

# Analysis of complement system gene expression and outcome across the subtypes of glioma

Ojasvi Mudda<sup>1</sup>, Elisabeth Wilson<sup>2</sup>

<sup>1</sup> Evergreen Valley High School, San Jose, California

<sup>2</sup> Washington University School of Medicine in St. Louis, St Louis, Missouri

## SUMMARY

Glioma is a malignant and poorly understood cancer that occurs in the glial cells of the brain. Gliomas consist of two clinical subcategories: low-grade glioma (LGG) and glioblastoma (GBM). Based on gene expression profile analysis, gliomas are also divided into four transcriptional subtypes and two mutational subclasses based on *IDH* expression. These subtypes are associated with differing prognoses, histology, and gene profiles. The complement system, a branch of the innate immune system traditionally associated with inflammation and opsonization, has various pro-tumor effects, including immunosuppression, maintenance of glial stem cells, and hypoxic signaling. We analyzed the expression of complement system genes across the transcriptional and *IDH*-mutational subtypes of LGG and GBM to determine whether these genes are differentially expressed concerning mutation status and transcriptional subclasses. We performed differential gene expression analysis and analyzed the results for gene set enrichments and correlations with outcome status. The results showed that complement system genes are differentially expressed with varying outcomes across different glioma subtypes. Within the transcriptional subtypes, the complement system genes tended to be overexpressed in subtypes with poor response to treatment or increased tumor malignancy. These results suggest that dysregulation of the complement system may significantly contribute to the categorization of transcriptional subtypes and further, may play a role in treatment response and/or overall patient outcome, although more research is needed to confirm this. These findings could help elucidate the interplay between the immune system, gene expression, and glioma pathogenesis.

## INTRODUCTION

Glioma is a highly malignant and poorly understood cancer that develops in the glial cells of the brain. Patients with glioma typically have poor survival: a dismal 40% of patients survive a year post-diagnosis, and only 17% survive through a second year. Current treatment options involve surgery, radiation, and chemotherapy, which often involve many dangerous side effects (1).

There are two clinically defined subcategories of gliomas: low-grade glioma (LGG) and glioblastoma (GBM). LGG is defined as grade 2 or grade 3 glioma and is less aggressive

and metastatic than GBM, which is grade 4 (2,3). Based on integrative genomic analyses involving gene expression levels and copy number event analysis, gliomas have further been divided into four transcriptional subtypes (proneural, neural, mesenchymal, and classical) and two mutational subtypes based on *IDH 1/2* mutation status (wildtype and mutant) (4,5). Mutations in *IDH1* and *IDH2* are used to define distinct subtypes of gliomas and are associated with better prognosis. The wildtype subtype is associated with a worse prognosis compared to the mutant subtype (7). *IDH* mutations are associated with DNA hypermethylation, malignancy, and tumor cell multipotency; even though *IDH*-mutant is associated with a better prognosis, it is also associated with higher chances of recurrence and malignant transformation to a higher grade of glioma (8,9).

The transcriptional subtypes for glioblastoma and lower-grade glioma are characterized by mutations in *PDGFRA*, *IDH1*, *EGFR*, and *NF1* (4,6). The classical subtype is characterized by the best response to aggressive treatment and consists of the most common mutations seen in gliomas, including chromosome 7 amplifications and chromosome 10 deletions, *EGFR* amplification, and *INK4A/ARF* locus deletion (4). It is also characterized by a lack of additional mutations in *TP53*, *NF1*, *PDGFRA*, or *IDH1*, which is why it was used as the control level in gene expression analysis (4). The proneural subtype is associated with and has a similar mutational profile to secondary GBM, which involves *PDGFRA*, *IDH1*, and *TP53* and reduced efficacy of aggressive treatment (4). Despite being resistant to treatment, this subtype is also associated with greater survival as age increases (4). The mesenchymal subtype in glioma is associated with the expression of mesenchymal markers which indicate an epithelial-to-mesenchymal transition and a propensity for metastasis. This subtype is also characterized by the overexpression of genes in the tumor necrosis pathway (4). The neural subtype in glioma is associated with a gene expression pattern most similar to normal gene expression in the brain, mirroring the expression signature of differentiated cells (like astrocytes and oligodendrocytes) (4).

The complement system, a branch of the innate immune system, is a proteolytic, self-activating cascade of various proteins, which traditionally has been thought to have a role in fighting cancer. There are three major pathways of the complement system: the classical pathway (activated by antigen-antibody complexes), the alternative pathway (activated by permissive surfaces), and the lectin pathway (activated by mannose-binding lectins). These pathways converge to the formation of C3 convertase, which is cleaved into C3a, an anaphylatoxin, and C3b, an opsonin. Further activation results in the formation of the membrane attack complex (MAC), which initiates the process of cell lysis and

death. The complement system's role in immune surveillance and cytotoxicity initially pointed to an anti-tumor role of the complement system (9).

In addition to the well-characterized role of the complement system in innate immunity, recent studies have discovered potential pro-tumor roles of the complement system, which refuted its proposed anti-tumor role (10). In tumor cells, the complement system is involved in immunosuppression and is mutated in various cancers (11). Hypoxic gene signatures are more enriched in patients with complement system mutations, resulting in poorer prognoses (11). In gliomas, the complement system is involved in the maintenance of glial stem cells (GSCs) and promoting metastasis by influencing GSC-mediated angiogenesis, increasing tumor proliferation in the tumor microenvironment, and promoting the epithelial-mesenchymal transition (10,12). The complement system has also been shown to impact the efficacy of cancer treatment by nanoparticle drug-induced complement activation, which interferes with the drug's efficacy (13).

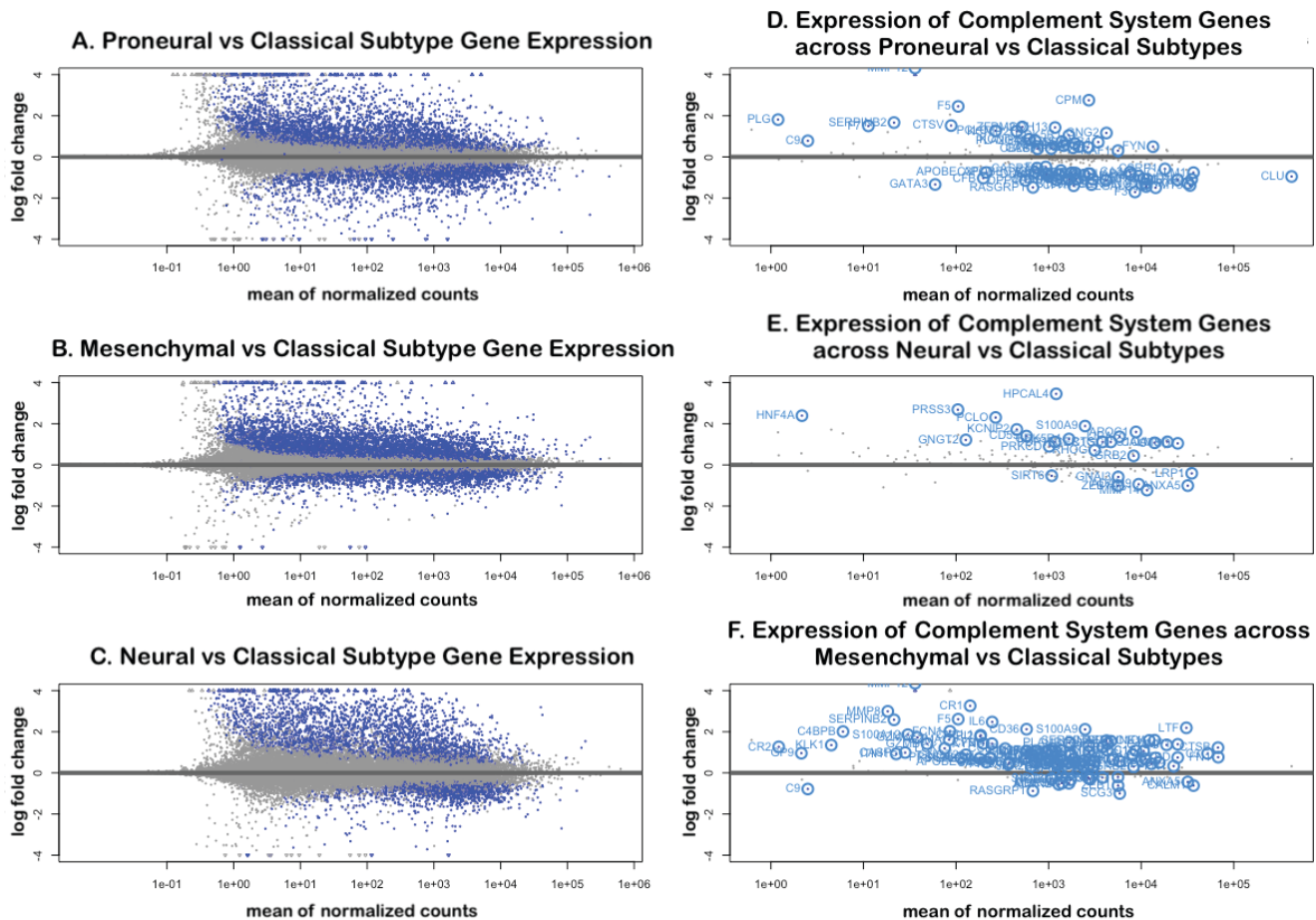
Here, we analyzed the expression of complement system genes across the transcriptional and *IDH*-mutational status subtypes of LGG and GBM to determine whether these genes are differentially expressed across mutational and/or transcriptional subclasses. We hypothesized that there would be a correlation between increased levels of complement

system gene expression and prognosis across the different glioma subtypes. All primary sample data analyzed were publicly available and accessed from The Cancer Genome Atlas (TCGA) database. Overall, the proneural subtype had the most genes associated with a worse prognosis, while the mesenchymal subtype had the most differentially expressed genes overall. Compared with *IDH*-mutant, *IDH*-wildtype had the most complement system genes overexpressed and associated with a worse prognosis.

## RESULTS

### DESeq2 Results Reveal Varying Levels of Complement System Gene Expression across Different Subtypes

Initial analysis of gene expression changes was supervised using outcome status to determine whether there were any overall trends in complement system gene expression across outcomes. The outcome status was based on the last contact the investigator had with the patient. Overall, genes were defined as upregulated if they had an LFC (Log2 Fold Change) greater than 0 and downregulated if they had an LFC less than 0. First, the data was cleaned by removing any NAs, resulting in a dataset with 122 dead patients and 27 alive patients. For GBM, we identified only a small number of differentially expressed genes ( $n = 8$ ) and only 1 gene (*CD36*) from the complement system gene set (adjusted  $p < 0.01$ ). For



**Figure 1: Differential gene expression across transcriptional subtypes.** MAplots compare normalized counts data (x-axis) with log fold change (LFC, y-axis) for all genes (A-C) or just complement system genes (D-F). Upregulated genes have LFC > 0 and downregulated genes have LFC < 0. Generated by the DESeq2 package. Points with an adjusted  $p$ -value  $\leq 0.01$  are marked blue.

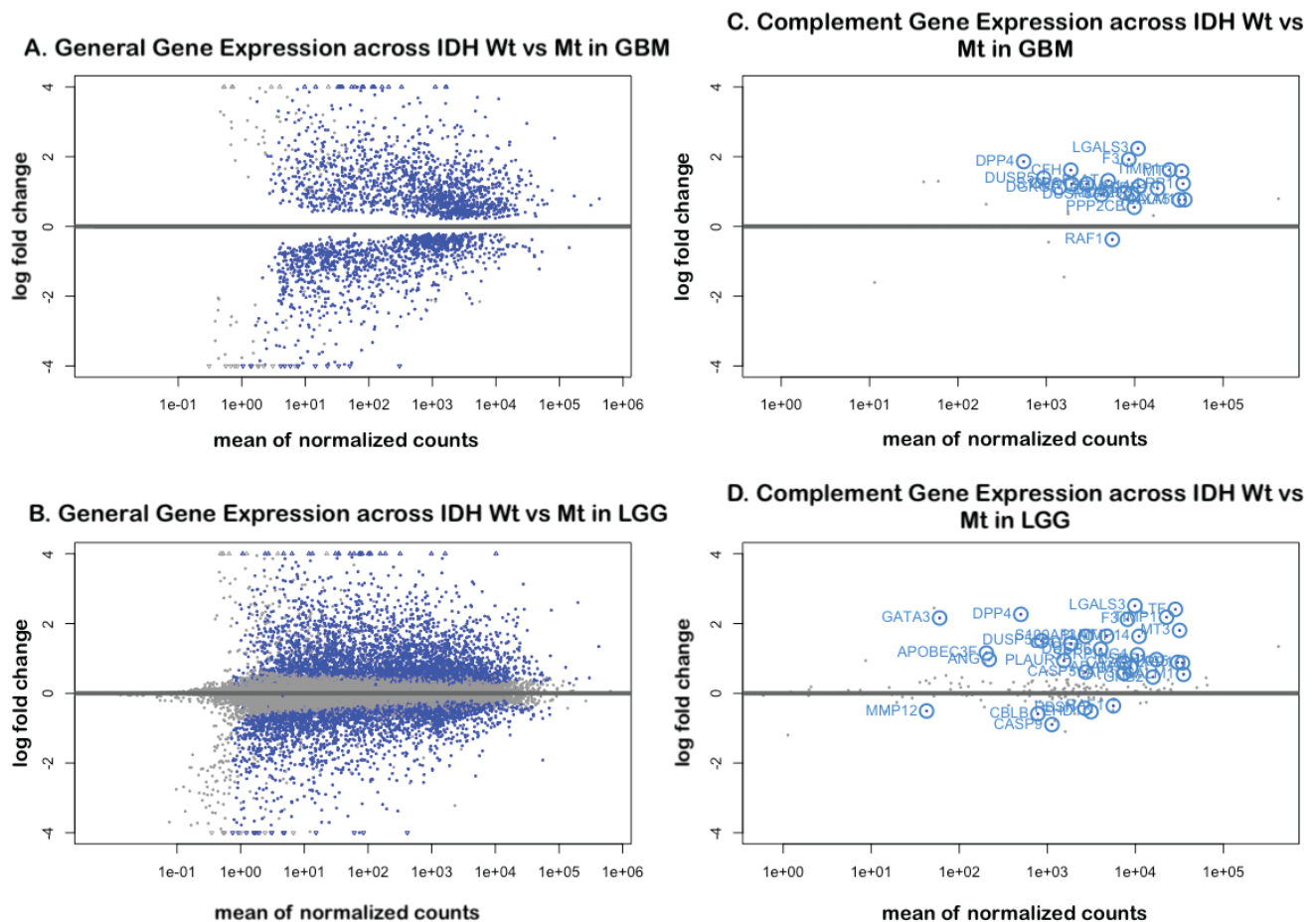
LGG (n=122 dead vs n=27 alive), there were 75 differentially expressed genes (n=8) and only 2 genes (*CD36* and *F5*) from the complement system gene set (adjusted  $p < 0.01$ ). Since complement gene expression could be dependent on other factors, like transcriptional subtype, comparing the overall complement gene expression by outcome might not reveal significant differences.

We next compared gene expression of the transcriptional subtypes within GBM (neural, mesenchymal, and proneural) with the expression profile of the classical subtype (n=40 dead vs n=8 alive) to determine whether transcriptional subtype could supervise complement system gene expression. Comparing the expression levels of proneural (n=15 dead vs n=3 alive) and classical subtypes revealed a total of 6331 differentially expressed genes (Figure 1A), including 67 complement system genes (adjusted  $p$ -value  $< 0.01$ ). Of the significantly differentially expressed complement system genes, 27 were upregulated and 40 were downregulated in proneural GBM compared to classical GBM (Figure 1D). Comparison of expression levels in neural (n=5 dead vs n=0 alive) and classical subtypes identified a total of 4029 differentially expressed genes (Figure 1C) with 27 complement system genes (adjusted  $p$ -value  $< 0.01$ ). Of the significantly differentially expressed complement system genes, 20 were upregulated and 7 were downregulated in

neural GBM compared to classical GBM (Figure 1E). When comparing the expression levels of mesenchymal (n=54 dead vs n=12 alive) and classical subtypes, there were a total of 8629 differentially expressed genes (Figure 1B) of which 107 were complement system genes. Of the significantly differentially expressed complement system genes, 92 were upregulated and 15 were downregulated in mesenchymal GBM compared to classical GBM (adjusted  $p$ -value  $< 0.01$ ) (Figure 1F).

We also analyzed gene expression differences between *IDH*-wildtype (*IDH*-wt) samples and *IDH*-mutant samples in GBM and LGG to determine whether *IDH* mutation status could supervise complement system gene expression. In GBM, there were a total of 1392 genes differentially expressed with 21 complement system genes (adjusted  $p$ -value  $< 0.01$ ). Of the significantly expressed complement system genes in GBM, 20 were upregulated and 1 was downregulated in *IDH*-mutant compared to *IDH*-wt samples (Figure 2A-B). In LGG, there were a total of 2520 genes differentially expressed with 33 complement system genes (adjusted  $p$ -value  $< 0.01$ ). Of the significantly expressed complement system genes in LGG, 27 were upregulated and 6 were downregulated in *IDH*-mutant compared to *IDH*-wt samples (Figure 2C-D).

Overall, we found that many genes in the complement system gene set were differentially expressed with respect



**Figure 2: Differential gene expression across *IDH* mutation status.** MAplots compare normalized counts data (x-axis) with log fold change (LFC, y-axis) for all genes (A-B) or just complement system genes (C-D). Upregulated genes have LFC  $> 0$  and downregulated genes have LFC  $< 0$ . Generated by the DESeq2 package. Points with an adjusted  $p$ -value  $\leq 0.01$  are marked blue.

to glioma subtypes and *IDH* mutation status. Overall, the mesenchymal subtype had the most upregulated complement system genes, followed by proneural and neural subtypes when compared with the classical subtype. Compared with GBM, LGG had more complement system genes overexpressed across *IDH*-mutant versus *IDH*-wt subtypes.

### Hierarchical Clustering of Complement System Gene Expression Supervises Glioma Subtypes

We performed hierarchical clustering of complement system gene expression in GBM, annotated with *IDH* mutation status, survival in months, vital status when the patient was in the last contact with the doctor, Karnofsky Performance Score, and transcriptome subtype to assess patterns of patient segregation from gene expression data from our DESeq studies. The Karnofsky Performance Score assesses disease progression by a scale that evaluates a patient's ability to perform day-to-day tasks on a scale of 0-100, with a higher score indicating a better prognosis (14). We determined where to cut the dendrogram tree by plotting the SSE (sum of squared error) across the number of cluster solutions and cutting where the SSE stopped decreasing exponentially. We observed clustering of patient samples sharing the same transcriptional subtype across three predominant clusters: a proneural cluster, which included all proneural cases, a classical cluster, which included 43% of all classical cases, and a mesenchymal cluster, which included 58% of all mesenchymal cases. The proneural cluster was composed of 48% proneural cases, 23% classical cases, 26% mesenchymal cases, and 3% neural cases. The classical cluster was composed of 100% classical cases. The mesenchymal cluster was composed of 76% mesenchymal cases and 24% classical cases. There was no discernible pattern for the other annotations included (Figure 3A).

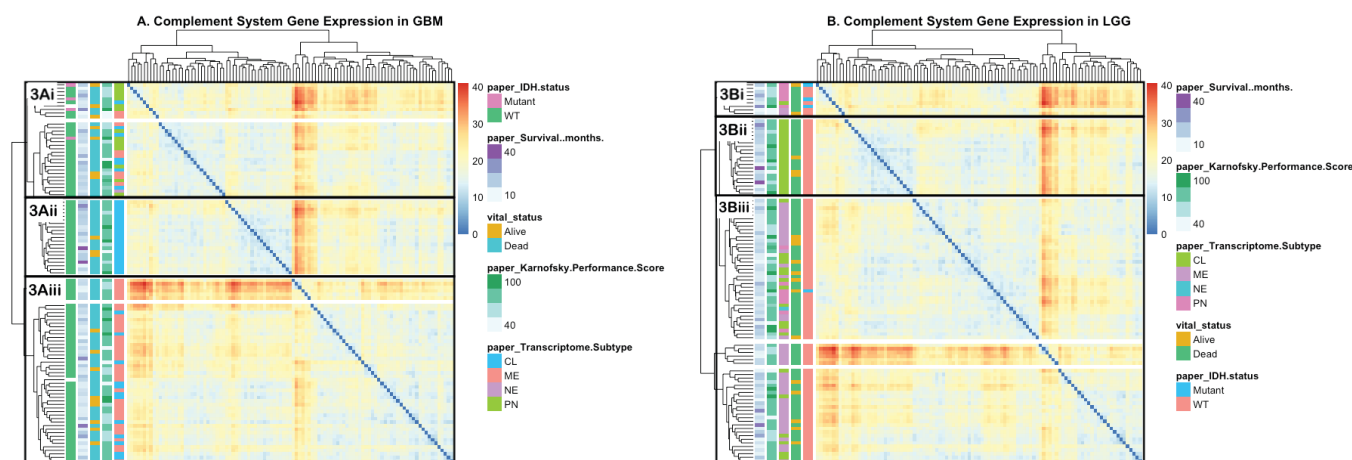
We also conducted hierarchical clustering of complement system gene expression in LGG, annotated with *IDH* mutation status, survival in months, vital status, Karnofsky Performance Score, and transcriptome subtype. Again, we determined the number of times the heatmap would be cut by plotting the SSE (sum of squared error) across the number of cluster solutions. Similar to the GBM cases, clustering was observed

to correlate with transcriptional subtypes defining 3 clusters: a proneural cluster (35% of all proneural cases), a classical cluster (43% of all classical cases), and a mesenchymal cluster (64% of all mesenchymal cases). The proneural cluster was composed of 67% proneural cases, 22% classical cases, and 11% mesenchymal cases. The classical cluster was composed of 100% classical cases. The mesenchymal cluster was composed of 66% mesenchymal cases, 25% classical cases, 11% proneural cases, and 1% neural cases. (Figure 3B). This data suggests that complement system gene expression can partially supervise transcriptional factor clustering.

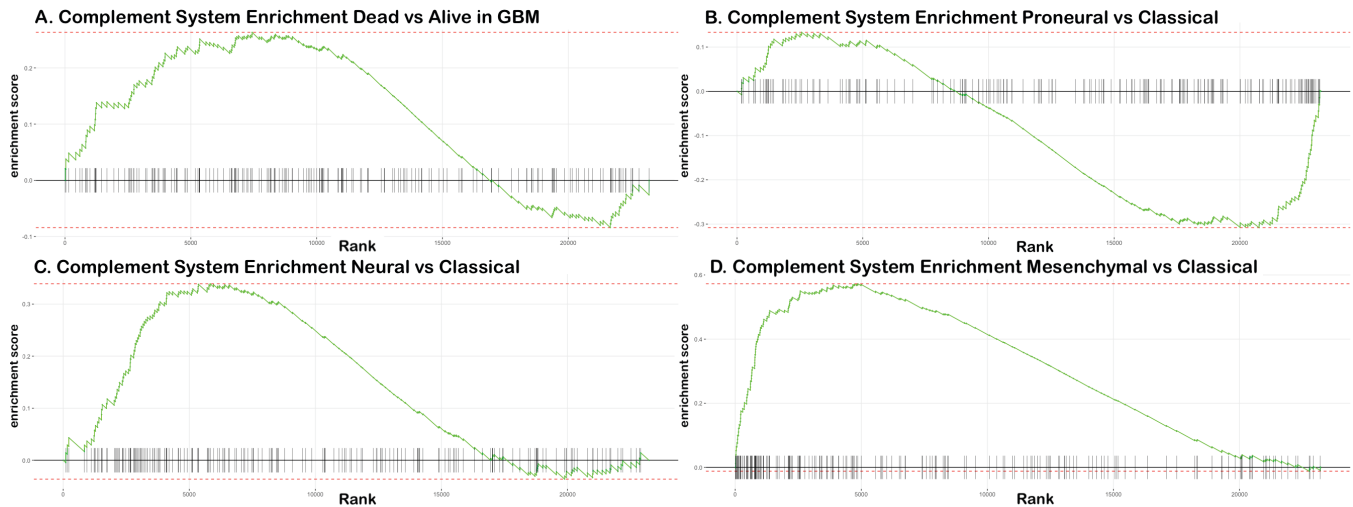
### GSEA Reveals Highest Pathway Enrichment Scores in Mesenchymal Subtype

We performed GSEA to assess whether the complement system pathway itself was differentially enriched across different subtypes. First, we performed Gene Set Enrichment Analysis (GSEA) to assess pathway enrichment between outcome groups (dead vs. alive) for GBM across 50 pathways defined by the hallmark dataset. In GSEA, the enrichment score shows the degree the genes in that specific gene set, which are ranked based on differential gene expression levels, are clustered at either the top of the list (overexpressed) or the bottom of the list (underexpressed). A non-enriched gene set would have its genes distributed evenly. This analysis revealed that the complement system pathway was the 11<sup>th</sup> most enriched pathway with the alive outcome as the control, with an adjusted *p*-value of 3.82E-01 and an enrichment score of 0.26 (Figure 4A).

We then performed GSEA to assess pathway enrichment between transcriptome subtypes (Figure 4B-D). When comparing both proneural and neural subtypes to classical, the complement system pathway was the 11<sup>th</sup> most enriched pathway with the classical subtype as the control, with an adjusted *p*-value of 0.0061 and an enrichment score of 0.34. In the analysis of mesenchymal versus classical enrichment, with the classical subtype as a control, the complement system pathway was the 11<sup>th</sup> most enriched pathway, with an adjusted *p*-value of 5.24E-18 and an enrichment score of 0.57.



**Figure 3: Annotated heatmap with complement system gene expression data.** Heatmap of complement system gene count data normalized using variance stabilizing transformation for GBM (A), with proneural (Ai), classical (Aii), and mesenchymal (Aiii) clusters, and LGG (B) with proneural (3Bi), classical (Bii), and mesenchymal (Biii) clusters. Generated using counts data with the pheatmap package.



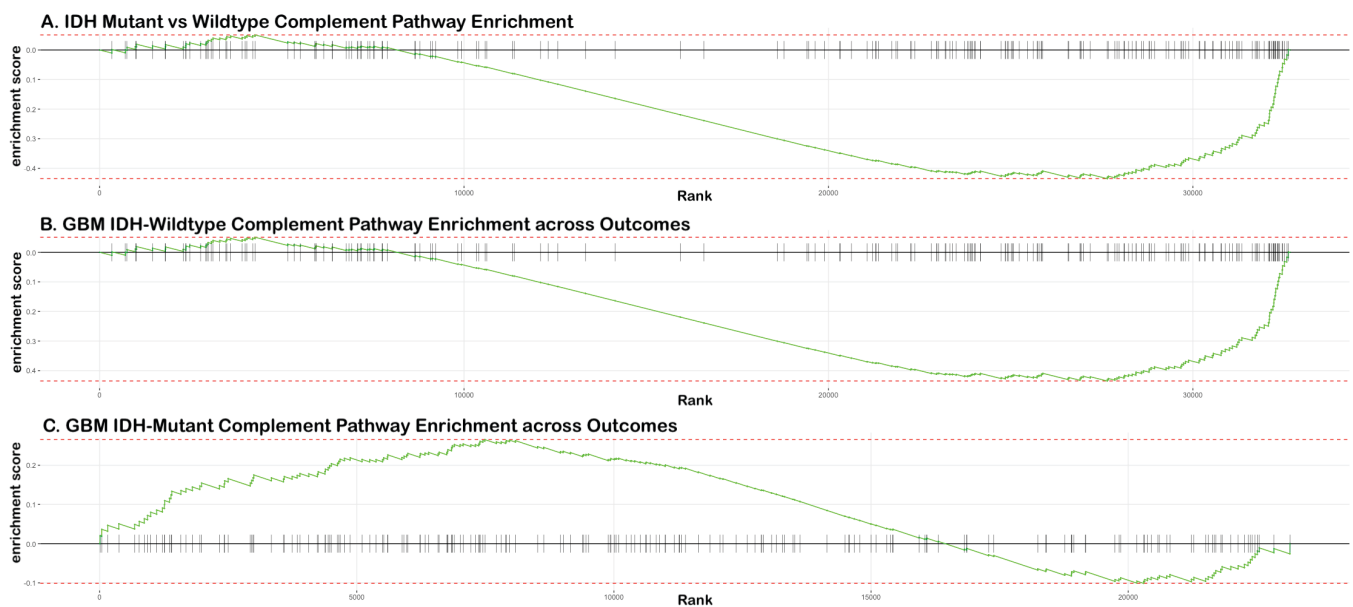
**Figure 4: Complement system enrichment in GBM.** GSEA enrichment plot comparing the complement system enrichment across outcome (Dead vs Alive) (A) and Subtypes (B-D). Generated with the FGSEA package.

We also performed GSEA testing for pathway enrichment between *IDH* mutation subclasses (*IDH*-mutant vs *IDH*-wt). Similarly, the complement system pathway was the 11th most enriched pathway with *IDH*-wt as the control, with an adjusted *p*-value of 3.50E-05 and an enrichment score of -0.44. Analysis was then performed comparing the outcome (dead vs alive) amongst the subtypes for GBM. For the *IDH*-wt, the complement system pathway was the 11th most enriched pathway, with an adjusted *p*-value of 1.97E-05 and an enrichment score of -0.44. For *IDH*-mutant, the complement system pathway was the 11th most enriched pathway, with an adjusted *p*-value of 0.41 and an enrichment score of 0.27 across dead versus alive outcomes (Figure 5A-C).

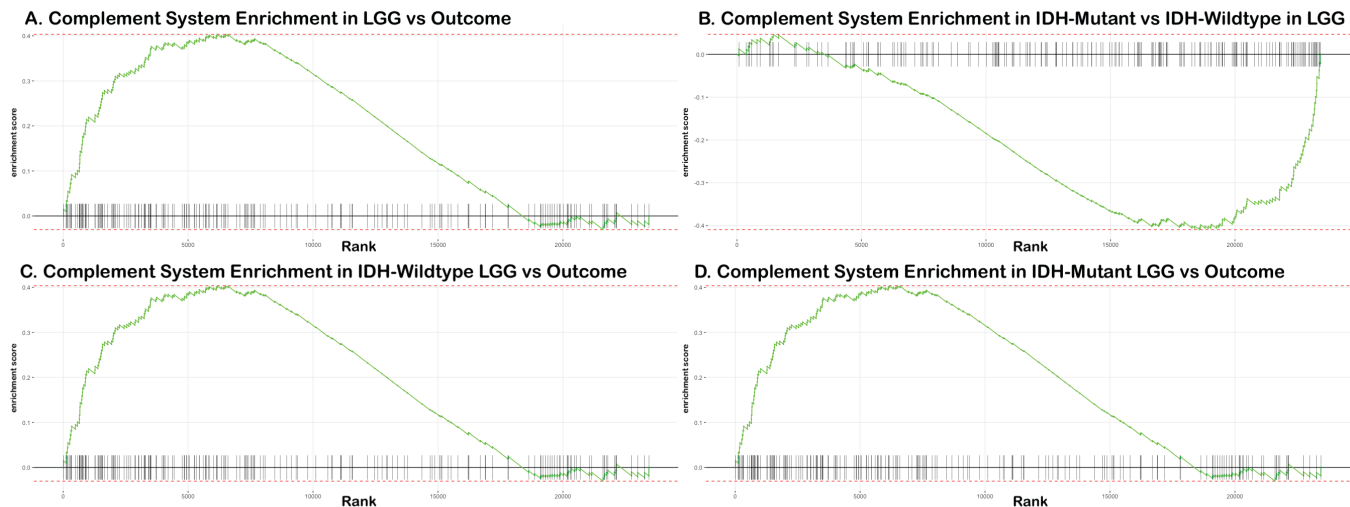
We performed GSEA comparing pathway enrichment with the outcome (dead vs alive (control)) amongst LGG samples. The complement system pathway was the 11th most

enriched pathway with the alive outcome as the control, with an adjusted *p*-value of 1.38E-06 and an enrichment score of 0.40 compared with the control of alive samples (Figure 6A).

Finally, we performed GSEA comparing pathway enrichment with *IDH* mutation status (*IDH*-mutant vs *IDH*-wt) of LGG samples, with *IDH*-wt as the control. This analysis revealed the complement system pathway as the 11th most enriched pathway, with an adjusted *p*-value of 1.79E-05 and an enrichment score of -0.41. Analysis was then performed comparing the outcome (dead vs alive) amongst the subtypes for LGG with the alive outcome as the control. For the *IDH*-wt, the complement system pathway was the 11th most enriched pathway, with an adjusted *p*-value of 2.40E-06 and an enrichment score of 0.40. For *IDH*-mutant, the complement system pathway was the 11th most enriched pathway, with an adjusted *p*-value of 4.42E-06 and an enrichment score of



**Figure 5: Complement system enrichment and *IDH* mutation status in GBM.** GSEA enrichment plot showing gene enrichment comparing *IDH*-mt vs *IDH*-wt (A) and comparing outcome (Dead vs Alive) across *IDH* Mutation Status (B-C). Generated with the FGSEA package.



**Figure 6: Complement System Enrichment in LGG.** GSEA enrichment plot comparing the complement system enrichment across outcome (Dead vs Alive) (A), across *IDH* mutation status (B), and comparing outcome (Dead vs Alive) across *IDH* Mutation Status (C-D). Generated with the FGSEA package.

0.40 (Figure 6B-D).

Overall, this reveals that the complement system pathway is differentially expressed across different subtypes of glioma. Compared with the other transcriptional subtypes, the mesenchymal subtype has the most complement system pathway enrichment. Across *IDH* mutation status in GBM, *IDH*-wt was negatively enriched compared with *IDH*-mt, which was positively enriched when comparing the outcome (dead vs. alive). Across *IDH* mutation status in LGG, both *IDH*-wt and *IDH*-mt were positively enriched comparing the outcome (dead vs. alive).

**SAMR Survival Analysis Reveals Differences in the Number of Genes Associated with Prognosis across Subtypes**

We performed a Significance Analysis of Microarrays (SAMR) survival analysis, focusing on the genes of the complement system, to determine whether complement system genes were associated with patient outcome. For GBM samples overall, 25 complement system genes were found to be associated with poorer prognosis (Figure 7A).

We next performed survival analysis on the four transcriptional subtypes in GBM. The proneural subtype had 27 complement system genes associated with decreased prognosis. The neural subtype had 15 complement system genes associated with poorer prognosis. The mesenchymal

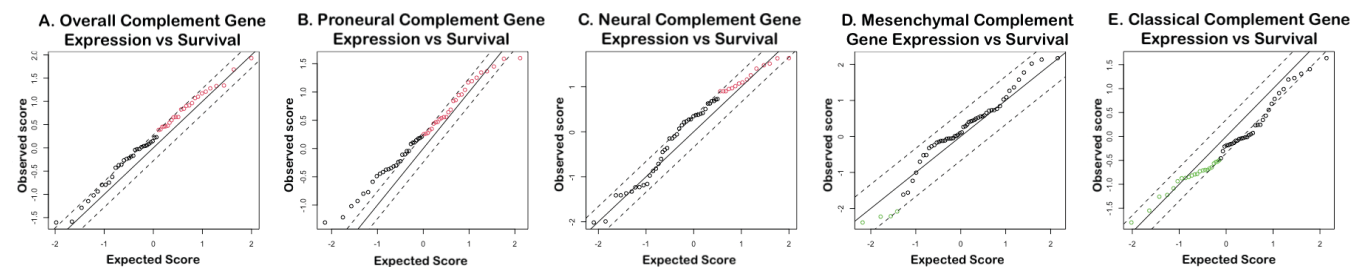
subtype had 4 complement system genes associated with increased prognosis. The classical subtype had 22 complement system genes associated with increased prognosis (Figure 7B-E). We also performed this analysis on *IDH* mutation status categories. *IDH*-wt samples had 62 complement system genes associated with decreased prognosis. *IDH*-mutant samples had 62 complement system genes associated with decreased prognosis (Figure 8A-B).

For LGG samples overall, 55 complement system genes were found to be associated with decreased prognosis. The analysis was also performed on *IDH* mutation status categories for LGG samples. *IDH*-wt samples had 76 complement system genes associated with decreased prognosis. *IDH*-mutant samples had 18 complement system genes associated with decreased prognosis (Figure 8C-E).

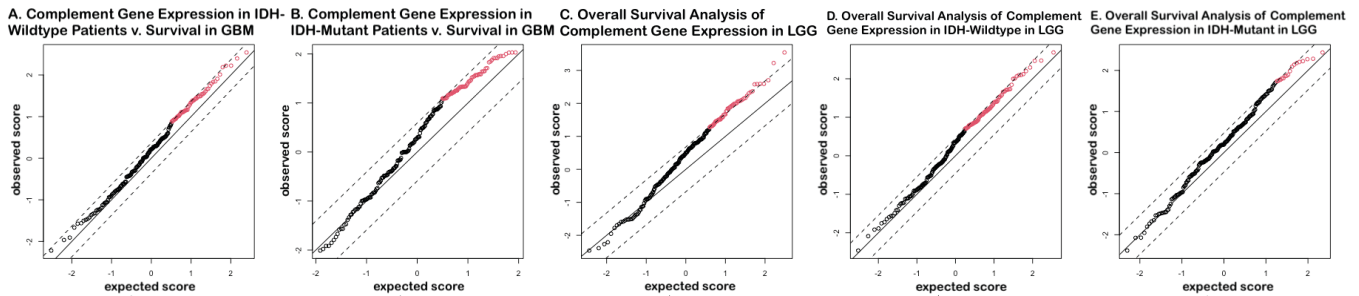
Overall, we determined that complement system genes are potentially associated with varying patient outcomes across different subtypes, as evidenced by the differing number of complement system genes associated with worse prognosis across subtypes.

**DISCUSSION**

In this analysis, we found that complement system genes were differentially expressed across the various subtypes of gliomas with variations in expression levels that could correlate with phenotype. Hierarchical clustering analysis



**Figure 7: Complement system gene expression vs survival for GBM.** Q-Q Plot shows genes associated with worse prognosis (red) and genes associated with better prognosis (green) for overall GBM (A), and across subtypes (B-D). Generated using SAMR.



**Figure 8: Complement system gene expression vs survival across *IDH* mutation status.** Q-Q Plot shows genes associated with worse prognosis (red) and genes associated with better prognosis (green) *IDH* Mutation Status for GBM (A-B), for the overall LGG gene expression (C), and comparing outcome (Dead vs Alive) across *IDH* Mutation Status (D-E). Generated using SAMR.

also suggested that the expression of the complement system genes may in part supervise sub-clustering of glioma patients based on mutational and global transcriptional subtypes.

Overall, the proneural subtype has the most genes associated with a worse prognosis, while the mesenchymal subtype had the most differentially expressed genes. When the gene expression levels of the proneural subtype were compared with the classical subtype (which does respond to aggressive treatment), the proneural subtype had 27 overexpressed complement system genes and 40 underexpressed complement system genes, making it the only subtype with more underexpressed genes than overexpressed genes in the complement pathway. The proneural subtype also had a positive enrichment score across the complement system based on GSEA compared with the classical subtype, and 27 genes in which increased expression were associated with worse prognosis. This was the highest number of genes associated with worse prognosis. This analysis suggests the greater number of underexpressed genes can potentially be associated with the proneural subtype's greater survival rates as age increases, as complement system gene expression has been associated with a poor prognosis due to increased expression of hypoxic gene signatures (10).

In the mesenchymal subtype, there were 107 complement system genes differentially expressed (the most out of the subtypes), with most of the genes being upregulated. This subtype also had the highest enrichment score across the complement system compared with the other subtype comparisons. This might suggest a correlation between the phenotypes observed with this subtype and gene expression. Surprisingly, there were only four complement system genes that were associated with a poorer prognosis. This is an unusual result considering the mesenchymal subtype had the most differentially expressed complement system genes, and it would be expected that the differentially expressed complement system genes would result in a worse prognosis. Further investigation is needed to fully determine the role of the complement system in this subtype.

The neural subtype was associated with the lowest level of complement system gene expression across the subtypes, which could be due to its small sample size. It also had a positive enrichment in the complement pathway when compared with the classical subtype. It had 15 genes associated with a higher prognosis.

The varying gene expression across the proneural subtype, which had the most genes associated with poor prognosis, and the mesenchymal subtype, which had the

most genes differentially expressed overall, with a less clear correlation with the outcome, could potentially indicate a correlation between complement system gene expression and the proneural-mesenchymal transition associated with recurrence, metastasis, and poor prognosis (15). This could suggest a correlation between metastasis and complement system gene expression, which should be further investigated.

Additionally, *IDH*-wt had the most complement system genes overexpressed and is associated with a worse prognosis. The *IDH* mutation status of glioma is known to be associated with prognosis, with *IDH*-wt being associated with a worse prognosis than the *IDH*-mutant subtype. In LGG overall, there were 33 complement system genes differentially expressed across *IDH*-wt and *IDH*-mutant, most of which were upregulated, with the baseline being *IDH*-wt. There was negative complement gene set enrichment in the mutant vs wildtype. When comparing gene enrichment across outcomes, there were more complement system genes associated with a worse prognosis in the wildtype subtype versus the mutant subtype. In GBM, there were 21 complement system genes differentially expressed across the wildtype subtype versus the mutant subtype, with most of the genes being upregulated. When considering patient outcomes alone (dead vs alive), there was the same number of genes associated with a worse prognosis in both *IDH*-mutant and *IDH*-wildtype. This indicates that there might be a correlation between the expression of complement system genes and the different phenotypes of the subtypes in LGG.

Overall, this analysis shows that the role of complement system gene expression is mixed and should be investigated further. Our results do indicate that there might be a correlation between complement system gene expression and prognosis amongst the different subtypes.

Further studies should be conducted to determine if there are any actionable targets in the complement system for treatment. Additional analysis should also be performed to determine if there is any correlation between complement system gene expression and metastasis and/or recurrence.

Overall, this research found that the genes of the complement system were differentially expressed across the various subtypes of gliomas in our dataset. Within the transcriptional subtypes, the complement system genes were more overexpressed in subtypes associated with a worse response to treatment or increased tumor malignancy. Comparing the *IDH* mutation subtype status, most complement system genes were overexpressed when comparing the wildtype subtype, associated with a worse

prognosis, with the mutant subtype, associated with a better prognosis. This suggests a potential connection between outcomes of the glioma subtypes with the expression of complement system genes, although more research needs to be done to confirm this.

### MATERIALS AND METHODS

All data were collected from the TCGA database and is publicly available. The TCGA project is a genomics project that has sequenced the genome, transcriptome, epigenome, and proteome of 33 cancer types characterized by poor prognosis and far-reaching impact on public health. A total of 326 RNASeq Data samples were downloaded using TCGABiolinks (16), a Bioconductor package, in the form of HTSeq Counts, which were then compiled into a RangedSummarizedExperiment, an R object, for downstream analysis in R.

#### Patient Cohort for GBM Samples

There was a total of 174 samples in the GBM cohort, all of which came from patients diagnosed with GBM. There was a total of 109 male patients (62.6%), 59 female patients (33.9%), and 6 patients whose gender was not reported (0.03%). This data set contained 4 Hispanic or Latino patients, 136 non-Hispanic or Latino patients, and 34 patients whose ethnicities were not reported. There were 5 patients of Asian descent, 11 patients of African American descent, 150 white patients, and 8 patients whose races were not reported. Overall, 30 patients were reported as being alive upon the last contact with a physician, 136 patients were reported as dead, and 8 patients did not have a vital status reported.

#### Patient Cohort of LGG Samples

There was a total of 152 patient samples in this cohort, all of whom were diagnosed with LGG. There was a total of 99 male patients, 52 female patients, and 1 patient whose gender was not reported. This data set contained 1 Hispanic or Latino patient, 125 non-Hispanic or Latino patients, and 26 patients whose ethnicities were not reported. There were 5 patients of Asian descent, 10 patients of African American descent, 135 white patients, and 2 patients whose races were not reported. Overall, 27 patients were alive, 122 patients were dead, and 3 patients did not have a vital status reported.

#### Statistical Analysis Software

R and Bioconductor packages were used to perform the data analysis and statistical tests. Bioconductor is a free, open-source software repository based on the R programming language (17). The packages used for this analysis are DESeq2 (18), pheatmap (19), FGSEA (20), and SAMR (21).

#### Complement System Gene Set

The complement system gene set was downloaded from the Hallmark Pathways Gene Sets (22) in the Molecular Signatures Database. There was a total of 200 genes in the complement system gene set used.

#### DESeq2

The DESeq2 package calculates differential gene expression using negative binomial generalized linear models. A DESeq2 dataset was generated with estimates of dispersion, logarithmic fold changes (LFC), and  $p$ -values

for each gene. Low-count genes (counts  $\leq 10$ ) were filtered out before performing the analysis. Apeglm LFC shrinkage (23) was used to control statistical noise across replicates by shrinking LFC estimates to zero for low-count genes. This dataset was used to compare gene expression between the four different transcriptional subtypes (proneural vs classical, mesenchymal vs classical, and neural vs classical) and *IDH* subtypes (*IDH*-wt vs *IDH*-mutant), the two outcomes (dead vs alive), and the outcomes within the subtypes. This analysis was used to generate an MA (log ratio (M) and log average (A)) plot with all the genes and with the subset of complement system genes.

#### pheatmap

This package was used to generate a heatmap with hierarchical clustering, focusing on the complement system genes. The count matrix from the DESeq Object generated by the DESeq functions first underwent a variance stabilizing transformation. A matrix of the sample-to-sample distances was created from the transformed matrix. This matrix then underwent hierarchical clustering, which grouped the samples based on similarities with each other. Finally, a heatmap was generated with the sample distances, the clustered dendrogram, and the heatmap annotations with the outcome (alive vs dead), transcriptional subtype, *IDH* mutation status, Karnofsky Performance Score, and survival in months. The number of times the tree would be cut was determined by plotting the sum of squared errors (SSE) across the cluster solutions and choosing the number of clusters based on when there was a dramatic change in the slope of the SSE ('the elbow'). Based on that, the heatmap was cut 5 times. The clusters were grouped into 'proneural', 'mesenchymal', and 'classical' groups based on the number of transcriptional subtype cases in each cluster.

#### SAMR (Significance Analysis of Microarrays)

This package was used to correlate complement gene expression patterns with survival. The gene expression matrix was extracted from the Summarized Experiment (an R Object) using the assay function, with only the rows with the complement genes selected. The survival in months was used as the outcome analysis, with the vital status (alive vs dead) used as the censoring status. The list of significant genes associated with better/poor prognosis was plotted on a Q-Q Plot. The expected score refers to the average gene expression distribution. The observed score refers to the actual distribution of gene expression data. Downregulated and upregulated genes were determined based on the False Discovery Rate (FDR < 0.2) and the respective delta values. This analysis was done across the glioma subtypes.

#### ACKNOWLEDGMENTS

We would like to acknowledge TCGA for the raw data used for this analysis and Polygence for providing us with this opportunity to work on this research.

**Received:** October 17, 2021

**Accepted:** September 21, 2022

**Published:** May 17, 2023



## REFERENCES

1. "Glioblastoma Multiforme." AANS, [www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme](http://www.aans.org/en/Patients/Neurosurgical-Conditions-and-Treatments/Glioblastoma-Multiforme).
2. "Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas." *New England Journal of Medicine*, vol. 372, no. 26, 2015, pp. 2481–2498., doi:10.1056/nejmoa1402121.
3. "Glioma vs. Glioblastoma: What's the Difference?" *Roswell Park Comprehensive Cancer Center*, [www.roswellpark.org/cancertalk/202105/glioma-vs-glioblastoma-whats-difference](http://www.roswellpark.org/cancertalk/202105/glioma-vs-glioblastoma-whats-difference).
4. Verhaak, Roel G.W., *et al.* "Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1." *Cancer Cell*, vol. 17, no. 1, 2010, pp. 98–110., doi:10.1016/j.ccr.2009.12.020.
5. Han, Sue, *et al.* "IDH Mutation in Glioma: Molecular Mechanisms and Potential Therapeutic Targets." *British Journal of Cancer*, vol. 122, no. 11, 2020, pp. 1580–1589., doi:10.1038/s41416-020-0814-x.
6. Guan, Xiaowei, *et al.* "Molecular Subtypes of Glioblastoma Are Relevant to Lower Grade Glioma." *PLoS ONE*, vol. 9, no. 3, 2014, doi:10.1371/journal.pone.0091216.
7. Yap, Joshua. "Isocitrate Dehydrogenase: Radiology Reference Article." *Radiopaedia Blog RSS*, Radiopaedia.org, 13 Oct. 2022, [radiopaedia.org/articles/isocitrate-dehydrogenase?lang=us](https://radiopaedia.org/articles/isocitrate-dehydrogenase?lang=us).
8. Tommasini-Ghelfi, Serena, *et al.* "Cancer-Associated Mutation and beyond: The Emerging Biology of Isocitrate Dehydrogenases in Human Disease." *Science Advances*, vol. 5, no. 5, 2019, doi:10.1126/sciadv.aaw4543.
9. Han, Sue, *et al.* "IDH Mutation in Glioma: Molecular Mechanisms and Potential Therapeutic Targets." *British Journal of Cancer*, vol. 122, no. 11, 2020, pp. 1580–1589., doi:10.1038/s41416-020-0814-x.
10. Afshar-Kharghan, Vahid. "The Role of the Complement System in Cancer." *Journal of Clinical Investigation*, vol. 127, no. 3, 2017, pp. 780–789., doi:10.1172/jci90962.
11. Olcina, Monica M., *et al.* "Mutations in an Innate Immunity Pathway Are Associated with Poor Overall Survival Outcomes and Hypoxic Signaling in Cancer." *Cell Reports*, vol. 25, no. 13, 2018, doi:10.1016/j.celrep.2018.11.093.
12. Bouwens van der Vlis, T. A., *et al.* "The Complement System in Glioblastoma Multiforme." *Acta Neuropathologica Communications*, vol. 6, no. 1, 2018, doi:10.1186/s40478-018-0591-4.
13. La-Beck, Ninh M., *et al.* "Nanoparticle-Induced Complement Activation: Implications for Cancer Nanomedicine." *Frontiers in Immunology*, vol. 11, 2021, doi:10.3389/fimmu.2020.603039.
14. "NCI Dictionary of Cancer Terms." *National Cancer Institute*, [www.cancer.gov/publications/dictionaries/cancer-terms/def/karnofsky-performance-status](http://www.cancer.gov/publications/dictionaries/cancer-terms/def/karnofsky-performance-status).
15. Fedele, Monica, *et al.* "Proneural-Mesenchymal Transition: Phenotypic Plasticity to Acquire Multitherapy Resistance in Glioblastoma." *International Journal of Molecular Sciences*, vol. 20, no. 11, 2019, p. 2746., doi:10.3390/ijms20112746.
16. Colaprico, Antonio, *et al.* "TCGAbiolinks: An R/ Bioconductor Package for Integrative Analysis of TCGA Data." *Nucleic Acids Research*, vol. 44, no. 8, 2015, doi:10.1093/nar/gkv1507.
17. Huber, Wolfgang, *et al.* "Orchestrating High-Throughput Genomic Analysis with Bioconductor." *Nature Methods*, vol. 12, no. 2, 2015, pp. 115–121., doi:10.1038/nmeth.3252.
18. Love, Michael I, *et al.* "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with *deseq2*." *Genome Biology*, vol. 15, no. 12, 2014, doi:10.1186/s13059-014-0550-8.
19. "Package Pheatmap." *CRAN*, Comprehensive R Archive Network (CRAN), [cran.r-project.org/web/packages/pheatmap/index.html](http://cran.r-project.org/web/packages/pheatmap/index.html).
20. Korotkevich, Gennady, *et al.* "FAST Gene Set Enrichment Analysis." 2016, doi:10.1101/060012.
21. Tusher, Virginia Goss, *et al.* "Significance Analysis of Microarrays Applied to the Ionizing Radiation Response." *Proceedings of the National Academy of Sciences*, vol. 98, no. 9, 2001, pp. 5116–5121., doi:10.1073/pnas.091062498.
22. Liberzon, Arthur, *et al.* "The Molecular Signatures Database Hallmark Gene Set Collection." *Cell Systems*, vol. 1, no. 6, 2015, pp. 417–425., doi:10.1016/j.cels.2015.12.004.
23. Zhu, Anqi, *et al.* "Heavy-Tailed Prior Distributions for Sequence Count Data: Removing the Noise and Preserving Large Differences." *Bioinformatics*, vol. 35, no. 12, 2018, pp. 2084–2092., doi:10.1093/bioinformatics/bty895.

**Copyright:** © 2023 Mudda and Wilson. All JEI articles are distributed under the attribution non-commercial, no derivative license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>). This means that anyone is free to share, copy and distribute an unaltered article for non-commercial purposes provided the original author and source is credited.