the rim, whereas the MES gene *THBS1* is enriched in the core. This suggests a trend toward a MES signature in the tumour core and a PN signature in the rim. Altogether, these results indicate that the PN subtype genes *ERBB3* and *STMN4* and the CL subtype gene *ACSBG1* are enriched in the invasive tumour area and are related to the invasive nature of GB. In the end, these findings reveal differences between the GB rim and core regions that should be considered in the search for new therapeutic modalities targeting GB cells at the edge of tumours.

Future directions are envisioned for the molecular characterization of patient-derived cellular models and their comparisons with corresponding tumour tissues. We aim to compare tissues, GB cells, and organoids as well as GB cells and organoids from different tumour regions (core vs. rim).

Taken together, we expect that the translational platform GlioBank will foster further advances in cancer biology by enabling the determination of molecular signatures of aggressive and therapy-resistant tumours and the identification of new drivers of GB progression. Our findings suggest that DAB2, an adaptor protein in clathrin-mediated endocytosis, is a potential prognostic factor for GB, and further experiments are ongoing to determine its role in GB progression.

Funding

This work was supported by the Slovenian Research Agency (programme and research grants P1-0245, J3-4504, J3-2526, NC-0023, Z3-2649, Z3-1869, Z3-4510, P1-0390, and P2-0209, and young researcher grants 10040137 to BM and 10040147 to AH), European Program of Cross-Border Cooperation for Slovenia-Italy Interreg TRANS-GLIOMA, and European Union's Horizon project Twinning for excellence to strategically advance research in carcinogenesis and cancer (CutCancer; 101079113). MM has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie Innovative Training Network 2020 (Grant Agreement N° 956394; PARENT project).

Acknowledgments

We thank dr. Eva Lasic for editing and reviewing a draft of this manuscript.

Conflict of Interest

None declared.

Authorship

MN and BB designed the study. MN, BM, AP, JM, AH, MM, IJ, NŠ, AZ, AZ, AM, and BB designed and performed experiments and analysed and interpreted data. MH and STB set up SciNote and the database. MN, MM, AS, BM, and BB analysed and interpreted the data. AP, JM, MSV, TVV, DG, RB, and BP provided clinical data and samples and analysed the data. RK and TLT conceived the project and interpreted the data. MN, BM, MM, and BB wrote the first draft of the manuscript. All authors reviewed and approved the submitted manuscript.

Data Availability

All data generated or analysed during this study are included in this published paper and its Supplementary File. The datasets generated and analysed during the current study are also available from the corresponding author upon reasonable request.

References

1. Klopfenstein Q, Truntzer C, Vincent J, Ghiringhelli F. Cell lines and immune classification of glioblastoma define patient's prognosis. *Br J Cancer*. 2019;120(8):806-814. doi:10.1038/s41416-019-0404-y

2. Li Q, Aishwarya S, Li JP, Pan DX, Shi JP. Gene Expression Profiling of Glioblastoma to Recognize Potential Biomarker Candidates. *Front Genet*. 2022;13. doi:10.3389/FGENE.2022.832742

3. Tan AC, Ashley DM, López GY, Malinzak M, Friedman HS, Khasraw M. Management of glioblastoma: State of the art and future directions. *CA Cancer J Clin*. 2020;70(4):299-312. doi:10.3322/CAAC.21613

4. Delgado-Martín B, Medina MÁ. Advances in the Knowledge of the Molecular Biology of Glioblastoma and Its Impact in Patient Diagnosis, Stratification, and Treatment. *Adv Sci*. 2020;7(9):1902971. doi:10.1002/ADVS.201902971

5. Ohgaki H, Kleihues P. Genetic Pathways to Primary and Secondary Glioblastoma. *Am J Pathol*. 2007;170(5):1445-1453. doi:10.2353/AJPATH.2007.070011

6. Wu W, Klockow JL, Zhang M, et al. Glioblastoma multiforme (GBM): An overview of current therapies and mechanisms of resistance. *Pharmacol Res*. 2021;171:105780. doi:10.1016/J.PHRS.2021.105780

7. Alexander BM, Cloughesy TF. Adult glioblastoma. *J Clin Oncol*. 2017;35(21):2402-2409. doi:10.1200/JCO.2017.73.0119

8. Stupp R, Hegi ME, Mason WP, et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol*. 2009;10(5):459-466. doi:10.1016/S1470-2045(09)70025-7

9. Stupp R, Mason WP, Van Den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352(10):987-996. doi:10.1056/NEJMoa043330

10. Qazi MA, Vora P, Venugopal C, et al. Intratumoral heterogeneity: pathways to treatment resistance and relapse in human glioblastoma. *Ann Oncol Off J Eur Soc Med Oncol*.

2017;28(7):1448-1456. doi:10.1093/ANNONC/MDX169

11. Goenka A, Tiek D, Song X, Huang T, Hu B, Cheng SY. The Many Facets of Therapy Resistance and Tumor Recurrence in Glioblastoma. *Cells*. 2021;10(3):1-21. doi:10.3390/CELLS10030484

12. Ou A, Alfred Yung WK, Majd N. Molecular Mechanisms of Treatment Resistance in Glioblastoma. *Int J Mol Sci 2021, Vol 22, Page 351*. 2020;22(1):351. doi:10.3390/IJMS22010351

13. D'Alessio A, Proietti G, Sica G, Scicchitano BM. Pathological and Molecular Features of Glioblastoma and Its Peritumoral Tissue. *Cancers (Basel)*. 2019;11(4). doi:10.3390/CANCERS11040469

14. Gimple RC, Yang K, Halbert ME, Agnihotri S, Rich JN. Brain cancer stem cells: resilience through adaptive plasticity and hierarchical heterogeneity. *Nat Rev Cancer*. Published online 2022. doi:10.1038/s41568-022-00486-x

15. Behnan J, Stangeland B, Hosainey SAM, et al. Differential propagation of stroma and cancer stem cells dictates tumorigenesis and multipotency. *Oncogene*. 2017;36(4):570-584. doi:10.1038/ONC.2016.230

16. Bhat KPL, Balasubramaniyan V, Vaillant B, et al. Mesenchymal differentiation mediated by NF-κB promotes radiation resistance in glioblastoma. *Cancer Cell*. 2013;24(3):331-346. doi:10.1016/J.CCR.2013.08.001

17. Minata M, Audia A, Shi J, et al. Phenotypic Plasticity of Invasive Edge Glioma Stem-like Cells in Response to Ionizing Radiation. *Cell Rep*. 2019;26(7):1893-1905.e7. doi:10.1016/j.celrep.2019.01.076

18. Majc B, Sever T, Zarić M, Breznik B, Turk B, Lah TT. Epithelial-to-mesenchymal transition as the driver of changing carcinoma and glioblastoma microenvironment. *Biochim Biophys Acta - Mol Cell Res*. 2020;1867(10):118782. doi:10.1016/J.BBAMCR.2020.118782

19. Novak M, Krajnc MK, Hrastar B, et al. CCR5-Mediated Signaling Is Involved in Invasion of Glioblastoma Cells in Its Microenvironment. *Int J Mol Sci*. 2020;21(12):1-20. doi:10.3390/IJMS21124199

20. Darmanis S, Sloan SA, Croote D, et al. Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Rep*. 2017;21(5):1399-1410.

doi:10.1016/j.celrep.2017.10.030

21. Garcia-Diaz C, Pöysti A, Mereu E, et al. Glioblastoma cell fate is differentially regulated by the microenvironments of the tumor bulk and infiltrative margin. *Cell Rep*. 2023;42(5). doi:10.1016/j.celrep.2023.112472

22. Bagley SJ, Kothari S, Rahman R, et al. Glioblastoma Clinical Trials: Current Landscape and Opportunities for Improvement. *Clin Cancer Res*. 2022;28(4):594-602. doi:10.1158/1078-0432.CCR-21-2750

23. Gedye C, Sachchithananthan M, Leonard R, et al. Driving innovation through collaboration: development of clinical annotation datasets for brain cancer biobanking. *Neuro-oncology Pract*. 2020;7(1):31-37. doi:10.1093/NOP/NPZ036

24. Porčnik A, Novak M, Breznik B, et al. TRIM28 Selective Nanobody Reduces Glioblastoma Stem Cell Invasion. *Molecules*. 2021;26(17). doi:10.3390/MOLECULES26175141

25. Smith SJ, Diksin M, Chhaya S, Sairam S, Estevez-Cebrero MA, Rahman R. The Invasive Region of Glioblastoma Defined by 5ALA Guided Surgery Has an Altered Cancer Stem Cell Marker Profile Compared to Central Tumour. *Int J Mol Sci*. 2017;18(11). doi:10.3390/IJMS18112452

26. Breznik B, Ko MW, Tse C, et al. Infiltrating natural killer cells bind, lyse and increase chemotherapy efficacy in glioblastoma stem-like tumorospheres. *Commun Biol*. 2022;5(1). doi:10.1038/S42003-022-03402-Z

27. Majc B, Habič A, Novak M, et al. Upregulation of Cathepsin X in Glioblastoma: Interplay with γ-Enolase and the Effects of Selective Cathepsin X Inhibitors. *Int J Mol Sci*. 2022;23(3). doi:10.3390/IJMS23031784

28. Zottel A, Novak M, Šamec N, et al. Anti-Vimentin Nanobody Decreases Glioblastoma Cell Invasion In Vitro and In Vivo. *Cancers (Basel)*. 2023;15(3). doi:10.3390/CANCERS15030573

29. Jacob F, Salinas RD, Zhang DY, et al. A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell*. 2020;180(1):188-204.e22. doi:10.1016/j.cell.2019.11.036

30. Jacob F, Ming G li, Song H. Generation and biobanking of patient-derived glioblastoma organoids and their application in CAR T cell testing. *Nat Protoc*. 2020;15(12):4000-4033.

doi:10.1038/S41596-020-0402-9

31. Majc B. Patient-Derived Organoids Mimic Glioblastoma Resistance to Therapy and Increase DNA Damage Response. *iScience*. preprint. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4545256

32. Baebler Š, Svalina M, Petek M, et al. quantGenius: implementation of a decision support system for qPCR-based gene quantification. *BMC Bioinformatics*. 2017;18(1). doi:10.1186/S12859-017-1688-7

33. SciNote. Electronic Lab Notebook (ELN), accessed on November 21, 2023. https://www.scinote.net/

34. Puchalski RB, Shah N, Miller J, et al. An anatomic transcriptional atlas of human glioblastoma. Accessed July 21, 2023. www.swedish.org/services/neuroscience-institute

35. Hawkins AK. Biobanks: importance, implications and opportunities for genetic counselors. *J Genet Couns*. 2010;19(5):423-429. doi:10.1007/S10897-010-9305-1

36. Home - BBMRI-ERIC: Making New Treatments Possible. Accessed July 18, 2023. https://www.bbmri-eric.eu/

37. Peiró-Chova L, Ponce OB, Abril-Tormo C, Martínez-Santamaría J, López-Guerrero JA, Riegman PHJ. The Importance of Biobanks in Epigenetic Studies. *Epigenetic Biomarkers and Diagnostics*. Published online January 1, 2016:19-35. doi:10.1016/B978-0-12-801899-6.00002-4

38. Mullins CS, Schneider B, Stockhammer F, Krohn M, Classen CF, Linnebacher M. Establishment and Characterization of Primary Glioblastoma Cell Lines from Fresh and Frozen Material: A Detailed Comparison. *PLoS One*. 2013;8(8):e71070. doi:10.1371/JOURNAL.PONE.0071070

39. C. Jayakrishnan P, H. Venkat E, M. Ramachandran G, et al. In vitro neurosphere formation correlates with poor survival in glioma. *IUBMB Life*. 2019;71(2):244-253. doi:10.1002/IUB.1964

40. Zottel A, Jovčevska I, Šamec N, et al. Anti-vimentin, anti-TUFM, anti-NAP1L1 and anti-DPYSL2 nanobodies display cytotoxic effect and reduce glioblastoma cell migration. *Ther Adv Med Oncol*. 2020;12. doi:10.1177/1758835920915302

41. Singh A, Settleman J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene*. 2010;29(34):4741-4751. doi:10.1038/ONC.2010.215

42. Segerman A, Niklasson M, Haglund C, et al. Clonal Variation in Drug and Radiation Response among Glioma-Initiating Cells Is Linked to Proneural-Mesenchymal Transition. *Cell Rep*. 2016;17(11):2994-3009. doi:10.1016/J.CELREP.2016.11.056

43. Chow KH, Park HJ, George J, et al. S100A4 Is a Biomarker and Regulator of Glioma Stem Cells That Is Critical for Mesenchymal Transition in Glioblastoma. *Cancer Res*. 2017;77(19):5360-5373. doi:10.1158/0008-5472.CAN-17-1294

44. Bao S, Wu Q, McLendon RE, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nature*. 2006;444(7120):756-760. doi:10.1038/nature05236

45. Fei F, Qu J, Zhang M, Li Y, Zhang S. S100A4 in cancer progression and metastasis: A systematic review. *Oncotarget*. 2017;8(42):73219. doi:10.18632/ONCOTARGET.18016

46. Ambartsumian N, Klingelhöfer J, Grigorian M. The Multifaceted S100A4 Protein in Cancer and Inflammation. *Methods Mol Biol*. 2019;1929:339-365. doi:10.1007/978-1-4939-9030-6_22

47. Bowman RL, Wang Q, Carro A, Verhaak RGW, Squatrito M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro Oncol*. 2017;19(1):139-141. doi:10.1093/NEUONC/NOW247

48. Abdelfattah N, Kumar P, Wang C, et al. Single-cell analysis of human glioma and immune cells identifies S100A4 as an immunotherapy target. *Nat Commun*. 2022;13(1). doi:10.1038/s41467-022-28372-y

49. Tan MSY, Sandanaraj E, Chong YK, et al. A STAT3-based gene signature stratifies glioma patients for targeted therapy. *Nat Commun 2019 101*. 2019;10(1):1-15. doi:10.1038/s41467-019-11614-x

50. Tang Y, Qing C, Wang J, Zeng Z. DNA Methylation-based Diagnostic and Prognostic Biomarkers for Glioblastoma. doi:10.1177/0963689720933241

51. Fares J, Kanojia D, Cordero A, Ulasov I, Lesniak MS. Targeting the molecular mechanisms of glioma stem cell resistance to chemotherapy. *Glioblastoma Resist to Chemother Mol Mech Innov Reversal Strateg*. Published online January 1, 2021:587-634. doi:10.1016/B978-0-12-821567-8.00014-2

52. Sherry MM, Reeves A, Wu JK, Cochran BH. STAT3 is required for proliferation and maintenance of

multipotency in glioblastoma stem cells. *Stem Cells*. 2009;27(10):2383. doi:10.1002/STEM.185

53. Itami Y, Miyake M, Ohnishi S, et al. Disabled Homolog 2 (DAB2) Protein in Tumor Microenvironment Correlates with Aggressive Phenotype in Human Urothelial Carcinoma of the Bladder. *Diagnostics (Basel, Switzerland)*. 2020;10(1). doi:10.3390/DIAGNOSTICS10010054

54. Price ZK, Lokman NA, Yoshihara M, Kajiyama H, Oehler MK, Ricciardelli C. Disabled-2 (DAB2): A Key Regulator of Anti- and Pro-Tumorigenic Pathways. *Int J Mol Sci*. 2022;24(1). doi:10.3390/IJMS24010696

55. Adamson SE, Griffiths R, Moravec R, et al. Disabled homolog 2 controls macrophage phenotypic polarization and adipose tissue inflammation. *J Clin Invest*. 2016;126(4):1311-1322. doi:10.1172/JCI79590

56. Marigo I, Trovato R, Hofer F, et al. Disabled Homolog 2 Controls Prometastatic Activity of Tumor-Associated Macrophages. *Cancer Discov*. 2020;10(11):1758-1773. doi:10.1158/2159-8290.CD-20-0036

57. Verhaak RGW, Hoadley KA, Purdom E, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98-110. doi:10.1016/J.CCR.2009.12.020

58. Patel AP, Tirosh I, Trombetta JJ, et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. 2014;344(6190):1396-1401. doi:10.1126/SCIENCE.1254257

59. Jin X, Kim LJY, Wu Q, et al. Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nat Med*. 2017;23(11):1352-1361. doi:10.1038/nm.4415

Figure Captions

Figure 1. Schematic presentation of the GlioBank platform and representative images of cell cultures and tissues. A) GlioBank is a collection of biological samples and neurological and clinical data (provided by the Department of Neurosurgery, Department of Neurology, and Institute of Radiology, all at the University Medical Centre Ljubljana), histopathological and molecular data (provided by the Institute of Pathology, Faculty of Medicine, University of Ljubljana), and oncological data (provided by the Institute of Oncology, Ljubljana). Tumour tissues were collected at the time of surgery and stored in tissue banks. Part of the tissue was further processed to establish GB cells, GSCs, and GB organoids. GB: glioblastoma, GSC: glioblastoma stem cells. The scheme was created using BioRender.com. **B–E)** Representative images of primary GB cells **(B)**, GSCs **(C)**, GB organoids **(D)**, and formalin-fixed paraffin-embedded GB tissue stained with haematoxylin and eosin **(E)**. Scale bars: 100 µm (B, C, E) and 500 µm (D).

Figure 2. Frequency distribution of glioblastoma (GB) subtypes and principal component analysis. A) Visualisation of principal component analysis results, mapping the space of the first two principal components. The left part shows the decision tree structure highlighting the rules derived from the model. The colours in both schemes refer to the GB subtypes: CL (blue), MIX (red), PN (green), and MES (purple). **B)** Visualisation of the loadings for the first two principal components: PC1 and PC2. The values represent the offsets from the mean. Features are coloured according to the GB subtype. **C)** Kaplan-Meier survival curves with respect to four GB subtypes and **D)** the most frequent subtypes in the dataset: the CL and MIX subtypes. For each group, the number of patients (n) and median survival (MS) in months are reported. The log-rank test was used for statistical analysis of 90 samples.

Figure 3. High *S100A4*, *STAT3*, and *DAB2* gene expression is associated with worse prognosis of glioblastoma (GB) patients. Kaplan-Meier overall survival curves for *S100A4* **(A)**, *DAB2* **(B)**, and *STAT3* **(C)** are shown for the Q2 **(A)** and Q3 **(B, C)** quartile thresholds. The analysis was executed with 91 observations. Multivariate Cox regression analysis of factors contributing to overall survival of GB patients **(D)**. The analysis was executed with 86 observations.

Figure 4. Correlation heatmaps of gene expression of different groups in glioblastoma (GB) tissues: markers for GB stem cells and the epithelial-to-mesenchymal transition **(A)**, the epithelial-to-mesenchymal transition and immune response **(B)**, and GB stem cells and the immune response **(C)**. The

correlations highlighted with a star (*) are statistically significant after correction for multiple comparisons.

Figure 5. The expression of mRNA in patient-matched tumour rim and core samples. Relative mRNA expression of *ID1*, *STMN4*, *P2RX7*, *ERBB3*, *ACSBG1*, *SNAI1*, *THBS1*, and *IL-6* was determined using real-time quantitative PCR. Differences in gene expression in 27 patient-matched tumour core (pink dots) and rim (blue dots) samples are shown in the estimation plots. The mean differences are plotted on the right axis, and the error bars indicate 95% confidence intervals. The paired *t*-test was used for statistical analysis. Values of $p < 0.05$ are considered statistically significant.

Table

Table 1. Biological materials collected and stored in GlioBank. LGG: low-grade glioma, *IDHmut: isocitrate dehydrogenase mutant; GB: glioblastoma.*

Table 1. Biological materials collected and stored in GlioBank. LGG: low-grade glioma, *IDHmut: isocitrate dehydrogenase mutant; GB: glioblastoma.*

Biological material	***Quantity***
Tissues	
Glioma WHO grade 1 and 2 (LGG)	*19*
Glioma WHO grade 3	*16*
Gliosarcoma	*15*
Astrocytoma grade 4, IDHmut	*3*
GB Core	*167*
GB Rim	*38*
Blood samples	*103*
Plasma	*103*
Peripheral blood mononuclear cells	*101*
Tissue-derived cellular models	
GB cells	*64*
GB stem cells (GSCs)	*14*
Organoids	*17*

Figure 1

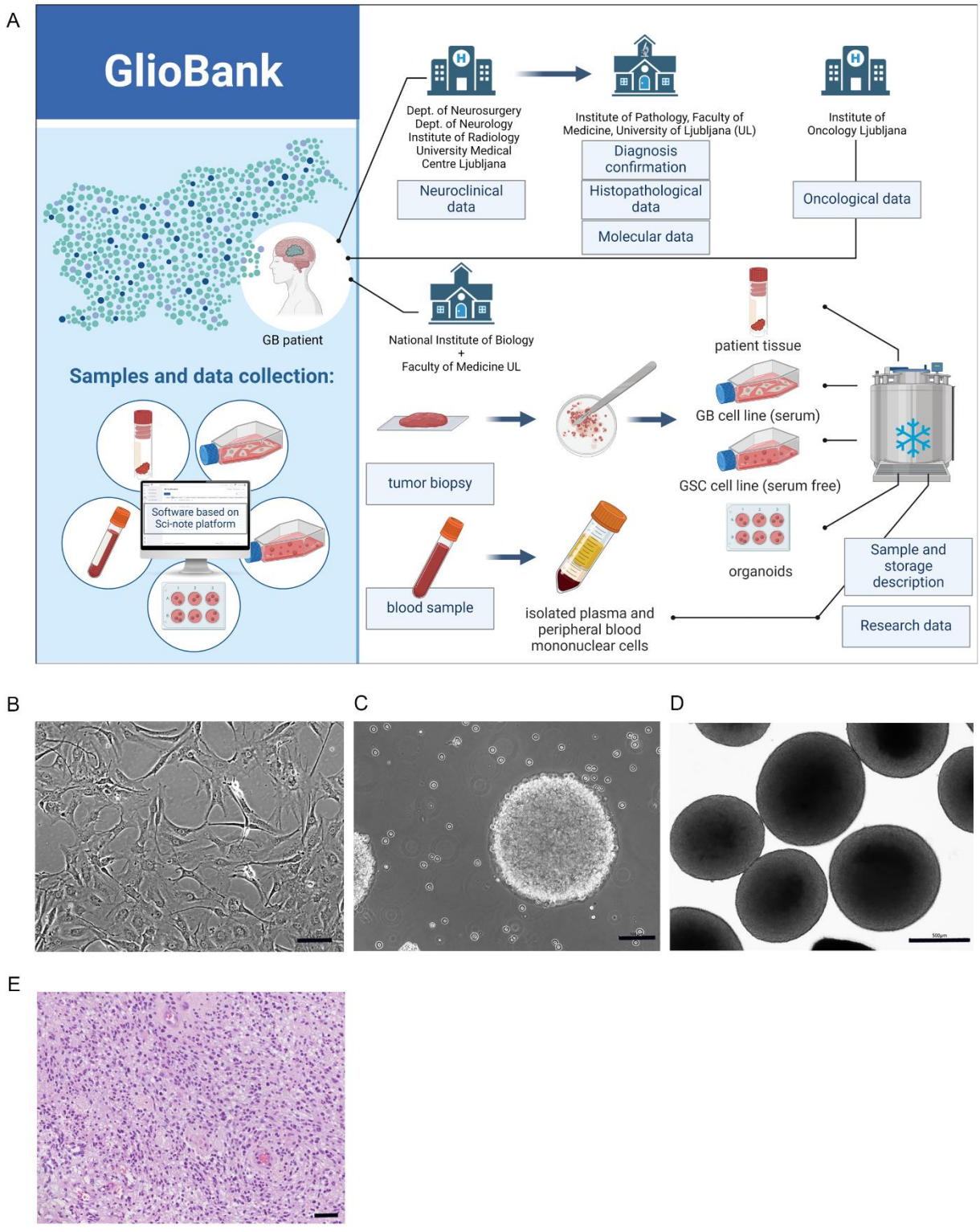


Figure 2

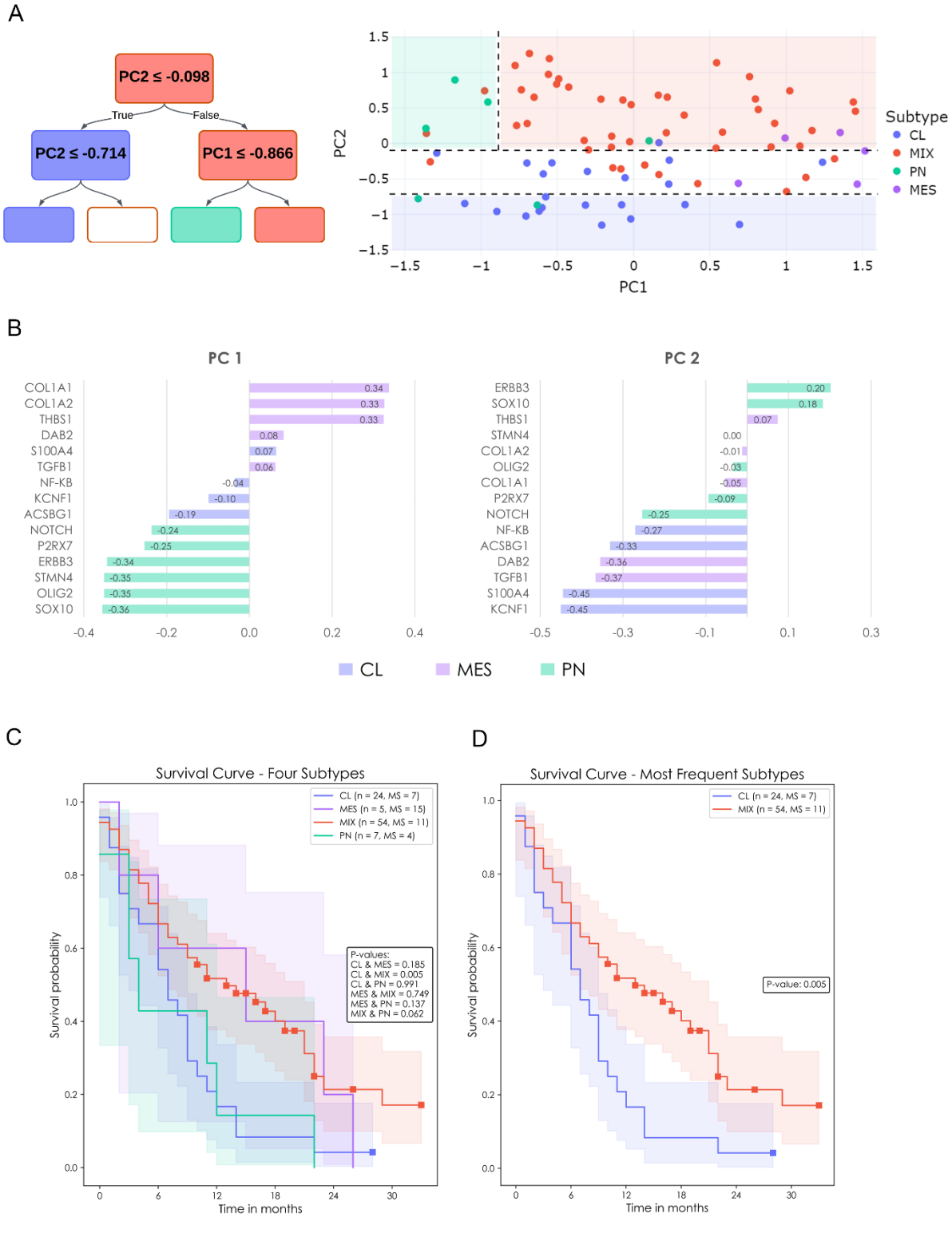
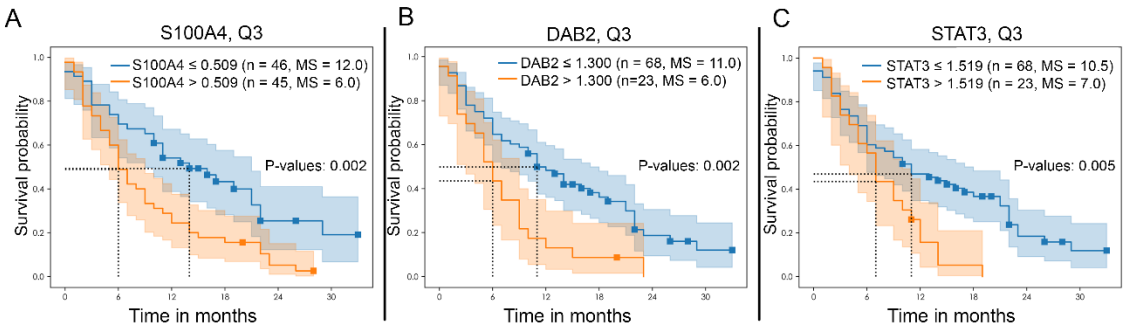


Figure 3



D

Covariates	***Full model HR (95% CI)***	***p-values***	***Category associated with more favourable prognosis***
Gender	2.01 (0.97 – 4.14)	0.06	Not significant
Age at time of surgery	1.02 (1 – 1.05)	0.08	Not significant
Pre-operative KPS	1.02 (0.97 – 1.09)	0.4	Not significant
Post-operative KPS	0.96 (0.91 – 1.00)	0.06	Not significant
Extent of surgical removal	1.31 (0.82 – 2.10)	0.25	Not significant
Treatment	2.20 (1.23 – 3.92)	**0.01**	**Complete treatment (IR+TMZ)**
S100A4	1.15 (0.88 – 1.5)	0.31	Not significant
DAB2	1.49 (1.16 – 1.91)	**<0.005**	**DAB2 expr. < 1.3**
STAT3	1.04 (0.74 – 1.47)	0.82	Not significant

Figure 4

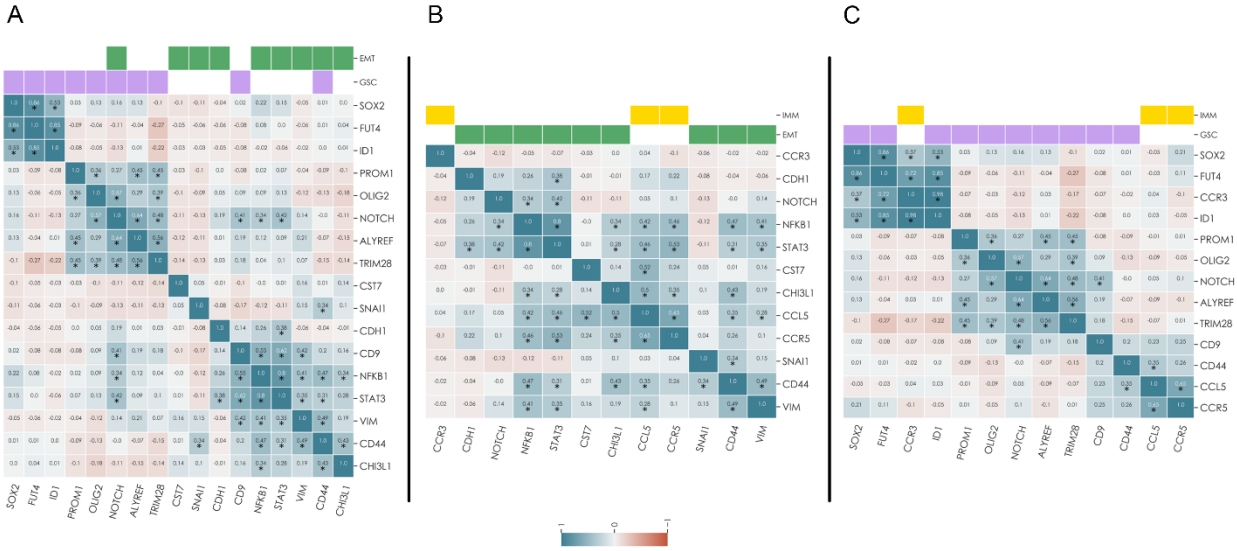


Figure 5