Neuronal and glioma-derived stem cell factor (SCF) expression on brain angiogenesis

Melissa CHUA Wan Ying

AS.410.671.82.SU21

Final Project, 8/17/21

Outline

- Introduction
- Background
- Dataset
- Initial central location and spread
- Outliers
- Gene filtering
- SCF (KITLG) gene expression
- Feature Selection and Gene Expression
- Dimensionality reduction
- Clustering analysis
- Sample Classification
- Top 5 discriminant genes
- Conclusions
- References

Introduction

- **Malignant gliomas** are among the most lethal tumors, with median survivals of less than a year for patients with the most common type of glioma (i.e., glioblastoma) despite aggressive treatment.
- High-grade gliomas differ significantly from low-grade gliomas both biologically and genetically.
- **High-grade gliomas** include **grade III anaplastic astrocytoma, grade IV astrocytoma, or glioblastoma**.
- **Low-grade gliomas** include **grade II astrocytoma**.

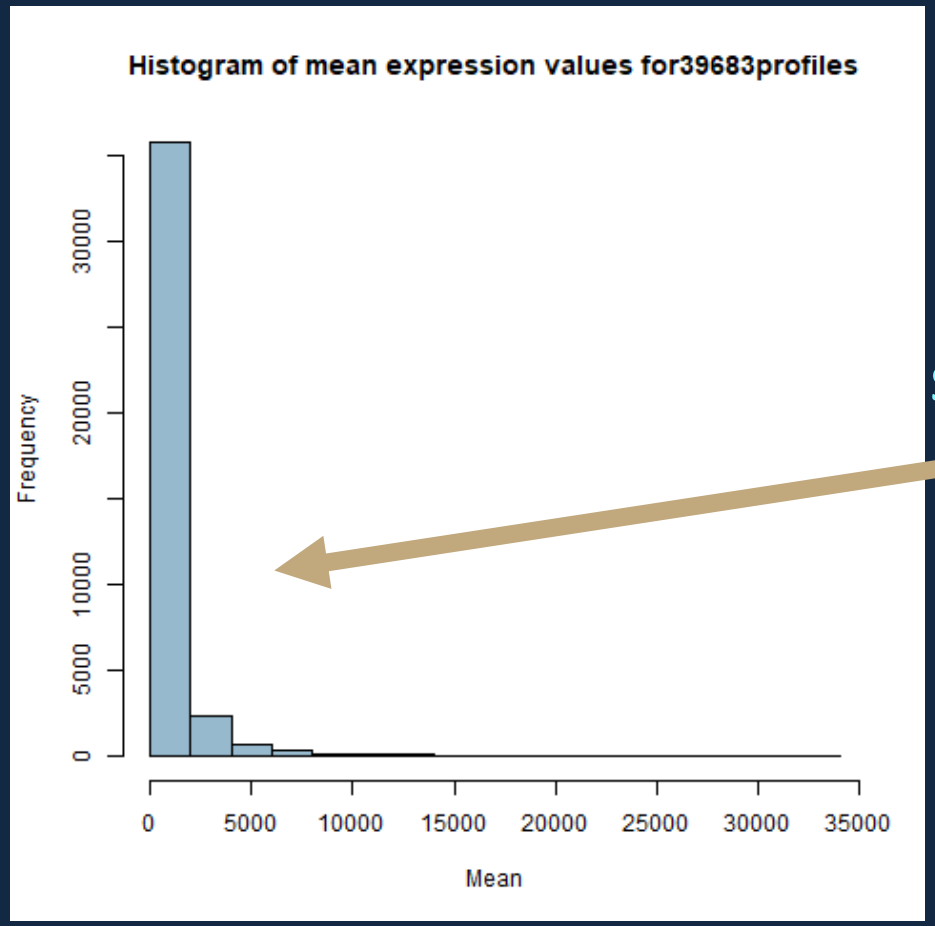
Background

- **Angiogenesis** is a hallmark of malignant gliomas. **Stem cell factor (SCF)** is overexpressed by neurons following brain injury as well as by glioma cells. Hepatic injury is shown to induce SCF production, which in turn promotes liver regeneration, an angiogenesis-dependent process.
- Sun et al., proposed that **SCF** is overexpressed by neurons following brain injury as well as by glioma cells by directly activating brain microvascular endothelial cells (ECs). They also proposed that primary human gliomas express SCF in a **grade-dependent manner** and induce non-tumor neurons to express SCF in brain regions infiltrated by glioma cells, areas that colocalize with prominent angiogenesis [1].
- They performed **mRNA profiling** of SCF expression levels in gliomas of different grades using the **Affymetrix Human Genome U-133 plus 2.0 GeneChip**. The **microarray** expression data were confirmed by quantitative real-time RT-PCR.
- For this analysis, glioma (grade II, III, and IV) samples were compared against non-tumor samples of epilepsy patients. Of the glioma samples, astrocytomas are grade II or III, glioblastomas are grade IV, whereas oligodendrogliomas are grade II or III. Differential expression of SCF (KITLG) gene is also analyzed across all the samples.

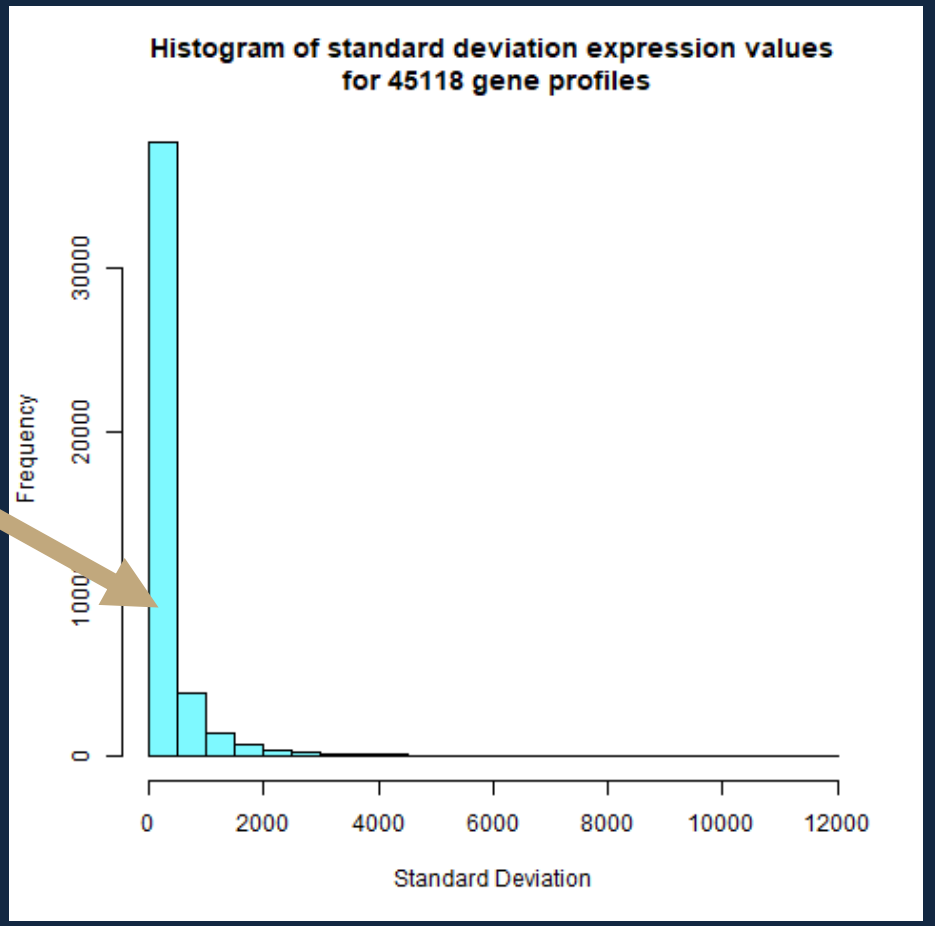
Dataset: GDS1962

- **Platform:** GPL570
- **Type:** Expression profiling by array
- **Organism:** Homo sapiens
- **Sample type:** mRNA from brain tumor tissues
- **Series published:** 10/04/2006
- **Genes:** 54,675
- **Samples:** 180
  - 4 groups:
    1. Non-tumor samples from epilepsy patients (n=23)
    2. Grade II Glioma (7 astrocytomas grade II, 38 oligodendrogliomas grade II)
    3. Grade III Glioma (19 astrocytomas grade III, 12 oligodendrogliomas grade III)
    4. Grade IV Glioma (81 glioblastomas grade IV)

Initial Central location and spread



Positive skewness

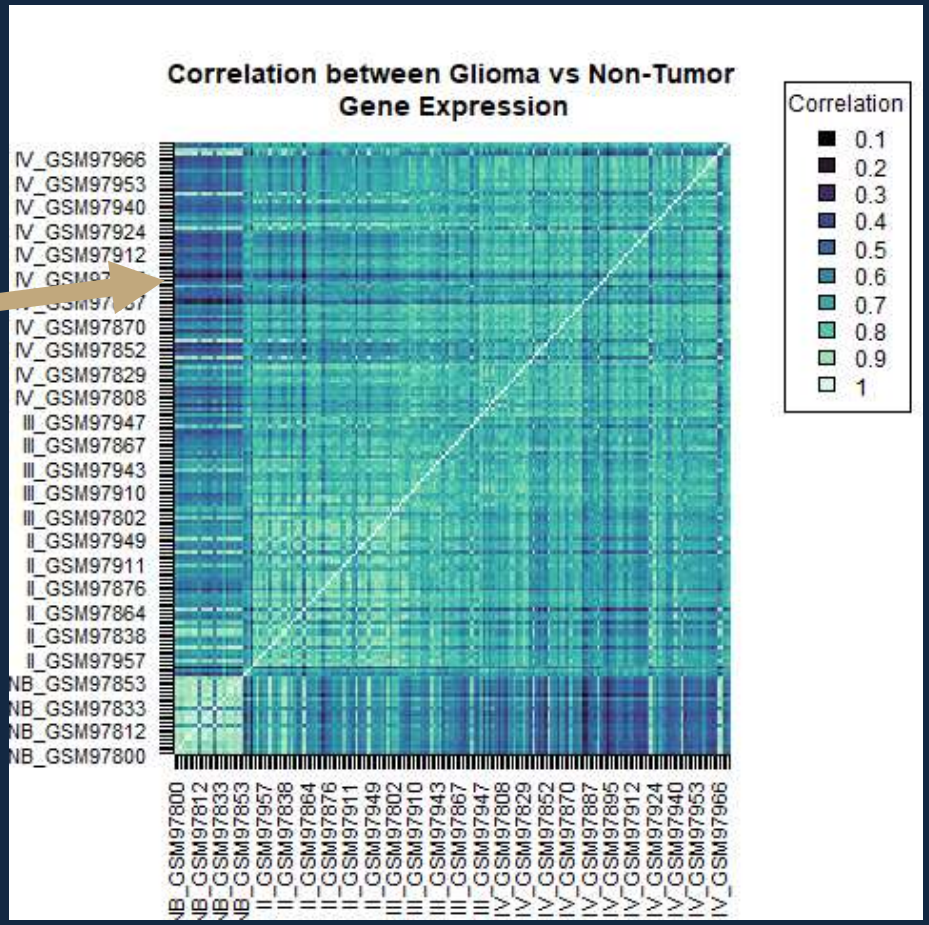


OUTLIERS

- Four methods used:
  - Correlation plot – heat map
  - Hierarchical clustering dendrogram
  - CV vs. Mean
  - Average correlation

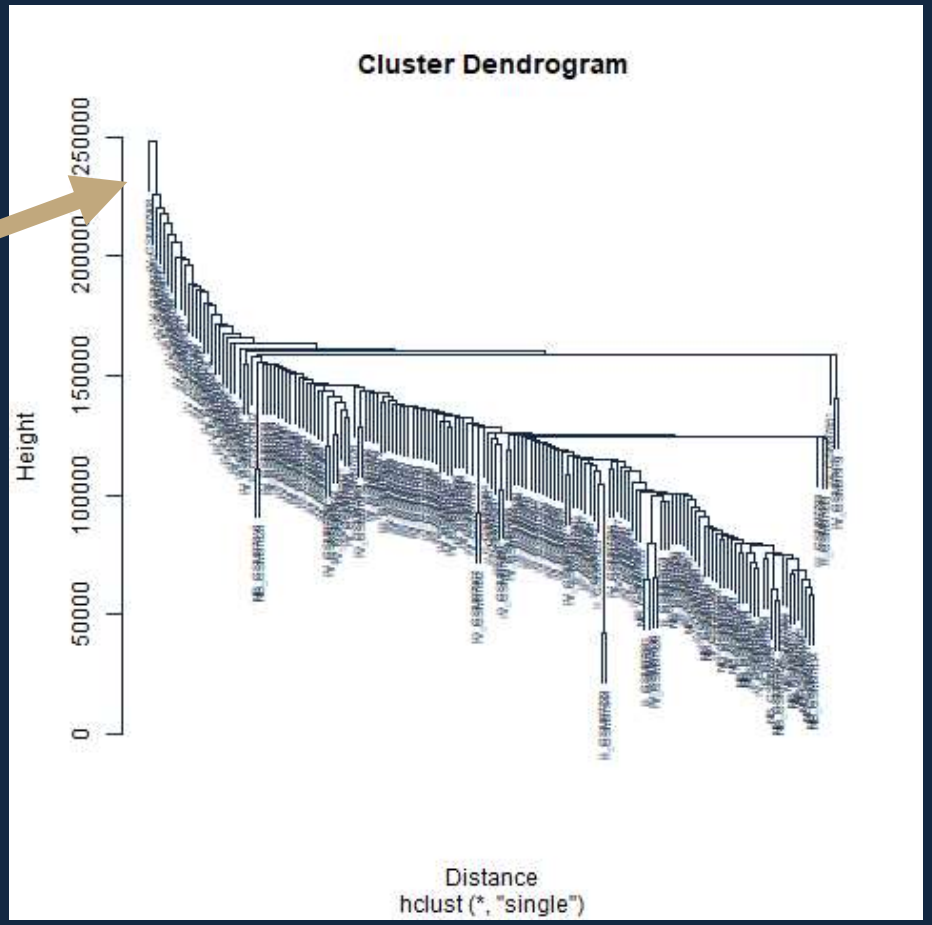
Outliers – Correlation plot

Potential outlier:
GSM97878



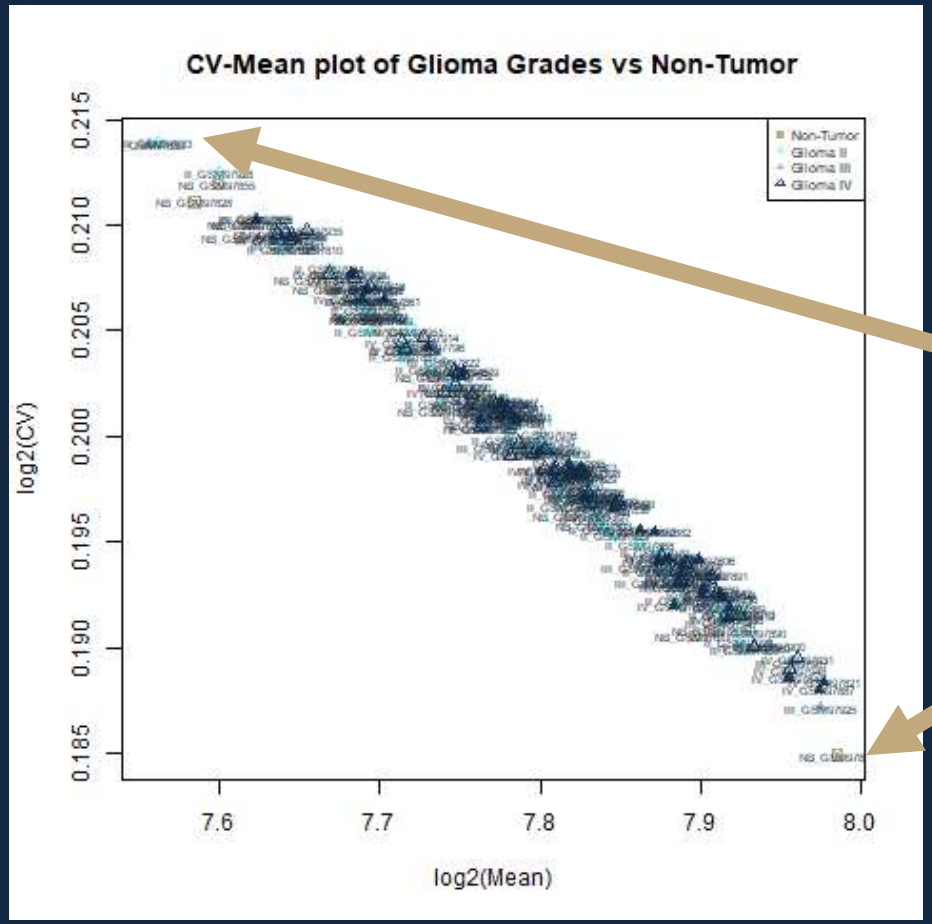
Outliers – Cluster Dendogram

Potential outlier:
GSM97878



Outliers – CV vs. Mean

The coefficient of variance (CV) is inversely correlated to mean

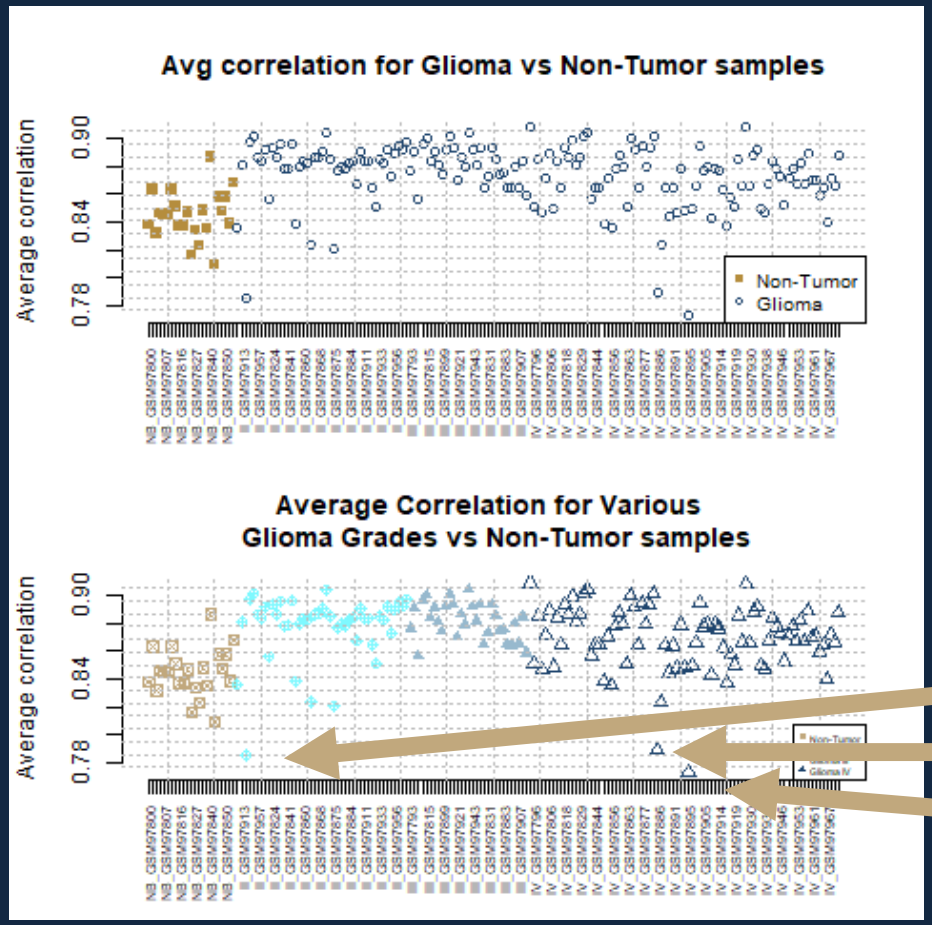


Potential outlier:

GSM97913

GSM97878

Outliers – Average correlation



The average correlation for non-tumor samples is relatively lower than that of glioma samples.

Potential outliers:

GSM97913

GSM97886

GSM97895

Outliers summary

- Correlation plot – no discernable outlier(s)
- Cluster dendrogram – GSM97878
- CV vs Mean – GSM97878, GSM97913
- Average correlation – GSM97913, GSM97886, GSM97895
- outlier() function – GSM97895

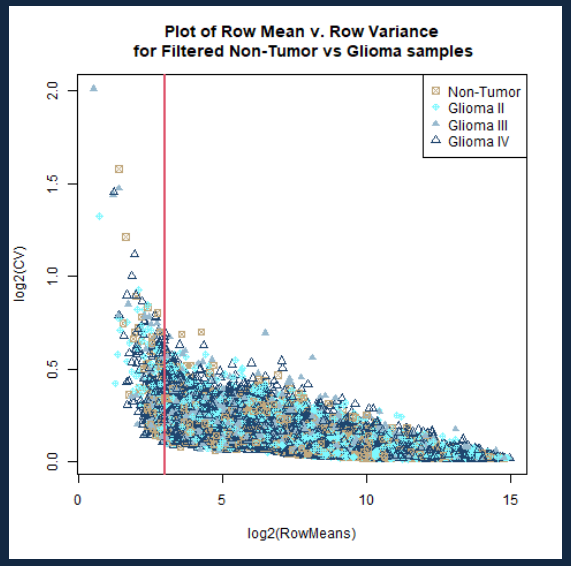
- All the potential outliers are removed
- New sample size = 176
  - Non-tumor (n=23)
  - Glioma (n=153)

Gene filtering

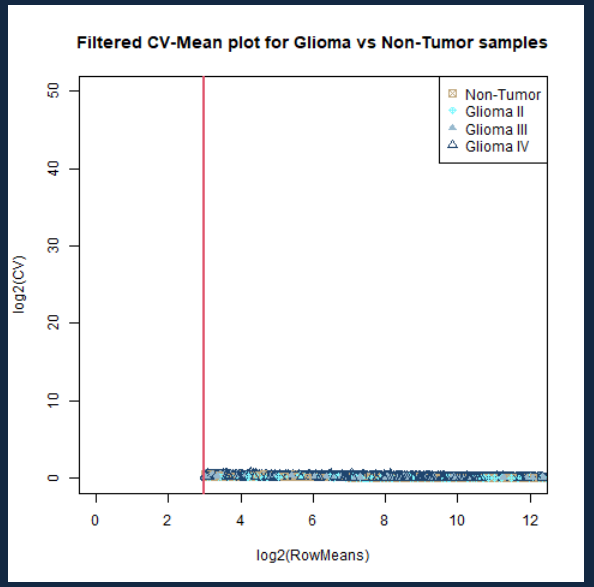
- Gene probes were removed if the average gene expression is more than 0 or less than 3 on a log2 scale
- 436 probes were removed

Quant' Dist'	0%	25%	50%	75%	100%
RowMeans	0.9473545	6.5153297	8.0221536	9.5758638	15.0504017

Mean vs Variance of Samples with Probes > 0 on log2 scale



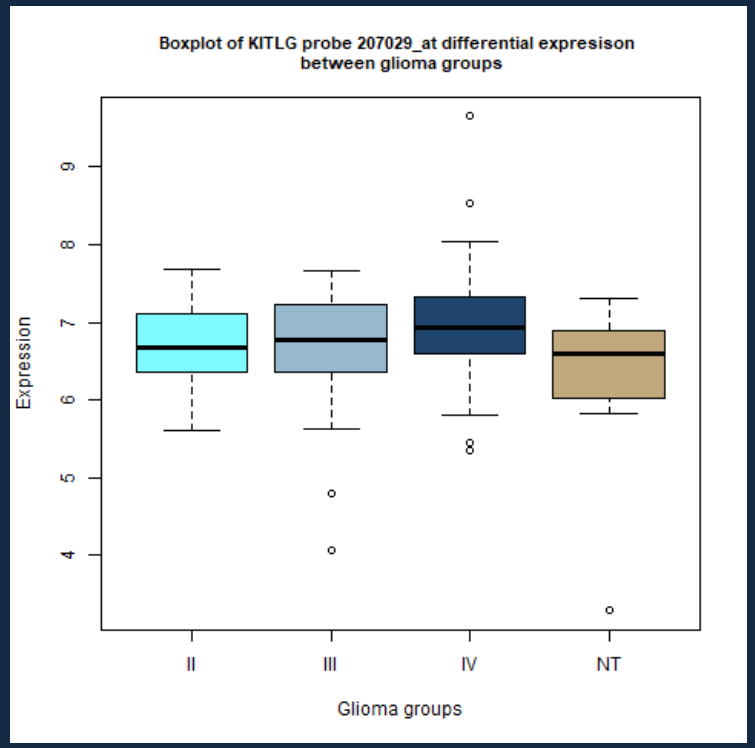
Probes (P) with rowMeans $0 \leq \log(P_{mean}) \leq 3$

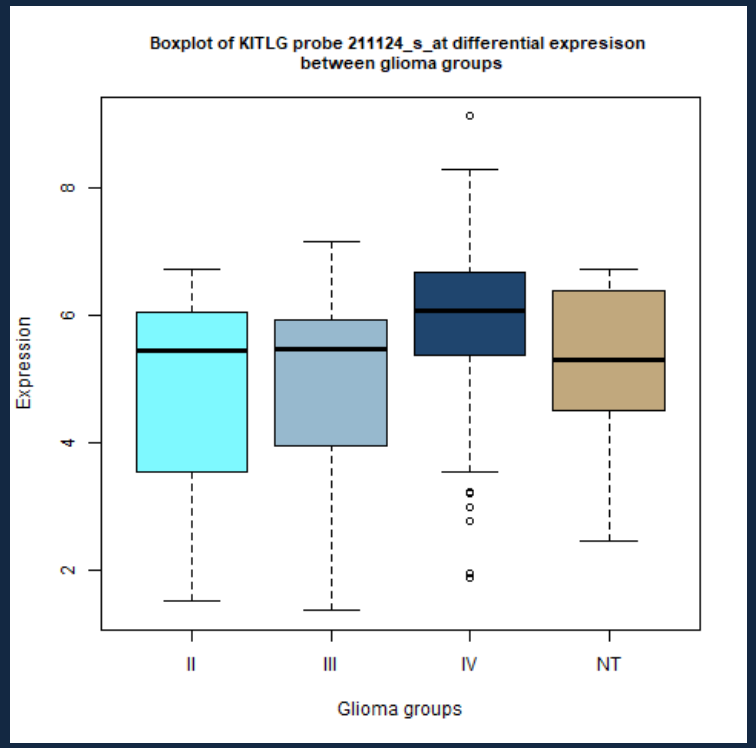


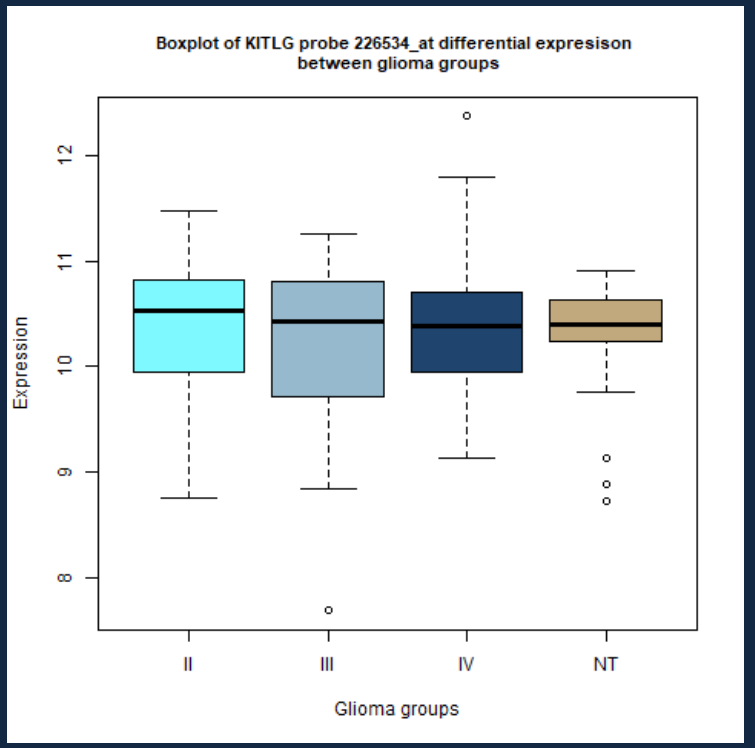
SCF (KITLG) gene expression

NT = Non-Tumor (Control)
II = Glioma Grade II
III = Glioma Grade III
IV = Glioma Grade IV

KITLG gene expression increases (marginally) in gliomas in a grade-dependent manner





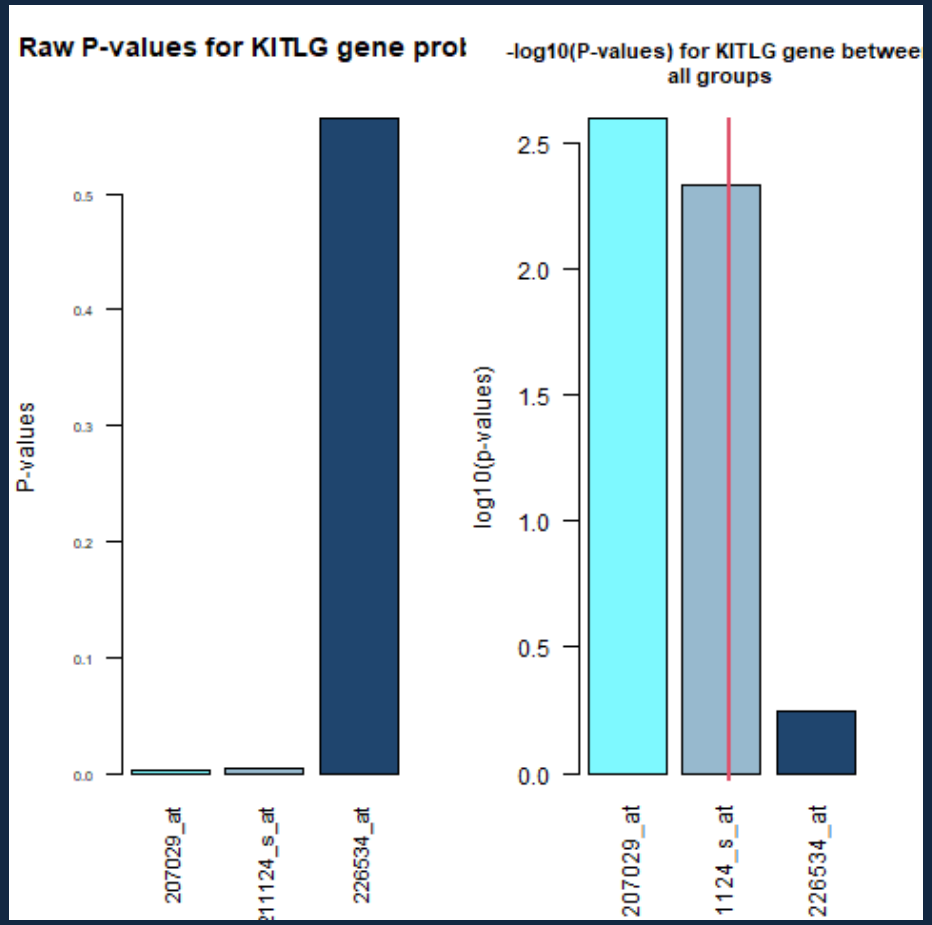


Findings:

Based on the 3KITLG probes, we cannot confidently corroborate previous findings that SCF gene is significantly expressed in glioma in a grade-dependent manner. This contradiction might be attributed to the control samples which were derived from brain tissues of epilepsy patients. Given that various forms of brain injury may induce neuronal expression of SCF, it is unsurprising that epilepsy-induced brain injury might contribute to the relatively high baseline SCF gene expression levels. All four KITLG gene probes were eliminated during feature selection.

SCF (KITLG) gene expression

One-Way ANOVA analysis showed that the **226534_at** KITLG gene probe is much higher p-value than **207029_at and 211124_s_at**



Note:

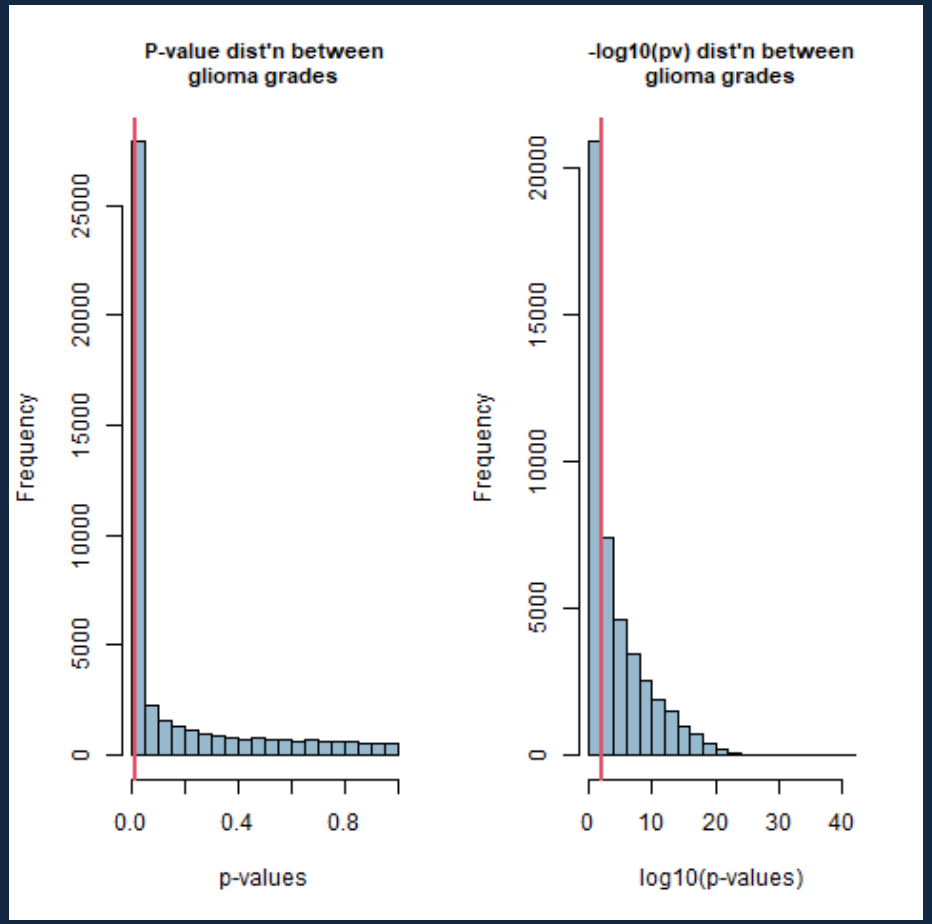
(1) Out of the four KITLG gene probes, only three passed the gene filtering process

(2) None of the KITLG probes are shortlisted by Feature Selection (next section)

Feature selection and Gene Expression - Methods

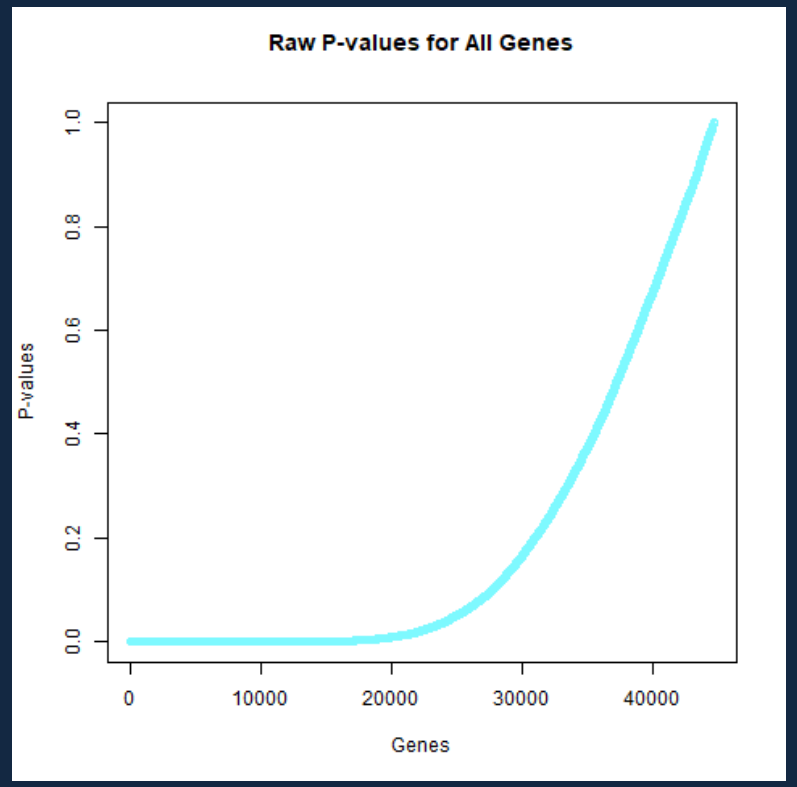
- **One-way Anova test** for comparing gene expression in different glioma grades
- **Student's T-test** for comparing gene expression between Non-Tumor (Control) vs Glioma tumor samples
- Genes with p-value <0.05: **8663**
- Genes with p-value <0.01: **7401**
- Genes with p-value <0.05 and linear fold >2: **3904**
- Linear fold change range: **0.03848994** (min) to **58.95058** (max)
- Feature selection methods: **Bonferroni, Holm-Bonferroni** and **SVM-RFE**
- Genes retained: **8481** (after feature selection)

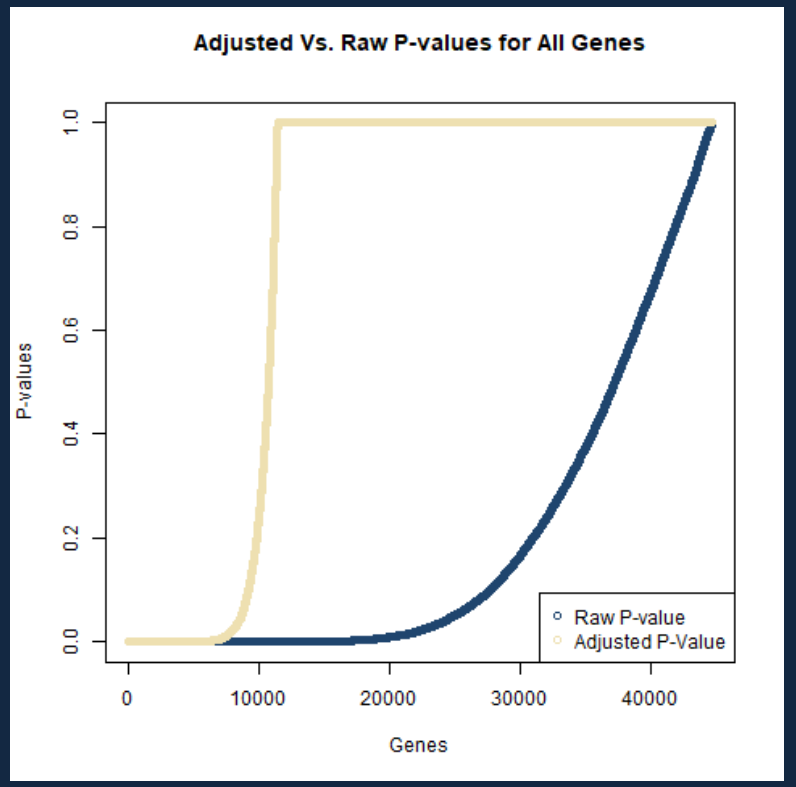
Feature selection: P-value distribution across glioma grades

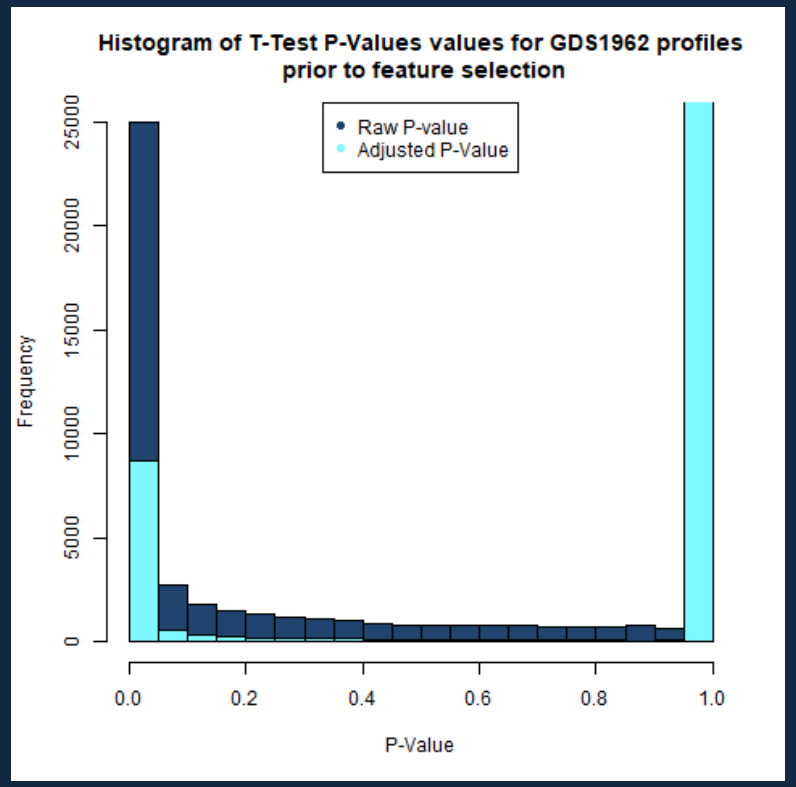


Feature selection: P-value distribution between control vs glioma samples

P-values adjusted using Bonferroni and Holm-Bonferroni methods resulted in an increase in P-value of 1

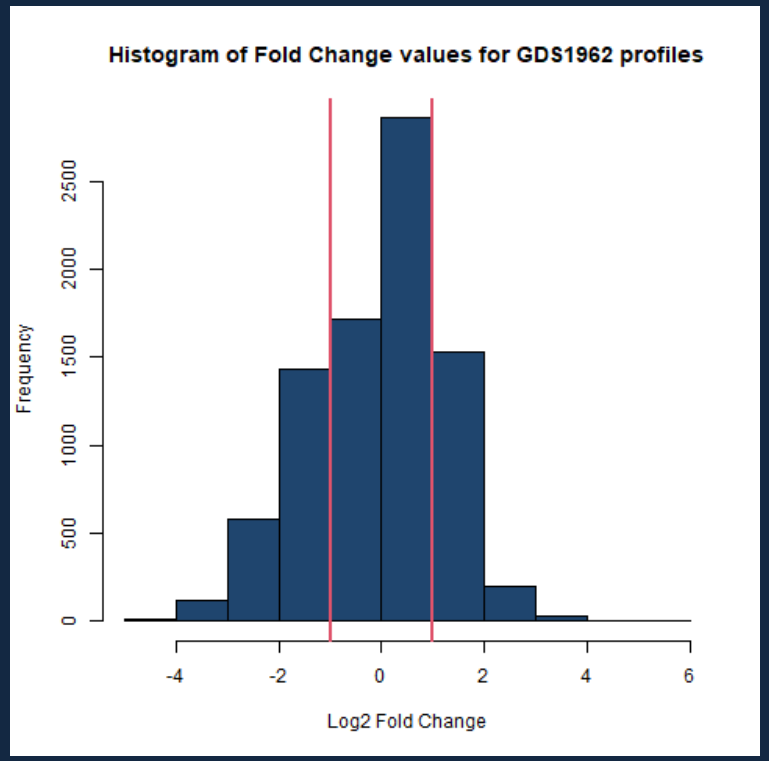


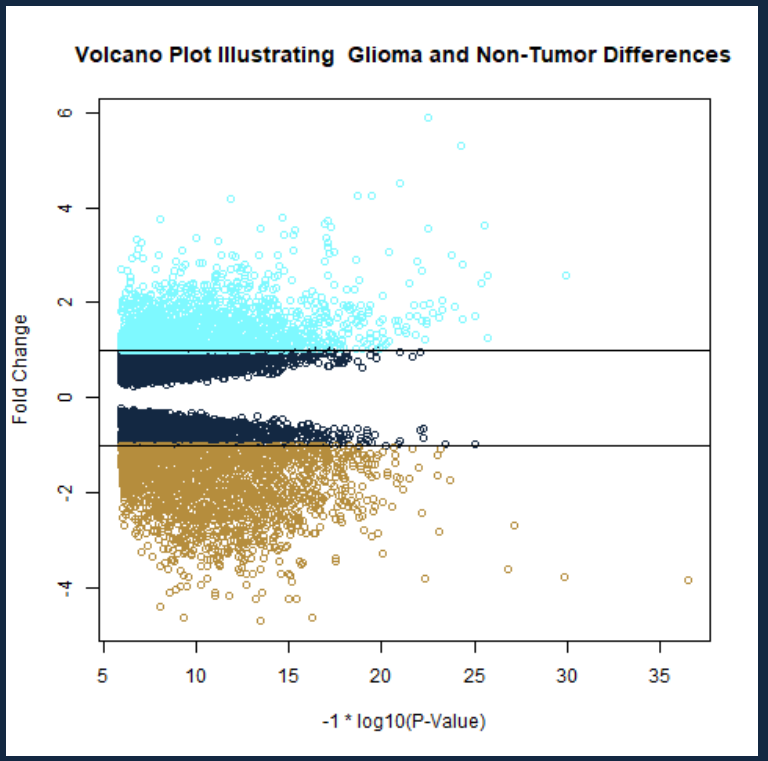




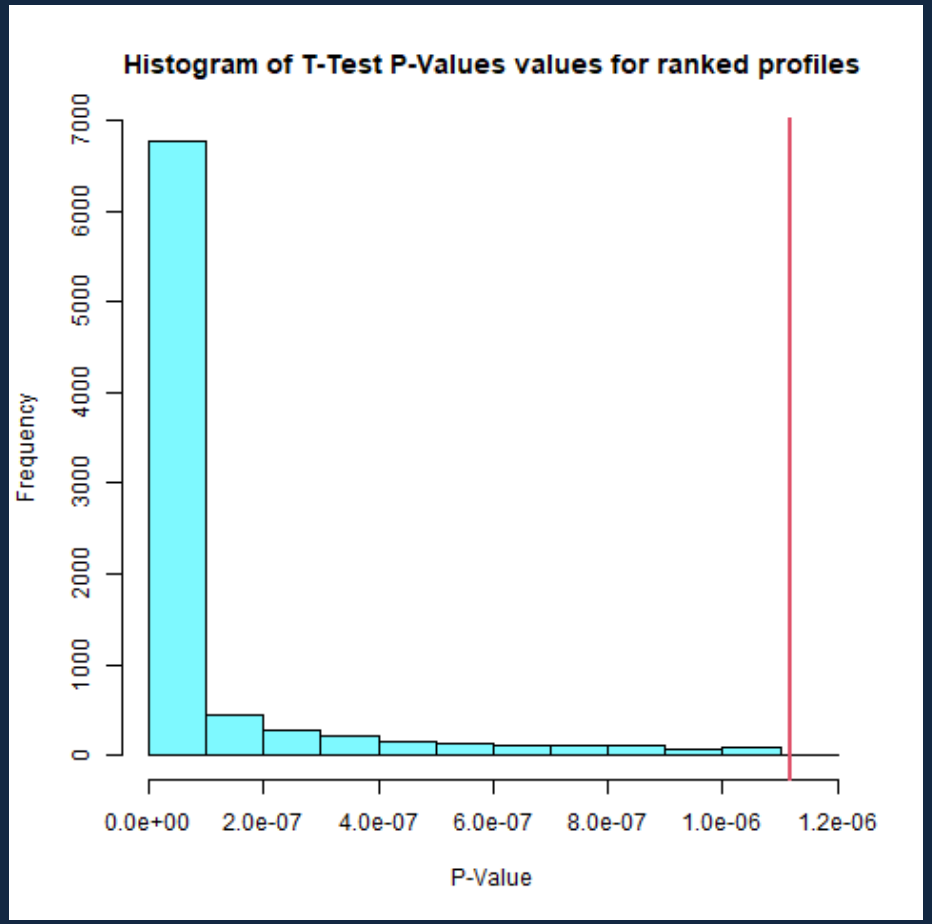
Gene distribution after Bonferroni Correction

Fold change values are normally distributed (slightly left skewed);
Symmetric, unimodal distribution; Center distribution p-value = 0.0



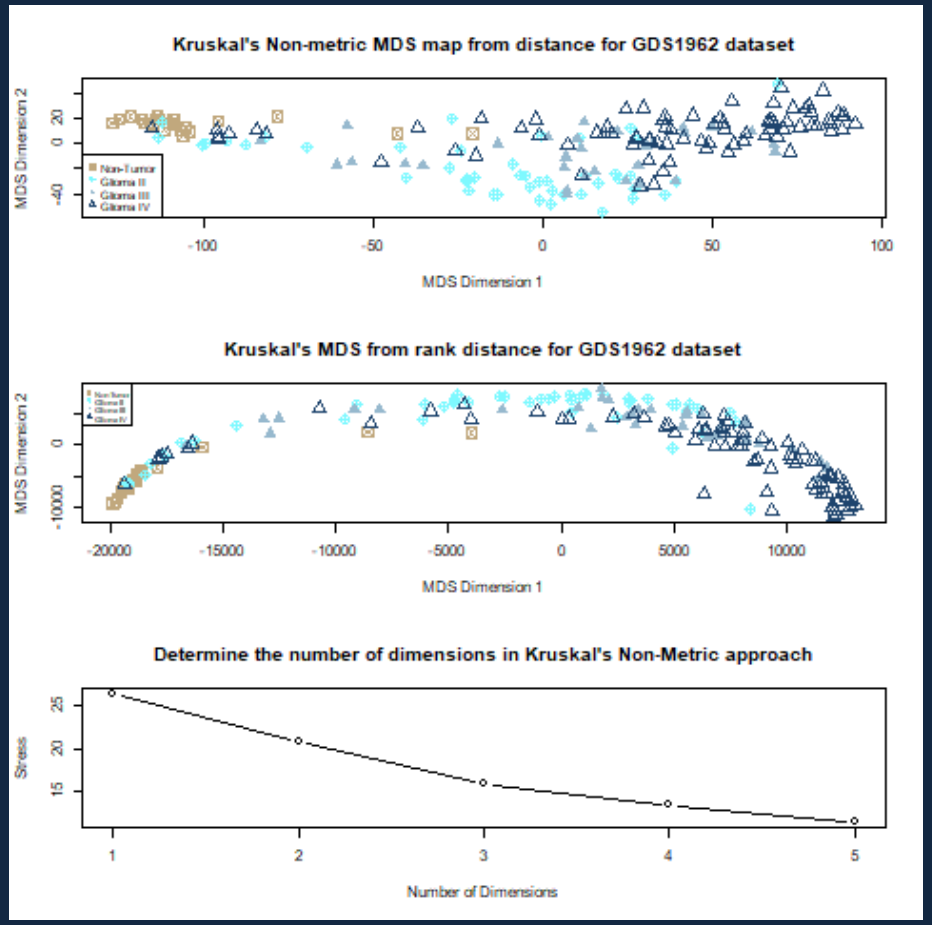


Gene distribution after SVM-RFE ranking



All the gene probes that pass the Holm-Bonferroni and Bonferroni correction methods are retained for further analysis (**n=8481**).

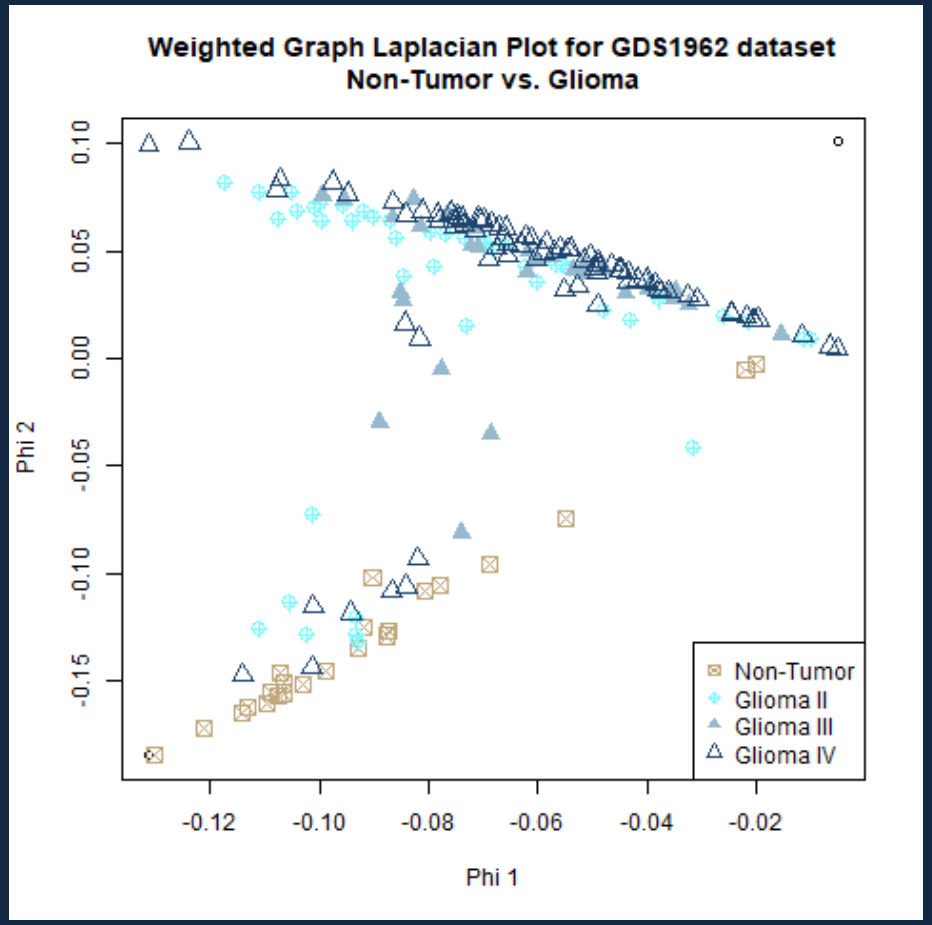
Dimensionality reduction: Kruskal non-metric MDS



The non-tumor control samples are close in proximity, while some of the glioma samples are further in distance.

Qualitatively, the biggest jump in stress happens between 2 and 3 dimensions.

Dimensionality reduction: Weighted Graph Laplacian Plot

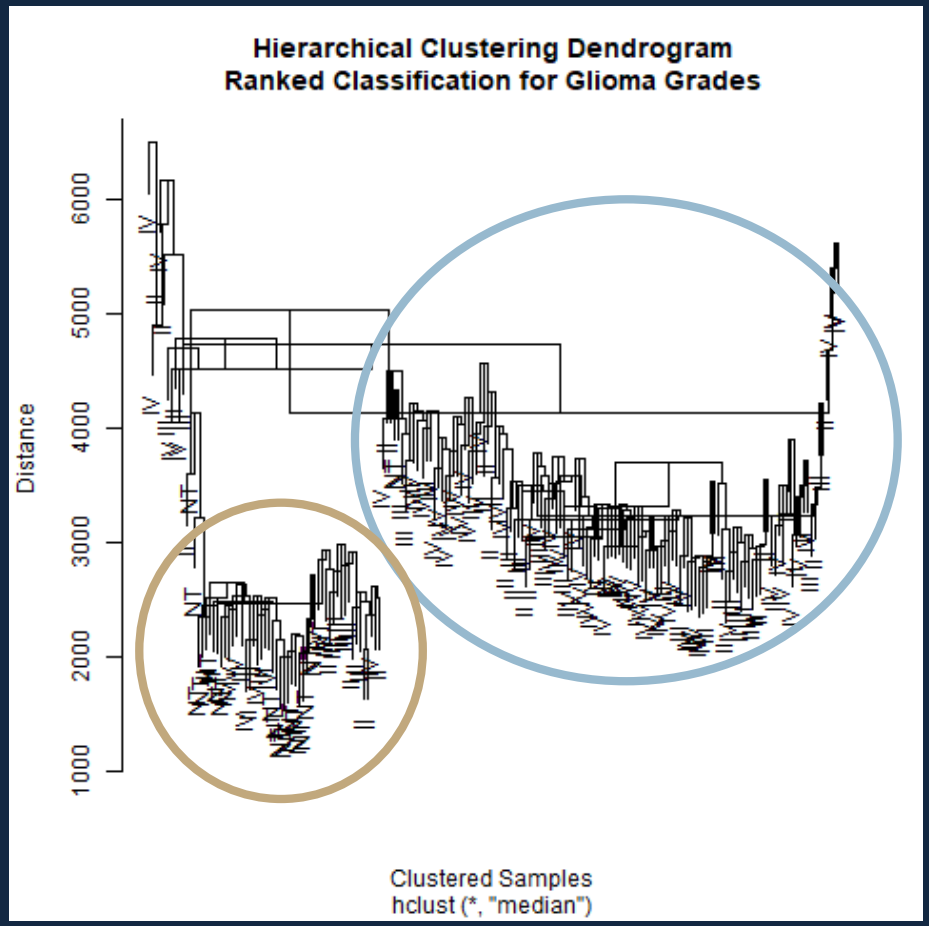


Qualitatively, the Laplacian Plot demonstrated that there are two distinct clusters – i.e., control vs glioma.

Clustering

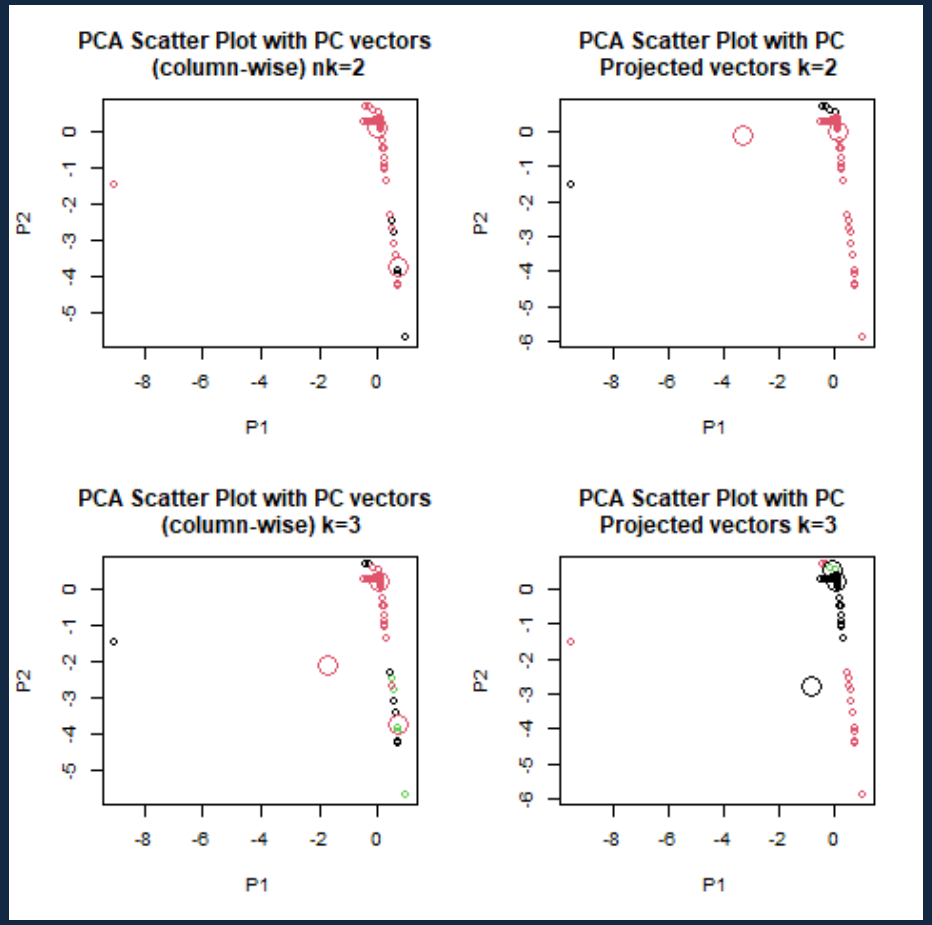
- Clustering methods:
  - Hierarchical clustering with Manhattan distance metric and median linkage method
  - K-means clustering with k=2
- Dendrogram produced two distinctive clusters. However, there are overlaps between non-tumor and glioma grade II clusters, suggesting that early glioma grade might incur a false negative error.

Clustering: Hierarchical Dendrogram



Two distinct clusters were observed in the hierarchical dendrogram.

Clustering: K-means



K-means clustering with K=2 or K=3 both resulted in two clusters

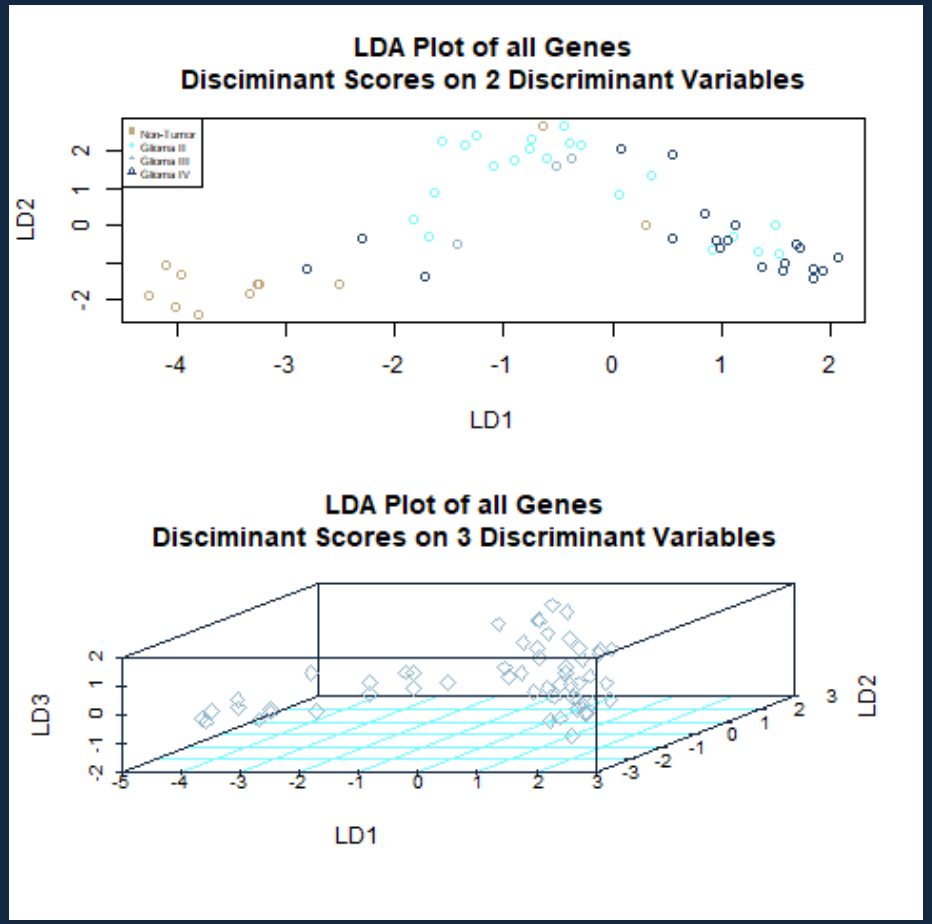
Sample Classification

- Two supervised methods
  - **Linear Discriminant Analysis**

	Control	Glioma II	Glioma III	Glioma IV
Training set	15	30	20	60
Test set	8	15	11	21

  - **Principal component analysis**
- Mismatches occurred with both LDA and PCA analysis. Support vector machine returns an overall lower misclassification rate and is therefore more reliable. There is a relatively accurate prediction of Control vs Glioma; however, the glioma grade classification is unreliable.

Classification: LDA plot



Classification: Confusion matrix -LDA

Relatively high misclassification rate

Raw LDA Confusion Matrix

	Non-Tumor	Glioma Grade II	Glioma Grade III	Glioma Grade IV
Non-Tumor	9	0	0	3
Glioma Grade II	0	13	6	1
Glioma Grade III	0	0	2	1
Glioma Grade IV	0	2	3	15

Classification rate

	Non-Tumor	Glioma Grade II	Glioma Grade III	Glioma Grade IV
Non-Tumor	0.164	0.000	0.000	0.055
Glioma Grade II	0.000	0.236	0.109	0.018
Glioma Grade III	0.000	0	0.036	0.018
Glioma Grade IV	0.000	0.036	0.055	0.273

Total misclassifications: **16**

Classification: Confusion matrix- PCA

Raw SVM Confusion Matrix

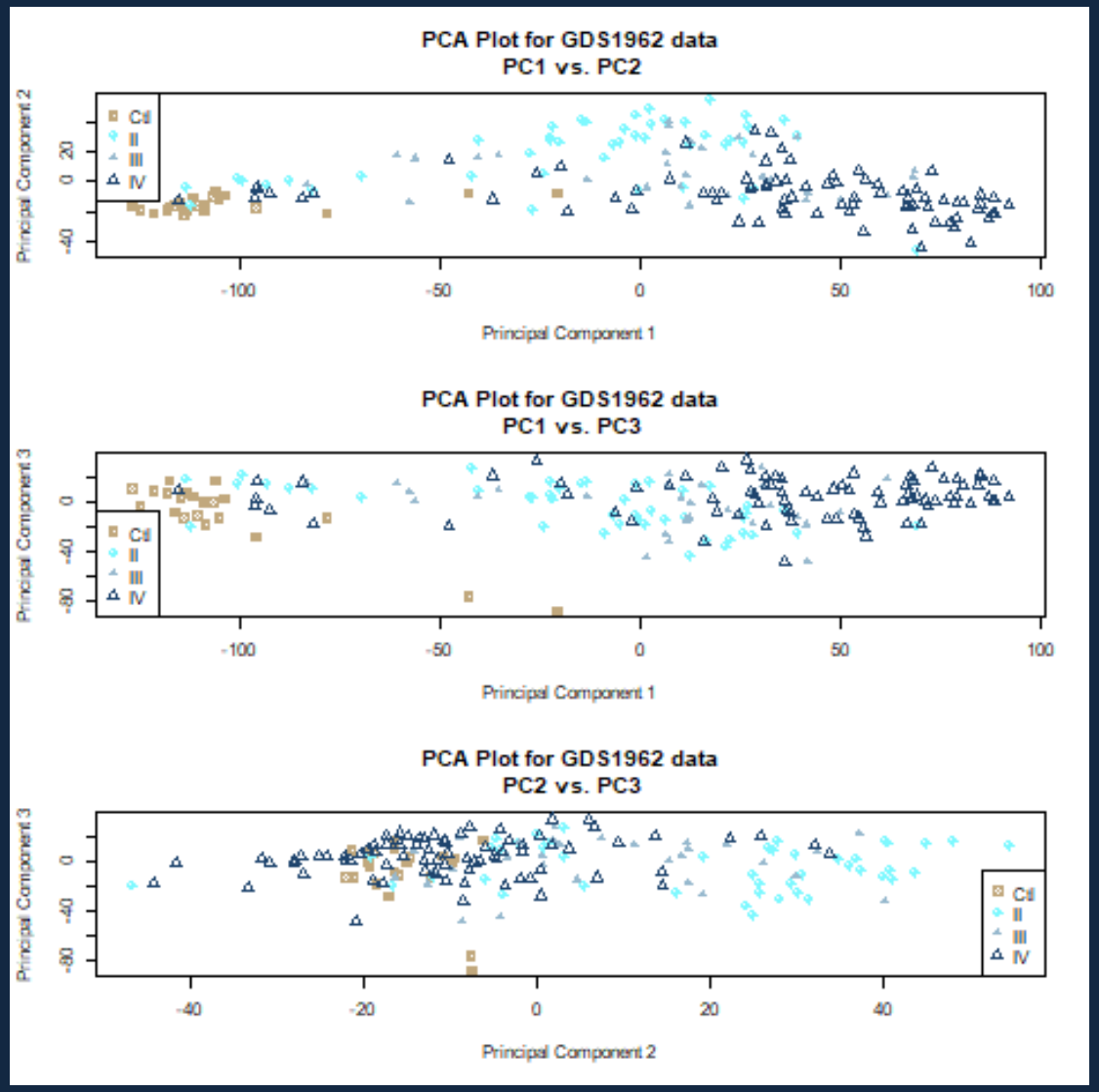
	Non-Tumor	Glioma Grade II	Glioma Grade III	Glioma Grade IV
Non-Tumor	19	2	2	0
Glioma Grade II	2	36	2	5
Glioma Grade III	0	11	5	15
Glioma Grade IV	3	7	2	69

Total misclassifications: **51**

Classification rate

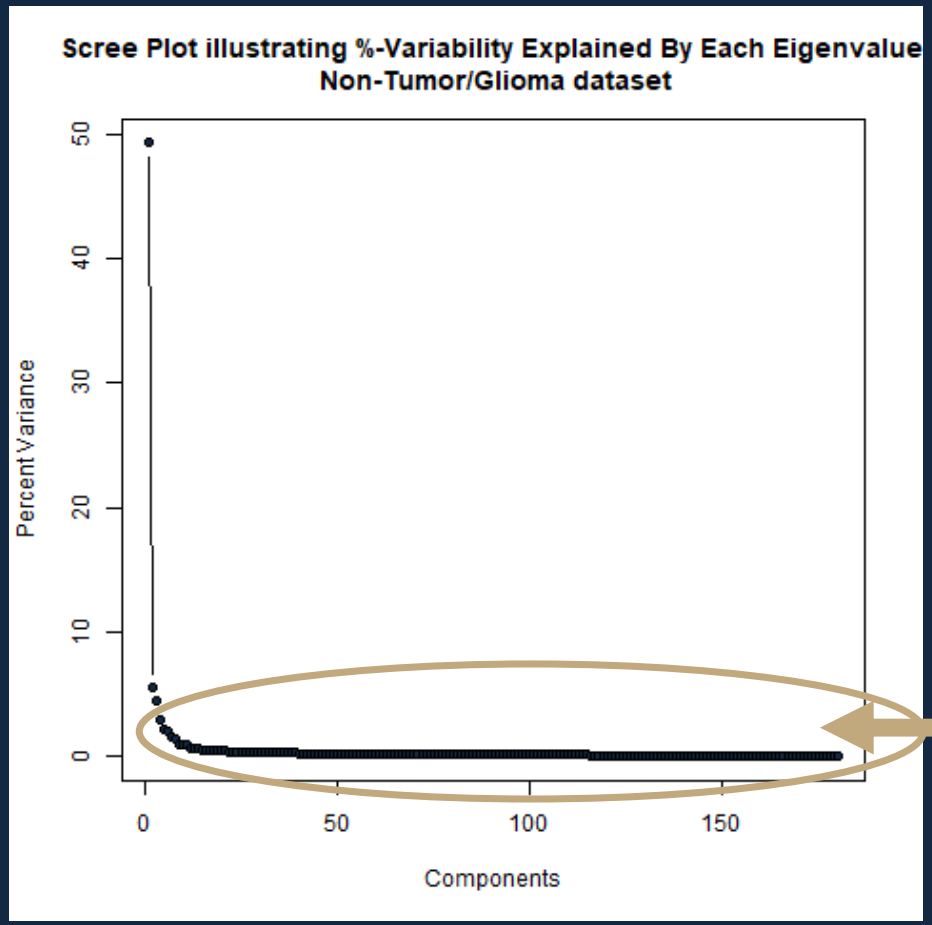
	Non-Tumor	Glioma Grade II	Glioma Grade III	Glioma Grade IV
Non-Tumor	0.106	0.011	0.011	0.000
Glioma Grade II	0.011	0.200	0.011	0.028
Glioma Grade III	0.000	0.061	0.028	0.083
Glioma Grade IV	0.017	0.039	0.011	0.383

Classification: Principal Component Analysis (PCA)



Approximately **54.74%** variability is explained using only the first two eigenvalues. The high eigenvalue potentially account for the high variance in PCA confusion matrix.

Classification: Scree plot (PCA)



The first three eigenvalues contributed to a significant level of variability, which were not removed during the gene filtering and feature selection steps. Hence, the number of factors that should be generated by the PCA analysis is **N-3 (i.e., 8478)**.

Top five discriminant genes: positive direction

Probeset	Gene	GO Biological Process	GO Molecular Function	Interpro
207035_at	SLC30A3	transport, response to zinc ion, positive regulation of transport, regulation of sequestering of zinc ion, zinc II ion transmembrane transport	protein binding, cation transmembrane transporter activity, zinc transporting ATPase activity	Cation efflux protein, Cation efflux protein transmembrane domain
222565_s_at	PRKD3	protein phosphorylation, protein kinase C-activating G-protein coupled receptor signaling pathway, peptidyl-serine phosphorylation, protein kinase D signaling	protein kinase C activity, protein binding, ATP binding, kinase activity, metal ion binding	Serine/threonine-protein kinase, Pleckstrin homology domain, Protein kinase C-like, phorbol ester/diacylglycerol binding, Pleckstrin homology-like domain, ATP binding site, Diacylglycerol/phorbol-ester binding
215020_at	NRXN3	angiogenesis, neuron cell-cell adhesion, signal transduction, neurotransmitter secretion, axon guidance, synapse assembly, learning, adult behavior, social behavior, etc.	transmembrane signaling receptor activity, calcium channel regulator activity, metal ion binding, cell adhesion molecule binding, neuroligin family protein binding	EGF-type aspartate/asparagine hydroxylation site, Epidermal growth factor-like domain, Laminin G domain, Neurexin/syndecan/glycophorin C, Concanavalin A-like lectin/glucanase, subgroup, Neurexin-1-alpha
233144_s_at	RASAL1	MAPK cascade, intracellular signal transduction, positive regulation of GTPase activity, negative regulation of Ras protein signal transduction, etc.	GTPase activator activity, phospholipid binding, metal ion binding	C2 calcium-dependent membrane targeting, Zinc finger, Btk motif, Pleckstrin homology domain, Ras GTPase-activating protein, Rho GTPase activation protein, Pleckstrin homology-like domain
228903_at	CES4A	NA	hydrolase activity, carboxylic ester hydrolase activity	Carboxylesterase, type B

Bottom 5 discriminant genes: Negative Direction

Probeset	Gene	GO Biological Process	GO Molecular Function	Interpro
224365_s_at	TIGD7	NA	DNA binding, protein binding	DDE superfamily endonuclease, CENP-B-like, HTH CenpB-type DNA-binding domain, DNA binding HTH domain, Psq-type, Homeodomain-like
207404_s_at	HTR1E	G-protein coupled receptor signaling pathway, phospholipase C-activating G-protein coupled receptor signaling pathway, serotonin receptor signaling pathway, chemical synaptic transmission,	G-protein coupled serotonin receptor activity, protein binding, neurotransmitter receptor activity	G protein-coupled receptor, rhodopsin-like, 5-Hydroxytryptamine receptor family5-Hydroxytryptamine 1E receptor
229626_at	CCDC184	NA	Protein binding	NA
229845_at	MAPKAP1	substantia nigra development, establishment or maintenance of actin cytoskeleton polarity, stress-activated protein kinase signaling cascade, T cell costimulation, TORC2 signaling, negative regulation of Ras protein signal transduction	protein binding, phosphatidylinositol-4,5-bisphosphate binding, phosphatidylinositol-3,4,5-trisphosphate binding, kinase activity, Ras GTPase binding, protein kinase binding, etc.	Stress-activated map kinase interacting 1
225104_at	ZNF598	NA	zinc ion binding, poly(A) RNA binding, metal ion binding	Zinc finger, RING-type, Zinc finger, C2H2, Zinc finger, RING/FYVE/PHD-type, Zinc finger, C2H2-like

Conclusions

- In contrary to the findings by Sun et al, we did not observe significant increase in SCF gene expression in glioma grade III and IV samples in the dataset. Furthermore, all four SCF (KITLG) gene probes were eliminated after the gene filtering and feature selection processes. The likelihood behind this discrepancy is because the control dataset is derived from brain tissues of epilepsy patients, which likely have a high baseline level of pro-angiogenic cytokines [2, 3].
- **8481** gene probes are differentially expressed across glioma grades, none of which are the KITLG gene probes. 436 gene probes were removed in the gene filtering step ($0 < \log_2(\text{expression}) \leq 3$), and 36201 gene probes were also removed in the subsequent feature selection process.
- The dataset can be divided into **two clusters**, based on hierarchical dendrogram and K-means clustering methods.
- While baseline gene expression levels in non-tumor samples can be confidently predicted, the gene expression across glioma grades are relatively ambiguous.
- The top 5 positive discriminant genes are **SLC30A3, PRKD3, NRXN3, RASAL1, and CES4A**. The top 5 **negative** discriminant genes are **TIGD7, HTR1E, CCDC184, MAPKAP1, and CES4A**.

References

1. Sun L, Hui AM, Su Q, Vortmeyer A, Kotliarov Y, Pastorino S, Passaniti A, Menon J, Walling J, Bailey R, Rosenblum M, Mikkelsen T, Fine HA. Neuronal and glioma-derived stem cell factor induces angiogenesis within the brain. Cancer Cell. 2006 Apr;9(4):287-300. doi: 10.1016/j.ccr.2006.03.003. PMID: 16616334.
2. Marchi, N., & Lerner-Natoli, M. (2013). Cerebrovascular remodeling and epilepsy. *The Neuroscientist : a review journal bringing neurobiology, neurology and psychiatry*, *19*(3), 304–312. https://doi.org/10.1177/1073858412462747
3. Rigau V, Morin M, Rousset MC, de Bock F, Lebrun A, Coubes P, Picot MC, Baldy-Moulinier M, Bockaert J, Crespel A, Lerner-Natoli M. Angiogenesis is associated with blood-brain barrier permeability in temporal lobe epilepsy. Brain. 2007 Jul;130(Pt 7):1942-56. doi: 10.1093/brain/awm118. Epub 2007 May 28. PMID: 17533168.