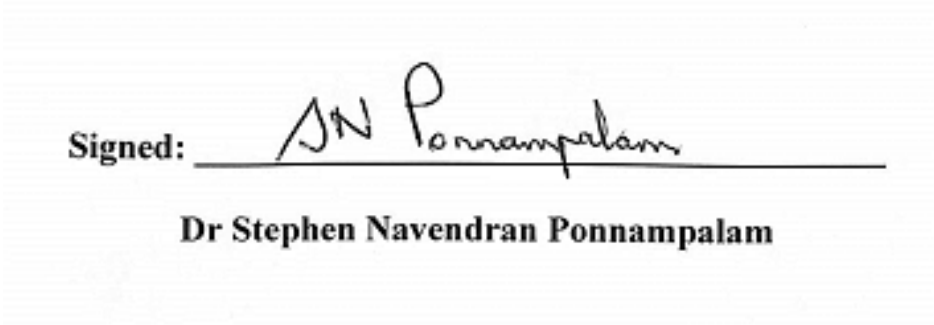
A Blood-Based Gene Expression and Signalling Pathway Analysis to Differentiate Between High and Low Grade Gliomas

Dr Stephen Navendran Ponnampalam

Submitted to the Faculty of Graduate Studies, University College London (UCL) for fulfillment of the requirements for award of the degree of Doctor of Medicine (MD)

I, Dr Stephen Navendran Ponnampalam, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



ABSTRACT

Introduction:
Brain tumours are the 17th most common cancer worldwide. Gliomas are the most common of the primary brain tumours and are highly malignant.

Objectives:
(a) To undertake gene expression profiling of the blood of glioma patients to determine key genetic components of signalling pathways
(b) To develop a panel of genes that could be used as a potential blood-based biomarker to differentiate between high and low grade gliomas, non-glioma and control samples.

Methods:
Blood samples were obtained from glioma patients, non-glioma and control subjects. Ten samples each were obtained from patients with high and low grade tumours respectively, ten samples from non-glioma patients and twenty samples from control subjects. Total RNA was isolated from each sample after which first and second strand synthesis was performed. The resulting cRNA was then hybridized with the Agilent Whole Human Genome (4x44K) microarray chip according to the manufacturer's instructions. Universal Human Reference RNA and samples were labeled with Cy3 CTP and Cy5 CTP respectively. Microarray data were analyzed by Agilent Gene Spring 12.1V software using stringent criteria which included at least a 2-fold difference in gene expression between samples. Statistical analysis was performed using the unpaired student T-test with a p-value < 0.01. Pathway enrichment was also performed with key genes within these pathways selected for validation with ddPCR.

Results:
The gene expression profiling indicated that were a substantial number of genes that were differentially expressed with more than a 2-fold change (FDR corrected value < 0.01) between each of the four different conditions. We selected key genes within significant pathways that were analyzed through pathway enrichment. These key genes included regulators of cell proliferation, transcription factors, cytokines and tumour suppressor genes.

Conclusion:
In this study, we have shown that key genes involved in significant and well established pathways, could possibly be used as a potential blood-based biomarker to differentiate between high and low grade gliomas, non-gliomas and control samples.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor, Professor John Hardy for accepting me to do the Doctor of Medicine (MD) degree under him. Many thanks to my co-supervisor, Dr Mina Ryten, for all the invaluable comments and suggestions that she has given over the course of my MD studies.

I would also like to thank the following individuals who have been a great help to me especially in the area of bioinformatics:

(1) Ms Supriya Karkra and Mr Ankit Potla from Strand Life Sciences, Bangalore, India who helped me with the GeneSpring analysis

(2) Dr Asif Khan, Dr Lloyd Low and Ms Atiqah Azhar from the Centre for Bioinformatics, Perdana University School of Medicine, Malaysia who helped me with formatting of the figures for the thesis and also pathway and statistical analysis.

Finally, I would like to thank the current Prime Minister of Malaysia, The Honourable Dato' Sri Haji Mohammad Najib bin Tun Haji Abdul Razak, who personally awarded me a scholarship for the MD degree through the Yayasan Rakyat 1 Malaysia (YR1M) foundation.

LIST OF ABBREVIATIONS

AQP4	Aquaporin 4
Bcl2	B-cell lymphoma 2
Bcl2L11	B-cell lymphoma 2L11
Bcl2A1	B-cell lymphoma 2A1
BCR-ABL	Breakpoint cluster region, Abelson Leukaemia Proto-Oncogene
B-H	Benjamini - Hochberg
BRT	Blood RNA tube
C	Control
CCNU	1-(2-Chloroethyl)-3-Cycohexyl-1-Nitrosourea (Lomustine)
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	Complementary deoxyribonucleic acid
CML	Chronic myeloid leukaemia
cRNA	Complementary ribonucleic acid
CML	Chronic Myeloid Leukemia
CT	Computed tomography

CTC	Circulating tumour cells
ddPCR	Droplet Digital Polymerase Chain Reaction
DTC	Disseminated Tumour Cells
EGFR	Epidermal Growth Factor Receptor
EGFRvIII	Epidermal Growth Factor Receptor Variant III
EMT	Epithelial-to-Mesenchymal transition
EpCAM	Epithelial cell adhesion molecule
FOS	FBJ Murine Osteosarcoma Viral Oncogene Homolog
FXR	Farnesoid X receptor
GBM	Glioblastoma Multiforme
GIST	Gastrointestinal Stromal Tumour
GTPases	Guanosine Triphosphate Hydrolase
GUSB	Beta-glucuronidase
HER2	Human Epidermal Growth Factor Receptor 2
HG	High Grade Gliomas
HGF	Hepatocyte Growth Factor
HIFa	Hypoxia Inducible Factor Alpha

HOXB2	Homeobox B2
H&E	Haematoxylin and Eosin
hTERT	Human Telomerase
HuPO	Human Acidic Ribosomal Protein
IDH1/2	Isocitrate Dehydrogenase 1 & 2
IL4	Interleukin 4
IL6	Interleukin 6
IL8	Interleukin 8
IL12RB1	Interleukin 12 Receptor, Beta 1
IPA	Ingenuity Pathway Analysis
K-RAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
KPS	Karnofsky Performance Score
LG	Low Grade Gliomas
LOH	Loss Of Heterozygosity
LXR	Liver X Receptors
MAPK12	Mitogen-Activated Protein Kinase 12
MAP3K8	Mitogen-Activated Protein Kinase 8

MET	Mesenchymal-to-Epithelial Transition
MGMT	O-6-Methylguanine-DNA Methyltransferase
MTC	Multiple Testing Correction
miR-17	MicroRNA 17
miR-184	MicroRNA 184
MMP9	Matrix Metallopeptidase 9
MREC	Medical Research Ethics Committee
MRI	Magnetic Resonance Imaging
NF-kB	Nuclear Factor-Kappa B
NG	Non-Gliomas
PCA	Principle Component Analysis
PDGFRa	Platelet-Derived Growth Factor Receptor Alpha
PEDF	Pigment Epithelium-Derived Factor
PKC	Protein kinase C
PTEN	Phosphatase And Tensin Homolog
RB1	Retinoblastoma 1
RhoG	Ras Homolog Family Member G

ROC	Receiver Operating Characteristic
RXR	Retinoid X Receptor
sIL4RA	Soluble IL-4 Receptor alpha
SOS1	Son of Sevenless Homolog 1
SOX8	SRY (Sex Determining Region Y)-Box 8
TBP	TATA Binding Protein
Tec kinase	Tec Protein Tyrosine Kinase
TP53	Tumor Protein P53
TNF	Tumor Necrosis Factor
VEGF	Vascular Endothelial Growth Factor
VEGFA	Vascular Endothelial Growth Factor A
WHO	World Health Organization
ZNF 649	Zinc Finger Protien 649
ZNF 205	Zinc Finger Protein 205

CHAPTER 1: INTRODUCTION

Cancer is the leading cause of morbidity and mortality worldwide with the number of cases expected to increase by 70% over the next 2 decades (1). Brain and central nervous system tumours are ranked 17th in incidence among all cancers worldwide, being the 13th and 15th most common tumour in men and women respectively (2). Cancer is the 2nd leading cause of death in the paediatric age group (3), with brain and other nervous system tumours ranked 2nd in incidence after leukaemias (4).

Brain tumours can be either primary or secondary. Gliomas are the most common of the primary brain tumours consisting mainly of oligodendroglioma and astrocytoma with a small number of mixed oligoastrocytoma. Currently, there is a simple World Health Organization (WHO) classification for these tumours. Oligodendrogliomas are divided into low grade (Grade II) and anaplastic (Grade III) tumours. For astrocytic tumours, the WHO classification system divides this tumour into pilocytic astrocytoma (Grade I), low grade diffuse astrocytoma (Grade II), anaplastic astrocytoma (Grade III) and glioblastoma multiforme (GBM) and its variants (Grade IV). Oligoastrocytomas are also classified into low grade (Grade II) and anaplastic (Grade III) tumours (5).

Low grade oligodendrogliomas grow slowly whereas anaplastic oligodendrogliomas grow more aggressively. These tumours are predominantly found in the cortex but may invade into other structures of the brain such as the corpus callosum, the ependyma and deep into the brainstem (6).

Oligodendrogliomas have their own unique set of genetic alterations compared to other gliomas. The most common genetic alteration found in oligodendrogliomas is loss of heterozygosity (LOH) on chromosomes 1p and 19q (7,8). Oligodendrogliomas with LOH at 1p are known to be more chemosensitive than tumours with retention of this chromosome. In addition, combined allelic losses at 1p and 19q are associated with longer recurrence-free survival after chemotherapy (9). Loss of chromosome 1p is also a predictor of longer recurrence-free survival in patients receiving radiotherapy with or without chemotherapy (10,11).

Besides LOH on chromosomes 1p and 19q, almost all oligodendrogliomas have mutations in the *IDH1/2* genes (12). In genome sequencing studies, *IDH1* appears to be a persistent mutation found in all patients with low grade glioma suggesting that this mutation is an early event and precedes the LOH on chromosomes 1p and 19q during glioma development and progression (13). In addition, patients with *IDH1/2* mutations are younger and have a more favourable prognosis compared to those without these mutations (14).

Astrocytic tumours also show many different genetic alterations. Some of the early genetic changes in low grade astrocytomas are mutations in *IDH1* and *TP53* (15,16). In addition, there are other genomic alterations that occur as a diffuse low grade glioma progresses to a high grade glioma which includes the anaplastic astrocytoma and secondary GBM. These changes include disruption of the *RB1* pathway (17), deletion of chromosome 9p including *CDKN2A* (18) and the presence of microRNAs miR-17 and miR-184 (19). On the other hand, primary GBM occurs *de novo*, has a short clinical

history and possesses unique genetic alterations such as *EGFR* amplification (20), *EGFRvIII* mutation (21), loss of the tumour suppressor gene *PTEN* (22) and deletion of chromosome 10 (23).

IDH1/2 mutations and *MGMT* promoter methylation deserve further mention as they have significant predictive and prognostic value in gliomas. *IDH1* mutations are mainly found in grade II and III gliomas and secondary GBM (12). *IDH2* mutations are commonly found in low grade oligodendrogliomas (24). Glioblastomas that have mutations in the *IDH1* gene generally show less oedema and necrosis with reduced disruption of the blood brain barrier (25). *VEGF* is responsible for vascular permeability in GBM resulting in oedema, increased vascularity and tumour enhancement. *HIFα* regulates *VEGF* expression and is downregulated in *IDH1/2* mutated cells resulting in tumours that have less oedema and decreased vascularity (26). In addition, tumours that have the *IDH1/2* mutations are more sensitive to oxidative stress induced by radiation (27) and chemotherapy (28,29) resulting in cell death. These phenomena result in better prognosis for glioma patients with *IDH1/2* mutations (14).

MGMT is a DNA repair enzyme that reverses the effects of alkylating chemotherapeutic agents such as temozolomide. Epigenetic silencing of the *MGMT* promoter by methylation, sensitizes GBM cells to temozolomide. The phenomenon is present in about half of all GBM patients resulting in marked survival when treated with temozolomide and radiation therapy (30).

The treatment for gliomas involves surgical resection, radiation and chemotherapy. Maximal safe surgical resection is recommended for all patients with

gliomas as it results in increased survival (31). Advanced neurosurgical equipment and techniques such as intraoperative MRI (32) and brain mapping (33) together with the use of 5-aminolevulinic acid which causes fluorescence of glioma tissue (34), has increased safety and resulted in more extensive resections of the tumour. Despite all these advances, it is still not possible to achieve complete resection of the tumour as gliomas are diffuse and infiltrate into surrounding brain tissue. As such, the majority of glioma patients have to undergo adjuvant therapy in the form of radiotherapy, chemotherapy or both. In addition, some patients are also offered more recently available treatments such as targeted immunotherapy.

Radiation in the form of external beam irradiation is given as an adjuvant therapy after surgical resection for glioma patients. Radiation therapy has been shown to prolong survival compared to surgery alone or supportive care only (35), especially when residual tumour is still evident (36). In addition to radiation therapy, chemotherapy is also used to treat malignant gliomas. The common chemotherapeutic agents used are procarbazine, CCNU and vincristine (PCV) for the treatment of oligodendroglial tumours (37,38) and temozolomide for the treatment of astrocytic tumours including GBM (39). Amongst targeted therapies for malignant gliomas, bevacizumab, a monoclonal antibody against vascular endothelial growth factor receptor (*VEGFR)* has shown promising activity (40).

Genome-wide characterization and classification of tumours as well as the development of specific inhibitors to treat such tumours, has given rise to the age of personalized medicine. Bevacizumab, besides its role in the treatment of malignant gliomas, is also effective in treating colorectal cancer patients whose tumours harbour the

wild-type *K-RAS* oncogene (41). Approximately 25 percent of breast cancers have amplification of the epidermal growth factor receptor 2 gene (*HER2*) which is involved in the growth and proliferation of both normal and malignant cells and a strong driver of tumorigenesis (42,43). This discovery led to the development of a recombinant humanized monoclonal antibody, trastuzumab, which inhibits the *HER2* receptor through antibody mediated cellular cytotoxocity (44,45). Multiple clinical trials have shown the superiority of using trastuzumab as an adjuvant and neoadjuvant agent in the treatment of early stage and metastatic breast cancers. In early stage and locally advanced breast cancer, the addition of trastuzumab to chemotherapy significantly improved both disease free survival and overall survival (46-48) whilst in metastatic breast cancer this led to higher response rates, prolonged time to progression and increased overall survival (44).

Non-small cell lung cancers, especially adenocarcinomas, have mutations in the intracellular tyrosine kinase domains of the epidermal growth factor receptor *(EGFR)* which results in activation of several signal transduction pathways promoting cell growth and proliferation (49). In addition, amplification of *EGFR* also promotes malignant transformation. Together, mutations in the tyrosine kinase domain and amplification of EGFR are found in 43-89 percent of cases of non-small cell carcinoma of the lung (50). 2 oral tyrosine kinase inhibitors, gefitinib and erlotinib, have been used successfully to treat mutated *EGFR* advanced lung cancer with a response rate of 65-90 percent (51).

The classical case of personalized medicine and targeted therapy can be found in the treatment of the Philadelphia positive chromosome in chronic myeloid leukaemia (CML). In this myeloproliferative disorder, there is a chromosomal rearrangement

between the long arms of chromosomes 9 and 22, forming the Philadelphia chromosome (52). This translocation gives rise to a fusion protein, *BCR-ABL*, demonstrating constitutive tyrosine kinase activity (52) resulting in uncontrolled cell growth and proliferation (53). Imatinib mesylate (Gleevec), a potent tyrosine kinase inhibitor which can be taken orally, is highly effective in treating CML (54) and induces a high haematologic and cytogenic response (55), in addition to prolonging survival (56). Imatinib mesylate is also used as adjuvant treatment in localized primary gastrointestinal stromal tumour (GIST) where it prolongs recurrence free survival after complete resection of the primary tumour (57). In GIST, imatinib mesylate inhibits activation of the *KIT* proto-oncogene and the platelet-derived gowth factor receptor alpha (*PDGFRα*) (58,59). It is also used to treat metastatic GIST where it achieves partial response and stable disease in up to 80 percent of patients with a median survival of 5 years (60). Thus, personalized medicine has resulted in improved treatments and prolonged survival for patients with different types of cancer.

In the case of gliomas however, despite advances in surgery, radiation and chemotherapy together with more recently available therapies such as molecularly targeted therapies, prognosis is generally poor. The median survival for patients with malignant gliomas is less than 15 months with GBM patients having the worst prognosis with less than 5% surviving after 5 years (61).

One of the reasons that cancers are detected at a late stage is because many tumours do not have symptoms until the disease has spread. There is also poor compliance in adhering to standard screening procedures, as some of these tests are

unpleasant in nature and have associated risks and complications (62,64). The current methods available for the detection of gliomas are computed tomography (CT) scan and magnetic resonance imaging (MRI) of the brain. However, the definitive diagnosis is by stereotactic guided biopsy of the tumour sample which is technically demanding and has its risks such as causing hemorrhage, infection, seizures and neurological deficits but is however considered relatively acceptable (64-66). Therefore, the development of a simple, non-invasive blood test which involves RNA profiling in whole blood, can be used as an addition to the more traditional methods of cancer screening and detection (67).

The inspiration for whole-blood, transcriptome profiling in the context of gliomas originates from the "sentinel" principle (67). Inherent in this principle is the fact that blood is in intimate contact and interacts with all human tissues including cancerous tissue. Blood is considered a connective tissue and is a transporter for various substances such as oxygen, nutrients, cells of the immune system including B cells, T cells, dendritic and natural killer cells, cytokines, growth factors and hormones (68). In addition, blood cells are affected in many disease processes such as haematological malignancies, solid tumours, asthma, autoimmune diseases such as rheumatoid arthritis to common chronic illnesses such as hypertension, diabetes and cardiovascular disease (69-72).

Peripheral blood cells have the ability to respond to changes that affect the physiology, microenvironment and systems biology of the human body. Perturbations or disturbances in the homeostasis of the system can also be subtly detected by peripheral blood cells (68,73). Thus blood, being easily accessible could serve as a molecular gene

expression profile reflecting changes that occur within tissues of the human body (67). The term "bloodomics" has thus been coined to reflect this function of blood in regulation of gene expression and in the molecular profiling of human diseases (68).

One of the earliest models where the sentinel principle has been studied is colorectal carcinoma where a 5 and 7-gene biomarker panel has been developed to assess the current relative risk of patients developing this cancer in Canada and Malaysia (74-76). Molecular gene profiling of the blood transcriptome has also been studied in other diseases including neurological disorders such as schizophrenia and bipolar disorders,. chronic fatigue syndrome, tuberous sclerosis complex 2, neurofibromatosis type 1, Down's syndrome, epilepsy, Tourette syndrome, ischemic stroke, migraine, Huntington's and Alzheimer's diseases (77-84).

Genome wide expression profiling of human blood in Huntington's disease resulted in a 12 gene biomarker panel that was clearly able to distinguish between normal controls and those with Huntington's disease. In addition, the genes in this panel showed varying expression in disease progression from early presymptomatic to late presymptomatic and finally to the symptomatic stage of Huntington's disease (83). In acute ischaemic stroke, blood expression profiling has identified 18 genes that are upregulated within 24 hours of the event. These genes were found to be upregulated in 67, 87 and 100 percent of patients after 3, 5 and 24 hours respectively (85). Some of the genes expressed after 3 hours are involved in the inflammatory process such as *MMP9* and also participate in the degradation of the extracellular matrix and breakdown of the blood brain barrier. *MMP9* is mainly secreted by neutrophils (86) and levels correlate

with infarct volume and hemorrhagic complications after thrombolytic therapy (87-89). S100 proteins are also released during acute ischaemia (90) bind to the vascular endothelium and induce a thrombotic state (91). These early changes in the expression levels of certain genes in blood, can be exploited as a diagnostic and predictive tool for acute stroke. Blood expression profiling in migraine has been able to distinguish between migraine and chronic migraine with a group of immediate early genes such as *c-fos* and *cox-2* being upregulated in migraine while specific mitochondrial genes were upregulated in chronic migraine (92).

Epilepsy in children treated with carbamezepine or valproic acid showed distinct patterns of expression in blood. In addition, patients treated with valproic acid formed distinct subclusters of those who were responsive versus refractory to treatment (81). Blood based gene expression has also been studied in schizophrenia and bipolar disorder. A set of putative biomarker genes to differentiate between schizophrenia and bipolar disorder has been validated with an accuracy of 95-97 percent by receiver operating characteristic (ROC) curve analysis (77). As can be seen, personalized medicine is revolutionizing the field of neurology including diseases with a pure brain pathology.

However, results obtained from gene expression profiling in blood have to be interpreted with caution. In colorectal carcinoma, the 7-gene biomarker panel assesses the current relative risk of a person developing the disease compared to the general population according to a scale (74). The panel of genes only indicates the probability of developing the disease and there is some overlap between high risk and low risk populations (75).

In Huntington's disease (HD), 4 of the 12 genes in the biomarker panel were upregulated in the Affymetrix platform but downregulated in the Amersham Biosciences platform (83). The reason for the discrepancy in results was most likely due to the probes on the different platforms hybridizing to different isoforms of the gene (83). This is clearly one of the drawbacks of gene expression profiling in blood unless custom made arrays are used where probes are specifically designed to detect all relevant exons.

Although gene expression profiling in blood has been undertaken in ischaemic stroke patients (85), the results need to be evaluated further. This is because all the patients in the study were treated with either tissue plasminogen activator (tPA) or tPA and eptifibatide. Thus, the increase in the number of genes that were upregulated after 3 hours could have been partly due to a direct effect of these medications on gene expression in blood.

Similarly for the psychiatric disorders, schizophrenia and bipolar disorder, the results of the studies have to be viewed critically, as all patients were receiving medications (77). It is likely that these medications, including antipsychotics and mood stabilizers, alter the gene expression profile in blood directly and therefore account for the differences seen. In addition the sample size was small and consisted of 30 patients with schizophrenia, 16 patients with bipolar disorder and 28 healthy control subjects (77) and has not been replicated in the literature to date.

Nonetheless given the accessibility of blood and the totality of available data on "bloodomics", blood-based profiling to differentiate between gliomas remains a promising approach. Furthermore, in considering cancers, there are compelling biological

reasons to believe that this may be a useful approach. It is known that during development of a tumour, substances are secreted by the tumour into the bloodstream and as a systemic response, there are subtle alterations in the level of expression of genes within peripheral blood cells in order to maintain homeostasis or as a reaction to the disease entity itself (67). Disruption of the blood-brain barrier is due to loss of substances such as the tight junction proteins claudin-1 and claudin-3, decrease in polarity of glioma cells, loss of the molecule agrin and upregulation of the aqueous channel protein, aquaporin 4 (AQP4) resulting in brain oedema formation (93-99). Since blood-brain barrier disruption occurs in brain tumours (100,101), substances that play a role in both homeostasis and tumourigenesis are likely to be secreted into the bloodstream under such conditions and may give a molecular profile signature.

In addition, cells may dislodge from the tumour and enter the peripheral circulation as circulating tumour cells (CTCs). These CTCs then colonise a distant tissue or organ and begin to form a new tumour mass. Although most CTCs do not survive in the circulation, a subset of cells known as disseminated tumour cells (DTCs) that have cancer stem cell properties are able to survive. These DTCs initially undergo an epithelial-to-mesenchymal transition (EMT), enter the circulation and then re-express their epithelial properties through the process of mesenchymal-to-epithelial transition (MET). They are then able to invade a distant tissue or organ site and form tumour cell clusters known as micrometastasis (102).

Since CTCs are found in extremely low levels in the circulation (less than 5 cells per 10mL of blood)(103), identification and detection of these cells require analytical

methods that are highly sensitive and specific combined with enrichment procedures. Among the procedures used for enrichment of CTCs are density gradient centrifugation, immunomagnetic procedures and filtration (102). Immunomagnetic procedures generally involve antibodies directed against tumour-associated antigens especially the epithelial cell adhesion molecule (EpCAM)(104) for positive selection or the common leukocyte antigen CD45 for negative selection(102).

EpCAM is found only on epithelial cells and therefore DTCs may be missed as they do not express this marker (105). CTCs have been extensively studied in tumours of epithelial origin such as breast, colon and prostate cancer including global gene expression profiling of CTCs of these tumours (106). CTCs have also been detected in patients with gliomas particularly those patients with high grade tumours such as GBM. However, glioma cells do not express EpCAM but instead express Nestin, both in, *in vitro* and *in vivo* studies. This suggests that Nestin could be used as a suitable marker for the detection of circulating glioma cells. In addition, glioma cells also express high levels of human telomerase (hTERT) which co-localizes with Nestin *in vivo* (107).

Detection of CTCs in glioma patients also has clinical utility. Many glioma patients who have undergo chemoradiaton therapy have necrosis of brain tissue which shows a persistence of signal on MRI. Clinicians are unable to accurately determine whether such signals are due to disease progression or pseudoprogression as current imaging (CT and MRI) cannot distinguish between them (107). However, the identification of glioma-derived CTCs in the circulation of such patients posttreatment is

prognostic, with a reduction in CTCs indicating treatment response and an increase in CTCs indicating disease progression (107).

As mentioned earlier, the prognosis for gliomas remains poor and has not changed much since the 1980s, despite advances in neuroimaging, neurosurgery, chemotherapy, radiotherapy and targeted therapy. It has been shown that patient survival depends on several clinical and biological factors such as tumour size and location (108), age at presentation (109), treatment (110,111), Karnofsky performance score (KPS)(112), histology (113), and molecular genetics of the tumour (114).

In this study, we have extrapolated the fascinating theory of the sentinel principle to the development of adult gliomas and to determine if such expression profiling in blood could be used to distinguish between high and low grade gliomas, non-gliomas and control samples. The justification for this study is that such profiling will help not only in the stratification of gliomas, but also in the early detection of tumours when they are far more amenable to complete surgical resection, thus improving prognosis and survival of the patient.

WHO Grade	Astrocytic Tumours	Oligodendroglial Tumours	Mixed Gliomas
I	Pilocytic Astrocytoma	–	–
II	Low Grade Diffuse Astrocytoma	Oligodendroglioma	Oligoastrocytoma
III	Anaplastic Astrocytoma	Anaplastic Oligodendroglioma	Anaplastic Oligoastrocytoma
IV	Glioblastoma Multiforme	–	–

Table 1: World Health Organization (WHO) classification of gliomas

CHAPTER 2: MATERIALS AND METHODS

2.1 Clinical patient data

Upon admission to the hospital, demographic data and a brief clinical history was elicited from 30 of the 50 patients. The demographic data included the age and sex of the patient and the state in which the patient was domiciled. The 30 patients comprised of 10 high grade gliomas (HG), 10 low grade gliomas (LG) (Table 2) and 10 non-glioma (NG) cases (Table 3). The remaining 20 subjects were normal, healthy controls (C) (Table 4) Due to the limited availability of samples, we were unable to fully age and sex match patients. The summary statistics for all 4 groups of patients is shown in Table 5.

The clinical history centered on whether there was a past history of brain tumours or other cancers in the family or the patient. In addition, patients were asked about any neurological symptoms they experienced prior to admission to the hospital, such as headache, nausea or vomiting, seizure, personality changes, slurring of speech, changes in vision or sensation and motor changes (Table 6). This data was documented on a patient consent form.

Informed consent was obtained prior to blood taking and brain tumour removal from the patient during surgery. The consent was obtained from the patient or a close relative the day before surgery. After obtaining consent, blood was immediately drawn from the patient. Surgery was performed the next day, typically within 12-24 hours of obtaining consent and drawing blood from the patient. The patient's consent was also obtained in order to perform further analysis and studies on the extracted samples. In addition, this study had received ethics approval from the Medical Research Ethics

Committee (MREC) of the Ministry of Health, Malaysia. The ethics approval reference was KKM/NIHSEC/08/0804/P10-622. Ethics approval for this project (NMRR-10-930-7461) was given on 24th January 2011 and was valid through 16th January 2015. A schematic diagram showing the process of blood collection, study design and procedure is shown in Figure 1.

2.2 Histopathological examination

Brain tumour tissue was sectioned onto glass slides and stained with haematoxylin and eosin (H&E). The slides were read by neuropathologists at the respective hospital. The diagnosis was made based on the World Health Organization (WHO) classification of tumours of the central nervous system (2007) (5). Of the 20 tumour samples, 10 each were high and low grade gliomas respectively (Table 1). Grade I and II tumours were classified as low grade while Grade III and IV tumours were classified as high grade. The incidence of gliomas is 2-3 new cases per 100,000 population per year (115). As such, the number of samples we were able to collect on our own was small.

2.3 Non-glioma and control samples

In addition to the 20 tumour samples, 10 non-glioma and 20 control samples were also obtained. The 10 non-glioma cases constituted patients with an inflammatory, non-malignant condition of the brain and included cases of hemangioblastoma, haemorrhagic and ischaemic stroke, inflammatory pseudotumour, arteriovenous malformation and

multiple sclerosis (Table 2). The 20 control subjects were healthy with no known illness (Table 3).

2.4 Blood sample collection

2.5ml of venous blood was drawn from each patient using the BD Vacutainer (Becton Dickinson, USA) with attached 21G x 3/4" x 12" butterfly needle directly into the PreAnalytiX (PAXgene) Blood RNA Tube (BRT) (Qiagen, Germany). The samples were kept at room temperature for 2 hours to allow for complete lysis of cell components after which they were stored at -20^0C.

2.5 RNA extraction

RNA was extracted from each blood sample using the PreAnalytiX (PAXgene)^{TM} Total RNA Blood Extraction Kit (Qiagen, Germany). After collection, the blood sample in the PAXgene Blood RNA Tube (BRT) was incubated at a minimum of 2 hours at room temperature to ensure complete lysis of blood cells. The BRT was then spun for 10 minutes at 3000-5000 x *g*. The supernatant was removed and the pellet containing the blood cells vortexed until dissolved in 4 ml of RNase- free water. The BRT was centrifuged again and the supernatant removed. 350µl of resuspension buffer was added and the pellet vortexed until dissolved. The sample was transferred into a 1.5ml microcentrifuge tube where 300µl of binding buffer was added to bind the RNA which was predominantly derived from leukocytes. 40µl of proteinase was also added to dissolve any protein present in the sample. The lysate was transferred directly into a PAXgene Shredder spin column and centrifuged to remove cell debris. The flow through

supernatant containing the total RNA was mixed with 350µl of 96-100% ethanol and vortexed. 700µl of the sample was pipetted into the PAXgene RNA spin column to which DNase I was added to remove any contaminating DNA. The PAXgene RNA spin column was washed several times with wash buffers 1 and 2 after which 40µl of elution buffer was added directly onto the PAXgene RNA spin column membrane. This was centrifuged for 1 minute at 8000-20,000 x *g* to elute the RNA. The eluate containing the total RNA was incubated at 65^0C for 5 minutes and then chilled immediately on ice.

The concentration and purity of the RNA was analyzed using the Spectrophotometer NanoDrop ND1000 (Thermoscientific, USA). The integrity of the RNA was analyzed using the Agilent 2100 BioAnalyzer RNA 6000 Nano Chip platform (Agilent Technologies, USA). The concentration of RNA obtained ranged from 37 to 442 ng/µl. The average value for the RNA Integrity Number (RIN) for the samples was 7.4 with a standard deviation of 0.87. The samples were stored at -80^0C until further use.

2.6 Microarray processing

Two-colour microarray-based gene expression utilizing the Agilent 4x44K whole human genome microarray, was performed on RNA isolated from the 50 blood samples. This array was chosen because it contained all the known genes in the human genome and was readily available at a competitive price. Standard protocols were followed for sample preparation, probe labelling and hybridization according to the Two-Colour Microarray-Based Gene Expression Analysis Protocol (Agilent Technologies, USA).

For sample preparation, the Two-Colour RNA Spike-In kit (Agilent Technologies, USA) was used. Spike A and Spike B Mix were thawed, mixed vigorously on a vortex mixer and then heated at 37^0C in a water bath for 5 minutes. 3 serial dilutions of 1:20, 1:40 and 1:4 were performed for each spike mix. For the labelling reactions, the Low Input Quick Amp Labelling Kit (Agilent Technologies, USA) was used. 150ng of total RNA to a volume of 1.5µl was labelled. 2µl of the Spike A Mix/Cy3-CTP was used to label the Universal Human Reference RNA (Stratagene, USA) while 2µl of the Spike B/Cy5-CTP was used to label the HG, LG, NG and C samples respectively. 1.8µl of T7 Promoter Primer Mix (consisting of 0.8µl T7 promoter primer and 1µl nuclease-free water) was added to the reaction containing 3.5µl of total RNA and diluted RNA spike-in mix. The primer and template were denatured by incubating the reaction in a water bath at 65^0C for 10 minutes. The reactions were then placed on ice for 5 minutes. 4.7µl of cDNA master mix (2µl 5X first strand buffer, 1µl 0.1M DTT, 0.5µl 10mM dNTP and 1.2µl AffinityScript RNase block mix) was added to each sample tube to a total volume of 10µl. Samples were incubated at 40^0C in a water bath for 2 hours after which they were moved to a 70^0C water bath and incubated for a further 15 minutes. The samples were then incubated on ice for 5 minutes. Finally, 6µl of transcription master mix (0.75µl nuclease-free water, 3.2µl 5X transcription buffer, 0.6µl 0.1M DTT, 1µl NTP mix, 0.21µl T7 RNA polymerase blend and 0.24µl Cy3-CTP/Cy5-CTP) was added to each sample tube for a total volume of 16µl and incubated at 40^0C in a water bath for 2 hours.

The resulting labelled/amplified cRNA was purified as per protocol using the RNeasy mini spin columns (Qiagen, Germany). The cleaned cRNA sample was eluted by

transferring the RNeasy column to a new 1.5ml collection tube. 30µl RNase-free water was added directly onto the RNeasy filter membrane and allowed to stand for 60 seconds. The RNeasy column in the collection tube was then centrifuged at 4^0C for 30 seconds at 13,000rpm. The flow-through containing the cRNA sample was maintained on ice. If not used immediately, the samples were stored at -80^0C.

The cRNA was quantified using the Nanodrop spectrophotometer as described previously. The yield and specific activity of each reaction was determined respectively as follows:

$$\frac{\text{(Concentration of cRNA) x 30µl (elution volume)}}{1000} = \text{µg of cRNA}$$

$$\frac{\text{Concentration of Cy3 or Cy5}}{\text{Concentration of cRNA}} \times 1000 = \text{pmol Cy3 or Cy5 per µg cRNA}$$

For the 4-pack microarray format, almost all yields obtained were ≥ 0.825µg and had specific activity (pmol Cy3 or Cy5 per µg cRNA) ≥ 6.

The initial step for the hybridization reactions involved the fragmentation of RNA. For the 4-pack microarray format, 825ng each of Cy3- and Cy5-labelled, linearly amplified cRNA, 11µl of 10x blocking agent were made up to a volume of 52.8µl with nuclease free water, after which 2.2µl of 25x fragmentation buffer was added to a total volume of 55µl. The samples were incubated at 60^0C for exactly 30 minutes to fragment the RNA and then immediately cooled on ice for 1 minute. 55µl of the fragmentation mix containing cRNA was mixed with an equal volume of 2x GE hybridization buffer HI-RPM. The samples were spun in a microcentrifuge at 13,000 rpm for 1 minute at room

temperature to drive any residual sample from the walls and lid of the tubes and to help with bubble reduction. The samples were then placed on ice and loaded onto the array immediately. 100µl of sample was pipetted into the gasket slide well of the Agilent SureHyb chamber and the "active side" of the array placed directly on top of the gasket slide to form a sandwich pair. The SureHyb chamber cover was placed on the sandwiched slides and the clamp assembly tightened onto the chamber. The assembled slide chamber was then placed in a rotisserie hybridization oven at 20rpm and the samples allowed to hybridize at 65^0C for 17 hours. The slides were then washed with Gene Expression Wash Buffer 1 followed by Prewarm Gene Expression Wash Buffer 2. In addition, 0.0005% Triton X-102 was added to both buffers which reduced the possibility of array wash artifacts. The microarray slides were scanned using the DNA Microarray Scanner (Agilent Technologies, USA).

2.7 Data extraction

Data was extracted using Agilent feature extraction software analyzed with Gene Spring version GX 12.5V (Agilent Technologies, USA). The data files were extracted in text (.txt) format after Lowess normalization. The sequence of events involved in processing of the data files were as follows: thresholding, summarization, dye swap, ratio computation, log transformation and baseline transformation.

Thresholding involved a substitution step where all expression values below a certain specified value were made constant. Thresholding was done to remove very small expression values or negative values in the dataset. This was to ensure that there were no very large negative numbers when the data was log transformed. Summarization was

done by calculating the geometric mean of the expression values. Raw signal values were then generated which essentially were linear data that had undergone thresholding and summarization for the individual channels (Cy3 and Cy5). Normalized signal values refer to the data after it has undergone ratio computation, log2 transformation and baseline transformation. Normalization was also done using the Human Reference RNA.

Dye-swapping accounts for dye related bias as different dyes (Cy3 and Cy5) bind DNA with different affinities. This dye related bias cannot be removed by standard normalization methods. In GeneSpring, samples that have been marked as dye-swapped were treated as follows: Cy3 was designated as "signal" and Cy5 as "control" and the signal was computed as Cy3/Cy5. For samples that have not undergone dye-swapping, GeneSpring treats Cy5 as "signal" and Cy3 as "control" and the signal is computed as Cy5/Cy3.

In baseline transformation, the baseline to median of control samples was performed. In the Agilent 4x44K Human Array, there are a set of samples designated as controls that can be used for all samples. In this baseline transformation, for each probe, the median of the log summarized values from the control samples was first computed, after which, this value was subtracted from the sample.

As mentioned previously, Lowess normalization was performed before the raw data was extracted. Lowess normalization is critical for reducing intra-array (within slide) variation. In 2 colour experiments, 2 fluorescent dyes, red and green, are used. The intensity-dependent variation in dye bias may introduce spurious variations in the dataset.

Lowess normalization is performed which merges the 2 colour data and applies a smoothing adjustment which removes such variations.

2.8 Investigating the effects of tumour status, age, gender and experimental array batch on gene expression

A linear mixed regression analysis was performed using the "R" statistical package, to investigate the effects of tumour pathology, age, gender and experimental batch effects on gene expression. The tumour status was defined by the four groups of samples, HG, LG, NG and C. The age and gender of patients are represented in Tables 2, 3 and 4. Samples were run on arrays in 4 experimental batches as follows: Batch 1: 6 HG and 6 LG samples, Batch 2: 4 HG, 4 LG and 4C samples, Batch 3: 8C and 4 NG samples, Batch 4: 8C and 6 NG samples. The explanatory power of each factor was assessed in a stepwise manner by examining the increase in the variation explained when a new covariate or set of covariates was added to the existing model. This resulted in the investigation of the following four models:

(1) Model 1: Gene expression as a function of tumour status

(2) Model 2: Gene expression as a function of tumour status and age

(3) Model 3: Gene expression as a function of tumour status, age and gender

(4) Model 4: Gene expression as a function of tumour status, age, gender and array batch

In total there were 50 microarray samples each with 29,092 gene expression values from 44,000 probe sets. In preparation of input data for multiple regression analysis, a table of 50 microarray samples (50 rows) x 29,092 gene expression values (29,092 columns) was generated. The metadata for each sample that included tumour status, age, gender and array batch were combined as columns in the prepared table. The input data was then read into the "R" software. For each gene, a linear model was fitted using the Im() function to its respective gene expression values versus variable(s) of interest as per the four models. For each model, there were 29,092 r^2 (coefficient of determination) values that were generated. Each r^2 value was then modified to generate an adjusted r^2 value to account for the number of variables and the sample size. A median, mean and range for r^2 was then calculated for each model as shown in Figure 2 and Table 7.

2.9 Unsupervised hierarchical clustering

Unsupervised hierarchical clustering using the Euclidean distance method and Ward's linkage was performed on each of the 4 different pairs of conditions and all 4 conditions.

One of the limitations in unsupervised hierarchical clustering is that this form of analysis could be influenced by noise and outliers particularly when sample sizes are small.

2.10 Principal component analysis (PCA)

Principal component analysis (PCA) was performed on the complete data set. The first step in PCA was to subtract the mean from each of the data dimensions. Then, the

covariance matrix and the eigenvectors and eigenvalues of the covariance matrix were calculated. Data compression and reduced dimensionality was performed when converting the data into components and to form feature vectors in 3 dimensions along the x, y and z axis.

2.11 Identification of significant differences in gene expression between the 4 different conditions

The moderated t-test, a modification of the Student's t-test, was used to identify significant differences in gene expression between the 4 sets of conditions (HG vs C, LG vs C, HG vs LG and NG vs C). While the Student's t-test calculates variance from the data that is available for each gene, the moderated t-test uses information from all of the genes to calculate variance. This is particularly useful when a small number of samples is available in each group (as in this case) making the variance estimates unstable.

When testing was performed across these different conditions, each gene was considered independently from the other as a moderated t-test was performed on each gene separately. Given that in this microarray experiment, the expression levels of 44,000 probes was measured simultaneously across each condition, multiple testing correction (MTC) was required. MTC is performed to control for occurrence of false positives that arise as a result of performing multiple tests. The purpose of MTC is to keep the overall error rate (false positive rate) to less than the p-value cutoff as specified by the user, even if hundreds or thousands of genes are being analysed. With this in mind, the Benjamini and Hochberg (B-H) false discovery rate was used to control for the large number of tests performed. This procedure is one of the less stringent methods of MTC but it provides a

good balance between identification of many genes that are statistically significant and protection against false positives (type I error).

The B-H multiple-testing correction of p-values was chosen because it is widely used, relatively easy to implement and computationally inexpensive. However, the B-H corrected p-values may be slightly over-corrected as the p-value decreases because they involve the upper boundary of the false discovery rate. This corresponds to the case where the null hypothesis is true for all tests. In addition, null distributions for the different tests are assumed to be independent which may not be the case.

2.12 Pathway analysis

For each group, genes were selected based on at least a 2-fold difference in expression and a B-H corrected p-value <0.01. Pathway analysis was performed using the Ingenuity Pathway Analysis (IPA) programme (Johns Hopkins University, Baltimore, Maryland, USA).

IPA is based on the Ingenuity Knowledge Base. In IPA, canonical pathways are well characterized pathways that have been curated and hand-drawn by PhD level scientists and the information comes from specific journal articles, review articles, text books, and Human Cyc, an encyclopaedia of human genes and metabolism (http://humancyc.org). Gene selection for the canonical pathways is based on this analysis. Although IPA is a single database, it is built upon a large collection of resources and the Ingenuity Knowledge base is one of the largest repositories in the world. It includes information from many databases such as the Biomolecular Interaction Network

Database (BIND), Biological General Repository for Interaction Databases (BIOGRID), ClinicalTrials.gov, Cognia, Database of Interacdting proteins (DIP), Gene Ontology (GO), GVK Biosciences, HumanCyc, Ingenuity Expert Findings, Ingenuity ExpertAssist Findings, INTACT, Interactome studies, Molecular Interaction Database (MINT), Munich Information Centre for Protein Sequences (MIPS), and microRNA database (miRBase). Furthermore, the IPA database is updated weekly. One drawback is that some very well established curated pathway databases are not included in IPA such as KEGG and REACTOME. Ideally, an analysis should be run against all canonical pathway databases. However, there is a very high level of overlap between many of these pathway databases.

2.13 Fisher's Exact Test and p-values

For a given gene list and pathway, the p-value associated with a pathway is a measure of its statistical significance in relation to the genes or pathway annotation for the gene list and a Reference Set of genes, which is defined by the total number of genes that could possibly be involved in all the pathways. The p-value for the pathways was calculated using the right-tailed Fischer's Exact Test.

In the right-tailed Fischer's Exact Test, only the over-represented functions or pathways will be shown. Under-represented functions or pathways (left-tailed p-values) will not be shown. Due to the large number of pathways tested, multiple testing correction (MTC) was performed using the Benjamini-Hochberg false discovery rate (FDR) with a corrected p-value <0.05. In some cases, the supporting evidence was further explored to grasp a better understanding of the biological significance.

2.14 cDNA synthesis

RNA from each sample was converted to cDNA using the high capacity RNA-to-cDNA kit (Applied Biosystems, USA). This kit used the reverse transcriptase enzyme from the multiscribe murine leukaemia virus (MuLV) mixed with RNase inhibitor protein. Optimal blend priming was performed with a mixture of random octamers and oligo dT primers.

200ng of total RNA was mixed with 10.0 µl of 2x Reverse Transcriptase (RT) buffer, 1.0 µl 20X enzyme mix and nuclease-free water to a total volume of 20.0 µl. The tube containing the reaction mix was then incubated in the T-Professional basic thermocycler (Biometra, Germany) at 37^0C for 60 minutes after which the reaction was terminated by heating to 95^0C for 5 minutes. The reactions were then used for droplet digital PCR (ddPCR) or stored long-term at -80^0C.

2.15 Droplet digital polymerase chain reaction (ddPCR)

Selected genes (Table 12 under Results section) from each of the 4 group pairs as mentioned previously, were verified using ddPCR. Reactions for each sample were done either singly or in duplicate. Beta- glucuronidase (*GUSB*) was used as the reference gene as it showed the least variation with gene expression amongst the other housekeeping genes used, namely TATA binding protein (*TBP*) and human acidic ribosomal protein (*HuPO*). All reagents and equipment used for ddPCR were from Bio-Rad, USA. 10ng of cDNA was mixed with 10 µl of 2x ddPCR Supermix for Probes (No dUTP), 1 µl 20x target primers/probe mix (FAM) or 20x reference primers/probe (HEX) and nuclease-free

water to a total reaction volume of 20 µl. The entire reaction mix of 20 µl was then loaded into a sample well of a DG8 Cartridge for the QX200/QX100 Droplet Generator. This was then followed by adding 70 µl of Droplet Generation Oil for Probes into the oil wells of the cartridge, according to the QX200/QX100 Droplet Generator Instruction Manual. The cartridge was then inserted into the Automated Droplet Generator. After droplet generation, the droplets were transferred to a 96-well plate and then sealed with foil using the PX1 PCR Plate Sealer.

Thermal cycling was then performed on the droplets using the C1000 Touch Thermal Cycler with 96-Deep Well Reaction Module according to the following protocol: enzyme activation at 95^0C for 10 minutes (1 cycle), denaturation at 94^0C for 30 seconds followed by annealing/extension at 55^0C for 1 minute (40 cycles), enzyme deactivation at 98^0C for 10 minutes (1 cycle) followed by hold at 4^0C. The ramp rate was set at 2^0C/sec, the heated lid to 105^0C and the sample volume at 40µl. After thermal cycling, the sealed plate was placed in a QX200/QX100 Droplet Reader and the absolute gene expression level per well for the probes and reference genes were quantitated using the QuantaSoft Software.

For analysis of the gene expression data, we assumed a normal distribution. Each gene was evaluated for its expression in a minimum of 3 to a maximum of 6 samples under each pair of conditions. The gene expression values for each sample were normalized to the housekeeping gene. The values for the absolute level of gene expression as obtained by ddPCR were then subjected to the t-test for the genes selected

under the 4 sets of conditions, with a resulting fold change and p-value. Statistical outliers were removed using the box and whisker plot.

Histopathology	Grade	Age	Gender
Pilocytic astrocytoma	I	31	Male
Diffuse astrocytoma	II	17	Male
Diffuse astrocytoma	II	32	Male
Fibrillary astrocytoma	II	62	Female
Recurrent astrocytoma	II	45	Female
Diffuse astrocytoma	II	36	Male
Low grade astrocytoma	II	59	Male
Low grade oligodendroglioma	II	45	Male
Low grade oligodendroglioma	II	56	Male
Recurrent oligodendroglioma	II	59	Male
Anaplastic oligoastrocytoma	III	37	Female
Anaplastic oligoastrocytoma	III	58	Male
Recurrent anaplastic oligoastrocytoma	III	66	Male
Anaplastic astrocytoma	III	29	Female
Anaplastic astrocytoma	III	43	Male
Glioblastoma multiforme	IV	24	Male
Glioblastoma multiforme	IV	54	Male
Glioblatoma multiforme	IV	24	Male
Glioblastoma multiforme	IV	34	Male
Glioblastoma multiforme	IV	56	Female

Table 2: WHO classification, histopathology of tumour samples and demographic data

Patient no	**Age**	**Gender**	**Sample Type**
1	40	Male	Hemangioblastoma
2	77	Male	Blood Clot
3	44	Female	Inflammatory pseudotumour
4	27	Male	Arteriovenous malformation (AVM)
5	51	Female	Ischaemic stroke
6	53	Female	Hemangioblastoma
7	61	Male	Haemorrhagic stroke
8	56	Female	Multiple sclerosis
9	34	Female	Ischaemic stroke
10	46	Female	Haemorrhagic stroke

Table 3: Demographics and types of non-glioma samples

Patient no.	Age	Gender
1	30	Female
2	38	Female
3	41	Male
4	57	Male
5	25	Male
6	57	Male
7	33	Male
8	51	Male
9	28	Male
10	25	Male
11	56	Male
12	32	Male
13	22	Male
14	59	Female
15	42	Female
16	55	Male
17	58	Male
18	48	Male
19	33	Male
20	55	Male

Table 4: Demographics of control samples

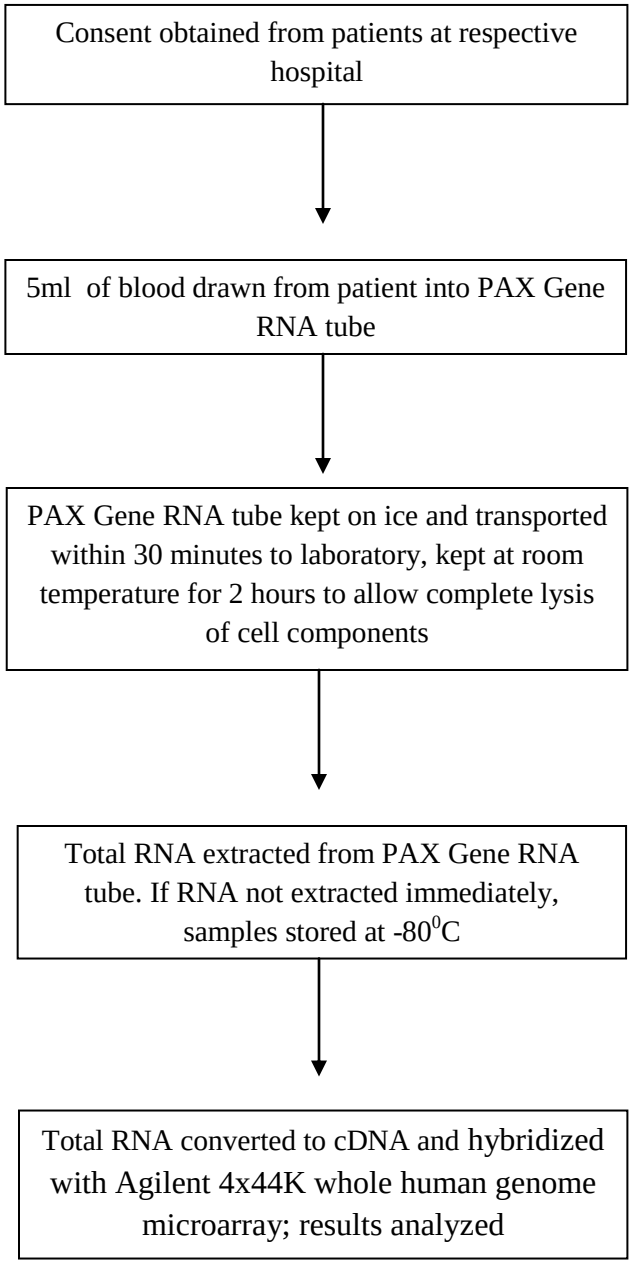
Patient/Subject type	No. of patients/subjects	Male	Female	Age range (years)	Median age (years)
Low Grade Glioma	10	8	2	17-62	45
High Grade Glioma	10	7	3	24-66	40
Non-glioma	10	4	6	27-77	48.5
Control	20	16	4	22-59	41.5

Table 5: Summary statistics of patients

Question	Information requested
1	Patients were asked if there was any history of brain tumours or other cancers in the family.
2	Patients were asked if they had any of the following neurological symptoms prior to presenting to the hospital: (a) Headaches (b) Seizures (c) Personality changes (d) Slurring of speech (e) Changes in vision (f) Changes in sensation (g) Motor weakness (h) Other symptoms such as vomiting

Table 6: Questionnaire

Figure 1: Schematic diagram of blood collection, study design and procedures



CHAPTER 3: RESULTS

3.1 Modelling the effects of tumour status, age, gender and experimental array batch effects on gene expression

Using a linear mixed regression analysis, the effect of tumour pathology, age, gender and experimental batch on gene expression was investigated. For models 1, 2, 3, and 4, the median and mean adjusted r^2 values did not vary significantly (Table 7, Figure 2). The change in median values for models 2 and 3 were 0.0077% and 0.0065% respectively as compared to model 1, suggesting that age and gender had a minimal impact on gene expression globally. For model 4, the change in the median adjusted r^2 value when compared to tumour status alone was 3.57%, indicating that array batch had some impact on gene expression globally, but that this was still small. The mean adjusted r^2 values showed a similar trend. Based on these findings, all subsequent analysis focused on the impact of tumour pathology alone on gene expression.

3.2 (a) Microarray analysis of samples

Unsupervised hierarchical clustering was performed on each of the 4 different pairs of conditions (HG vs C, LG vs C, HG vs LG and NG vs C) and all 4 conditions together, with the total gene input list, using the Euclidean distance method and Ward's linkage. The gene input list consisted of genes which were found to be differentially expressed with a corrected p-value < 0.01 and a fold change of at least 2. The results are shown in Figures 3(a-e). All the microarray analysis heat maps clearly showed the different groups clustering independently. This was seen clearly in the unsupervised

hierarchical clustering with all four conditions together which showed that there were 3 clusters of samples: one cluster for the high grade tumours, a second cluster for the low grade tumours and a third cluster for the non-glioma and control samples. In Figure 3(e), the non-glioma and controls clustered together and were distinct from the glioma samples, with the low grade glioma being furthest from the control and non-glioma samples. Therefore, not only were we able to show a distinction between the glioma and control samples but were also able to distinguish between high and low grade gliomas.

(b) Principal component analysis (PCA) of samples

The PCA plot (Figure 4) of the first 3 axes, showed results that were very similar to that of the microarray analysis, demonstrating clear separation into the 3 clusters of samples as mentioned in 3.2(a). Of specific note, is that the 2 sample types which were closest to each other were the control and non-glioma sets.

3.3 Volcano plots

Multiple testing correction using the Benjamini-Hochberg (B-H) analysis with a corrected p value of <0.01, and a two-fold change cut-off, for each of the four conditions were as follows: HG vs C: total number of genes: 1055, with 479 upregulated and 576 downregulated; LG vs C: total number of genes: 2708, with 713 upregulated and 1995 downregulated; HG vs LG: total number of genes: 1629, with 1287 upregulated and 342 downregulated; and NG vs C: total number of genes: 82, with 56 upregulated and 26 downregulated (Table 8). The results were represented on volcano plots [Figures 5(a)-(d)].

The results showed that there were relatively few genes which were differentially expressed between control and non-glioma samples. In comparing the glioma samples to the controls, the predominant effect was the downregulation of genes in the glioma samples. When comparing the high and low grade samples, there was in general an upregulation occurring in the high grade samples.

3.4 Venn Diagram of Differentially Expressed Genes

The Venn Diagram represented the genes with at least a 2 fold difference in expression and a p-value <0.01, that were unique to each condition and also those that overlapped between the various conditions (Figure 6). There were 104 genes common to both the HG vs C and the HG vs LG pairs. These included genes belonging to the zinc finger transcription factor, *ZNF 649* and *ZNF 205*, homeobox genes such as *HOXB2* and *SOX8*, a transcription factor involved in embryonic development and determination of cell fate. For the HG vs LG pair, there were a total of 1629 genes, of which 644 were unique to this pair and included EGFR, *TGFβ1* and *VEGFA*. There were 573 genes common to both the HG vs C and LG vs C pairs. These common genes included *IL12RB1*, *FOS, TP53* and *TNF*. One important gene common to the HG vs C, LG vs C and HG vs LG pairs was *IL6*.

For the NG vs C pair, there were 46 unique genes, 19 that overlapped with the HG vs C pair, 7 that overlapped with the LG vs C pair and another 7 that were common to the HG vs LG and LG vs C pairs. There were no genes common to all 4 conditions (Figure 6).

3.5 Canonical Pathways

The significance of association between differentially expressed genes and the canonical pathways (as annotated by the HumanCyc Pathway Database) were assumed to follow a normal distribution and assessed using the B-H multiple testing correction to calculate a p-value. Only those pathways with a corrected p-value <0.05 were selected. This determined the probability that the association between the genes and the pathways, relative to all functionally characterized human genes, were not explained by chance alone [Figure 7(a)-(d)]. The IPA also determines whether the pathways are activated or inhibited by assigning a z score. The ratio defined the proportion of differentially expressed genes from a pathway to the total number of genes that make up that particular pathway. For the HG vs C pair, 4 significant pathways were identified (ratios ranging from 0.084 to 0.136) with no evidence for significant activation or inhibition as shown by z scores close to zero. (Figure 7a). The 4 significant pathways included those involved in innate and adaptive immunity. For the LG vs C pair, the IPA predicted a mixed pattern of activity for the 46 significant pathways with 23 pathways having no activity pattern available, 5 pathways having a positive z-score (predicted activation), 16 pathways having a negative z-score (predicted inhibition) and 2 pathways having a z-score of zero (Figure 7b). The z-score of zero corresponded to the standard mean of the normal distribution curve. Pathways having no activity pattern available meant that a z-score could not be calculated. The significant pathways with a positive z-score included those involved in *LXR/RXR* activation, *RhoG*, *Ephrin B*, IL-8 and cholecystokinin/gastrin-mediated signalling. The z-score of zero included pathways involved in NF-κB activation by viruses and glioma invasiveness signalling. The significant pathways with a negative

z-score were signalling by the Rho family of GTPases, *Tec* kinase signalling, *HGF*, eicosanoid, integrin, acute phase response, *PEDF* and thrombin signalling. It also included pathways involved in *PKC*, actin nucleation and immune system signalling.

For the HG vs LG glioma pair, 9 significant pathways were predicted with 6 having no activity pattern and 1 each with a positive, negative and zero z-score respectively. The activity pattern referred to the differential expression of genes that made up the pathway. The 6 pathways with no activity pattern were those involved in *FXR/RXR* activation, superoxide radical degeneration, hepatic fibrosis/hepatic stellate cell activation, role of tissue factor in cancer, clathrin-mediated endocytosis and atherosclerosis signalling. The pathways that had a positive z-score, a z-score of zero and a negative z-score were pathways involved in *LXR/RXR* activation, coagulation system and acute phase response signalling respectively (Figure 7c). For the NG vs C pair, there was only one significant pathway, hepatic fibrosis/hepatic stellate cell activation that had no activity pattern available (Figure 7d).

In summary, the total number of pathways with differentially expressed genes for each pair of conditions after B-H multiple testing correction, is shown in Table 9. The list of pathways are listed in Table 10. In Tables 11(a)-(d), the significant pathways are shown for each pair of conditions together with the B-H multiple testing correction value and associated genes.

3.6 Venn Diagram of Significant Pathways

The Venn diagram for the pathways showed pathways that were unique to each pair of conditions and also pathways that overlapped between the 4 different groups

(Figure 8). For the HG vs C pair, there was 1 unique pathway and 3 pathways that overlapped with the LG vs C pair.. The LG vs C pair had 39 unique pathways. The pathways that overlapped between the HG vs C and LG vs C pairs were pathways involved in the innate and adaptive immune response. The HG vs LG pair had a total of 9 significant pathways, with 4 unique pathways, 4 overlapping with the LG vs C pair and 1 overlapping with the NG vs C pair. The 4 unique pathways were the superoxide radicals degradation, clathrin-mediated endocytosis, coagulation system and role of tissue factor in cancer pathways. The 4 pathways overlapping with the LG vs C pair were the acute phase response, *FXR/RXR* activation, *LXR/RXR* activation and atherosclerosis signalling pathways. The 1 pathway overlapping with the NG vs C pair was the hepatic fibrosis/hepatic stellate cell activation pathway (Figure 8).

3.7 Heat Map

A heat map (Figure 9) with genes commonly involved in tumour signalling pathways especially in high and low grade brain tumours was generated with the four types of samples, namely C, NG, LG and HG glioma respectively. The results showed a unique differential pattern of expression for each of the 4 sample types. In addition, genes commonly upregulated in high grade tumours such as *EGFR* and *VEGFC*, are also highly expressed in blood. On the other hand, these genes are downregulated in the low grade tumour heat map. Specific isoforms of *Bcl2* such as *Bcl2L11* and *Bcl2A1* are upregulated in the low grade but not high grade samples. None of the genes involved in tumourigenesis are significantly upregulated in the non-glioma and control samples.

3.8 Genes Chosen for Validation by ddPCR

10 genes were selected for statistical validation by ddPCR (Table 12). These genes were selected from the list of differentially expressed genes that were significant from the 4 pairs of conditions. These genes were selected because they were known to be common genes involved in pathways related to tumourigenesis including the pathogenesis of brain tumours. Only the NG vs C had no significant genes that were downregulated. The other 3 conditions had significant genes that were both upregulated and downregulated.

Each gene was evaluated for its expression in a minimum of 3 to a maximum of 6 samples under each pair of conditions. The values for the absolute level of gene expression as obtained by ddPCR was then subjected to statistical analysis. A normal distribution of the values was assumed and the t-test applied to each gene with a resulting p-value. 7 of the 10 genes had p-values <0.05 and 3 genes had p-values >0.05. The genes with a p-value <0.05 were *MMP, MAP3K8, TP53, SOS1, FOS, IL6* and *TNF*. The genes with a p-value >0.05 were *EGFR, VEGFA* and *MAPK12* (Table 12). Multiple testing correction of the p-value using the Bonferroni correction with a threshold p-value of 0.05 and 10 test samples, resulted in only 4 genes that were highly significant. The genes were *MAP3K8, TP53, SOS1* and *IL6* (Table 13).

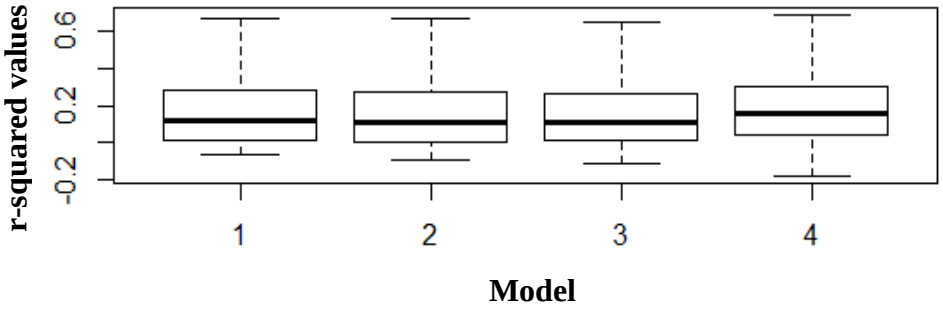


Figure 2: Box plot of adjusted r^2 values for the 4 models

Figure 3 (A)-(E): Unsupervised hierarchical clustering of the different conditions by use of the Euclidean similarity measure and Wards linkage to visualize variation in gene expression between the selected conditions: (A) HG vs C, (B) LG vs C, (C) HG vs LG, (D) NG vs C and (E) all 4 conditons together. The heat map showed the gene expression for the different conditions in columns, with a dendogram representing their similarity. The clustering was performed on a filtered gene list of normalized signal intensity values (averaged over replicates) for each condition.

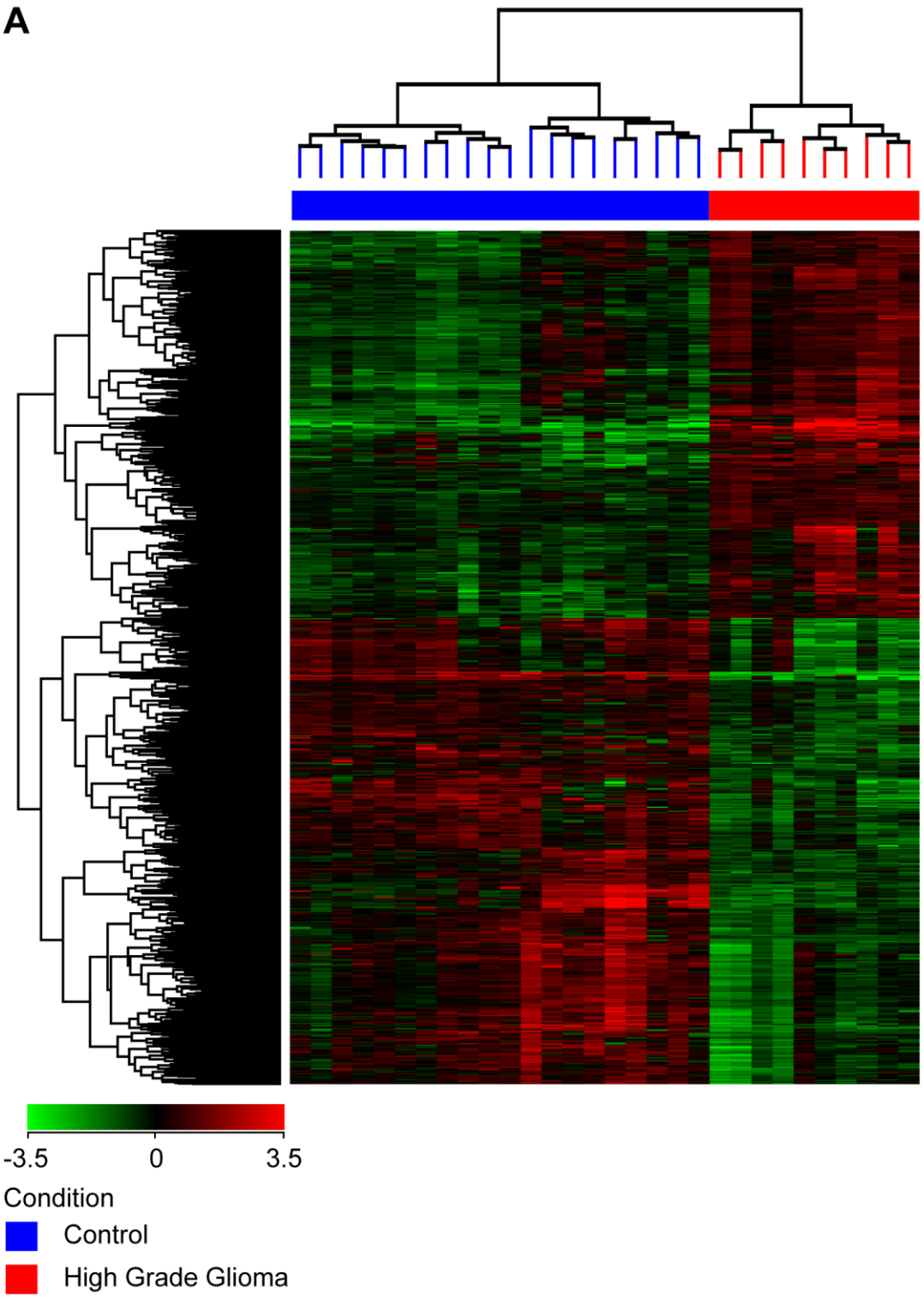


Figure 3(a): Unsupervised hierarchical clustering of HG vs C samples.

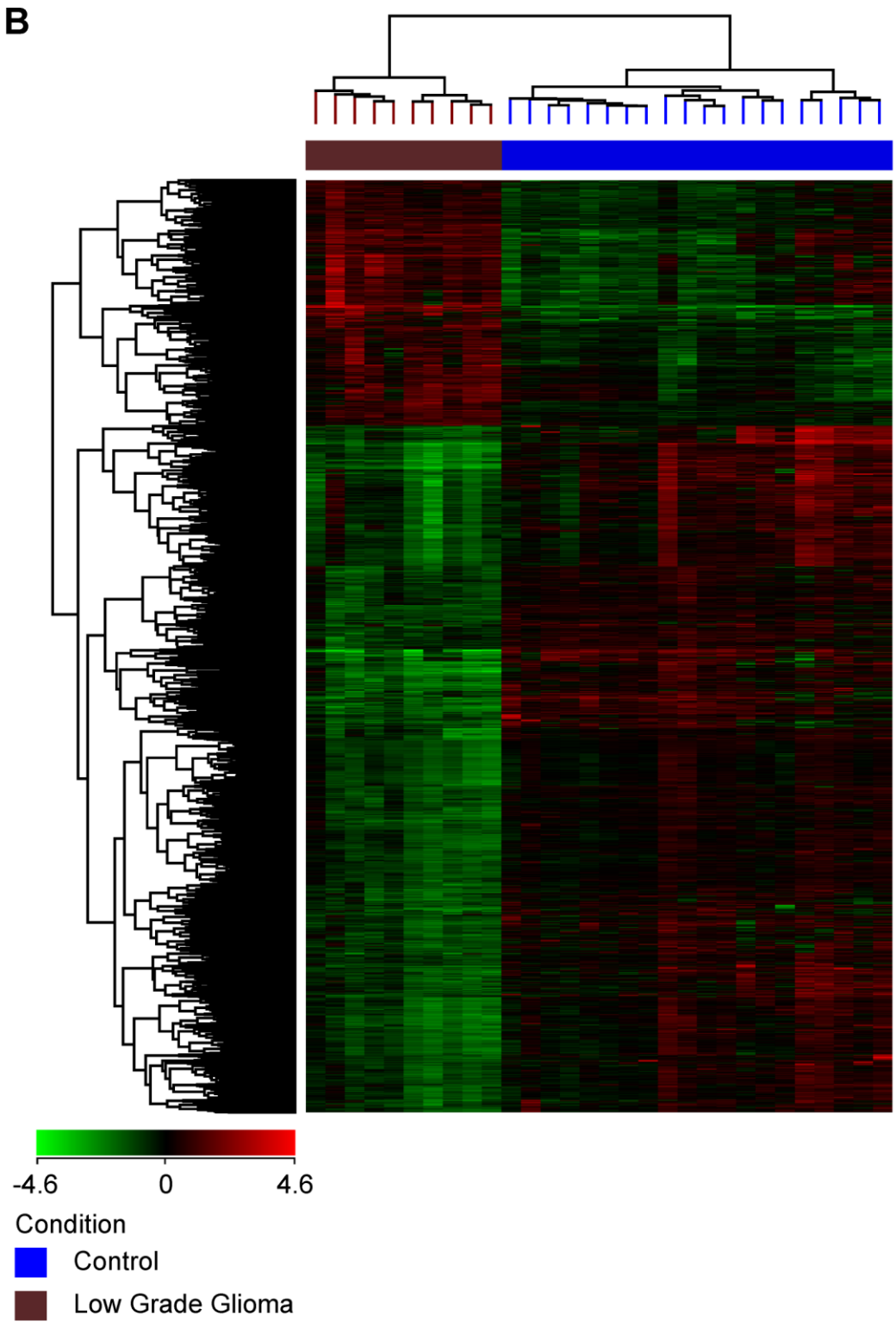


Figure 3(b): Unsupervised hierarchical clustering of LG vs C samples

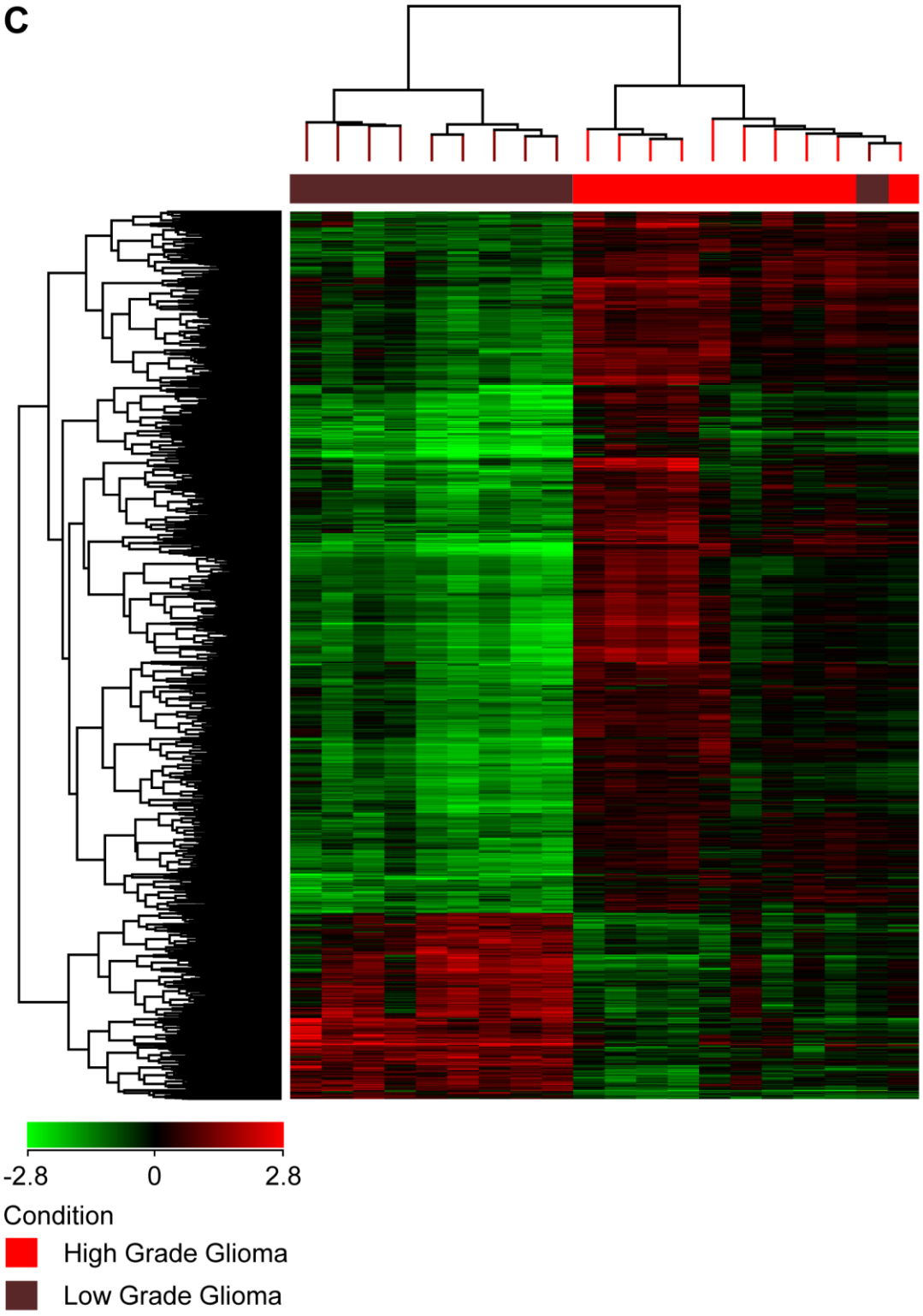


Figure 3(c): Unsupervised hierarchical clustering of HG vs LG samples

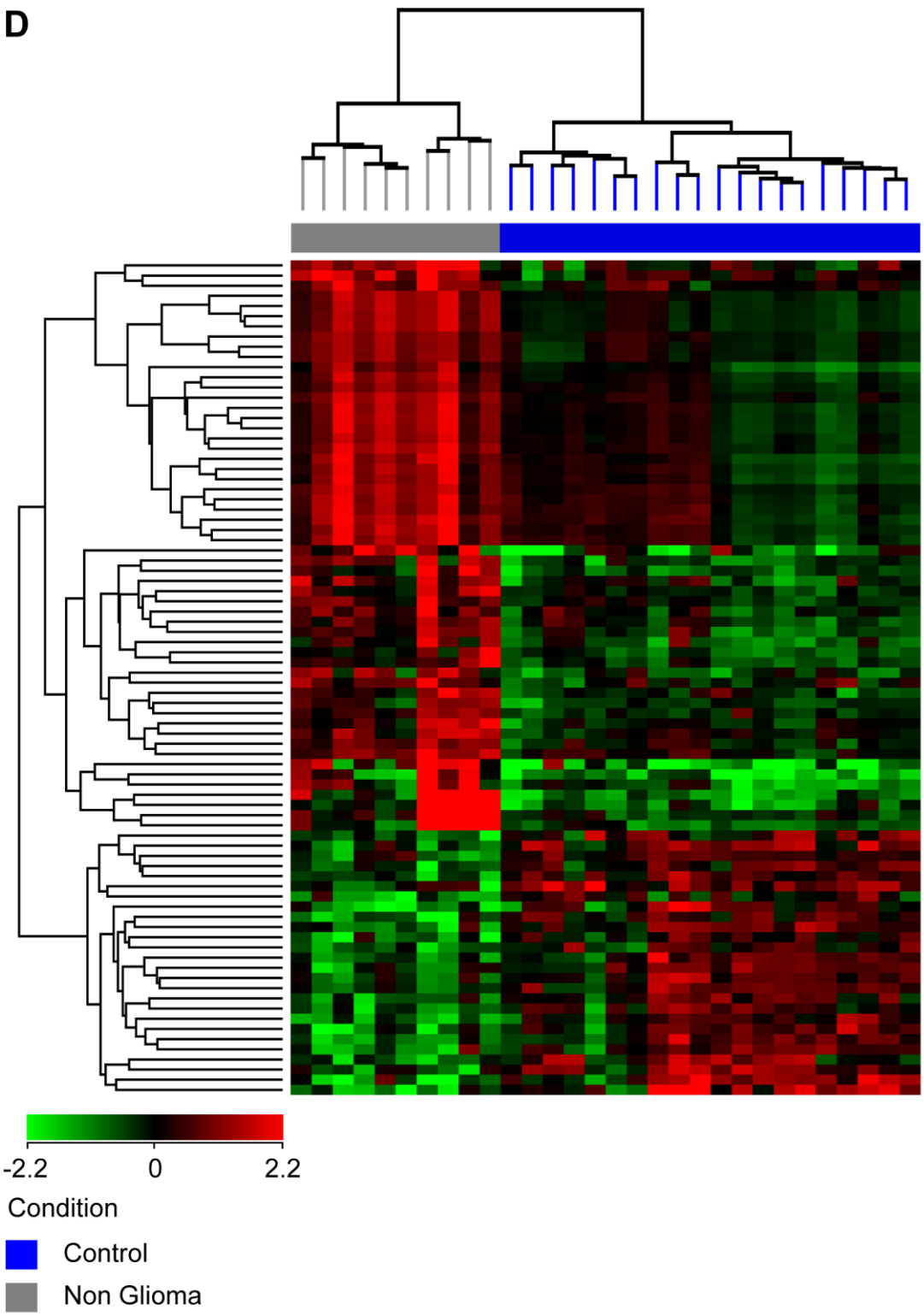


Figure 3(d): Unsupervised hierarchical clustering of NG vs C samples

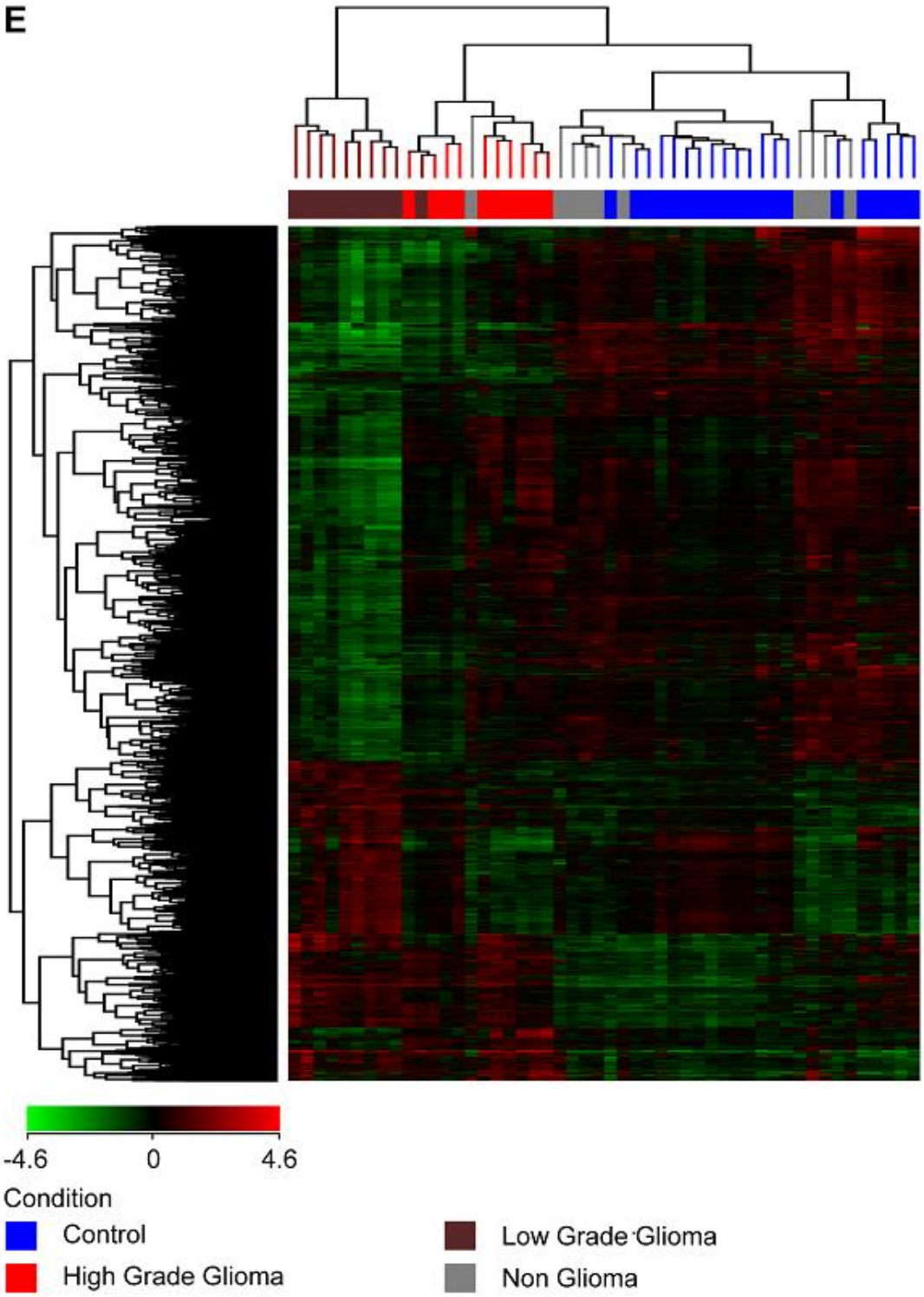


Figure 3(e): Unsupervised hierarchical clustering of all 4 condition pairs

Figure 4: Principal Component Analysis (PCA) plots for the different conditions. Panels (A) - (C) are plots based on the first 3 principal components in various orientations. The axes corresponds to principal component 1 (PC1, x-axis), PC2 (y-axis) and PC3 (z-axis). Panel D is a plot of PC axes 1 and 2. The ellipses (2 standard deviation coverage; see colour key for the different conditions) showed a distinct directionality in the different groups based on similarities in gene expression.

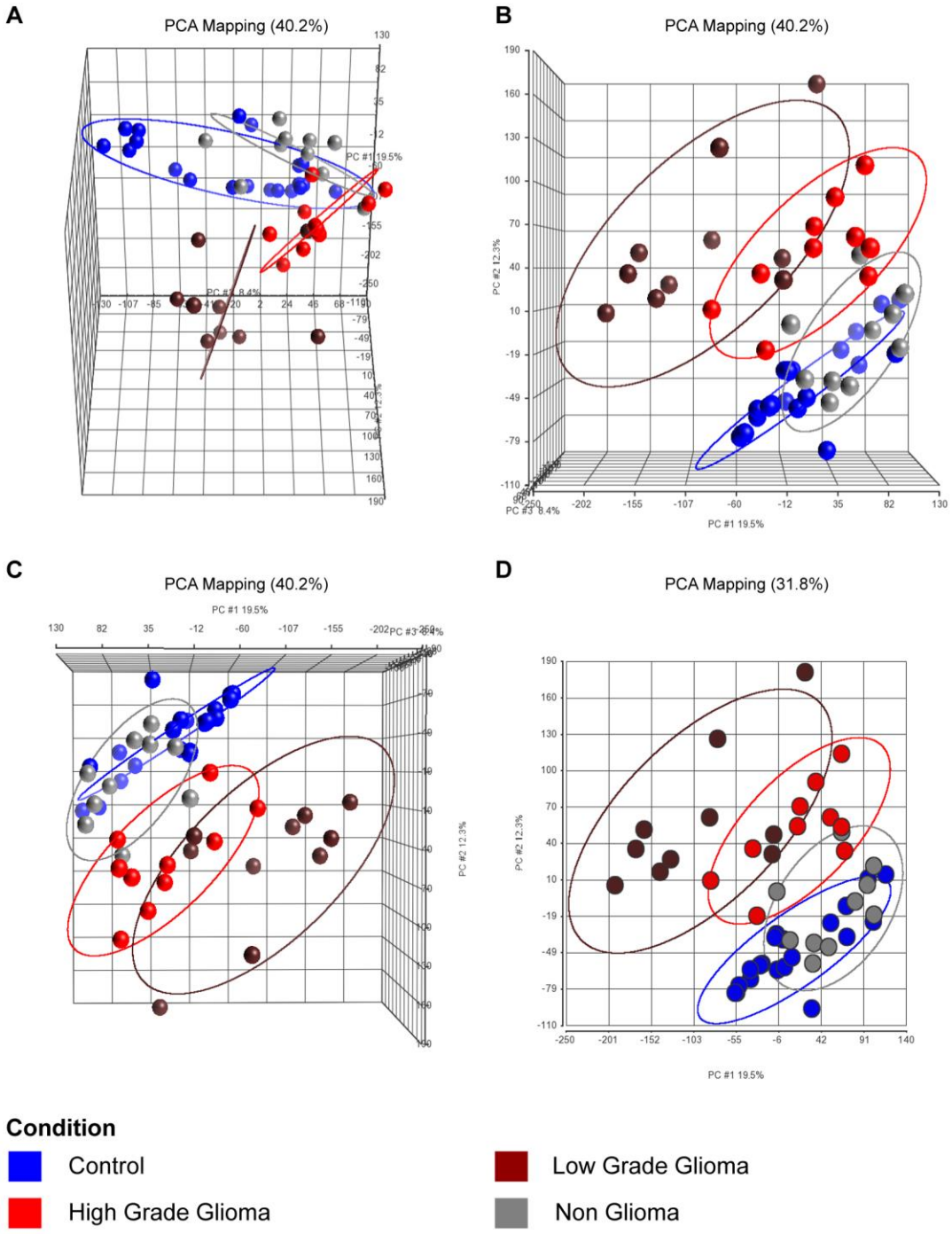


Figure 4: Principal Component Analysis (PCA) plots in various orientations for all 4 conditions together.

Figure 5: Volcano plots to determine differentially expressed genes for the individual pairs of conditions: (A) HG vs C, (B) LG vs C, (C) HG vs LG, and (D) NG vs C. The x-axis represents the log2 fold change of genes for the different condition pairs, while the y-axis represents the -log10 of the corrected p-values for the different pairs of conditions. Each dot represents a gene and the red coloured area represents the differentially expressed genes that meet the selection criteria of a fold change (FC) of at least 2 (FC≥ 2 or ≤ 2) and a p-value <0.01.

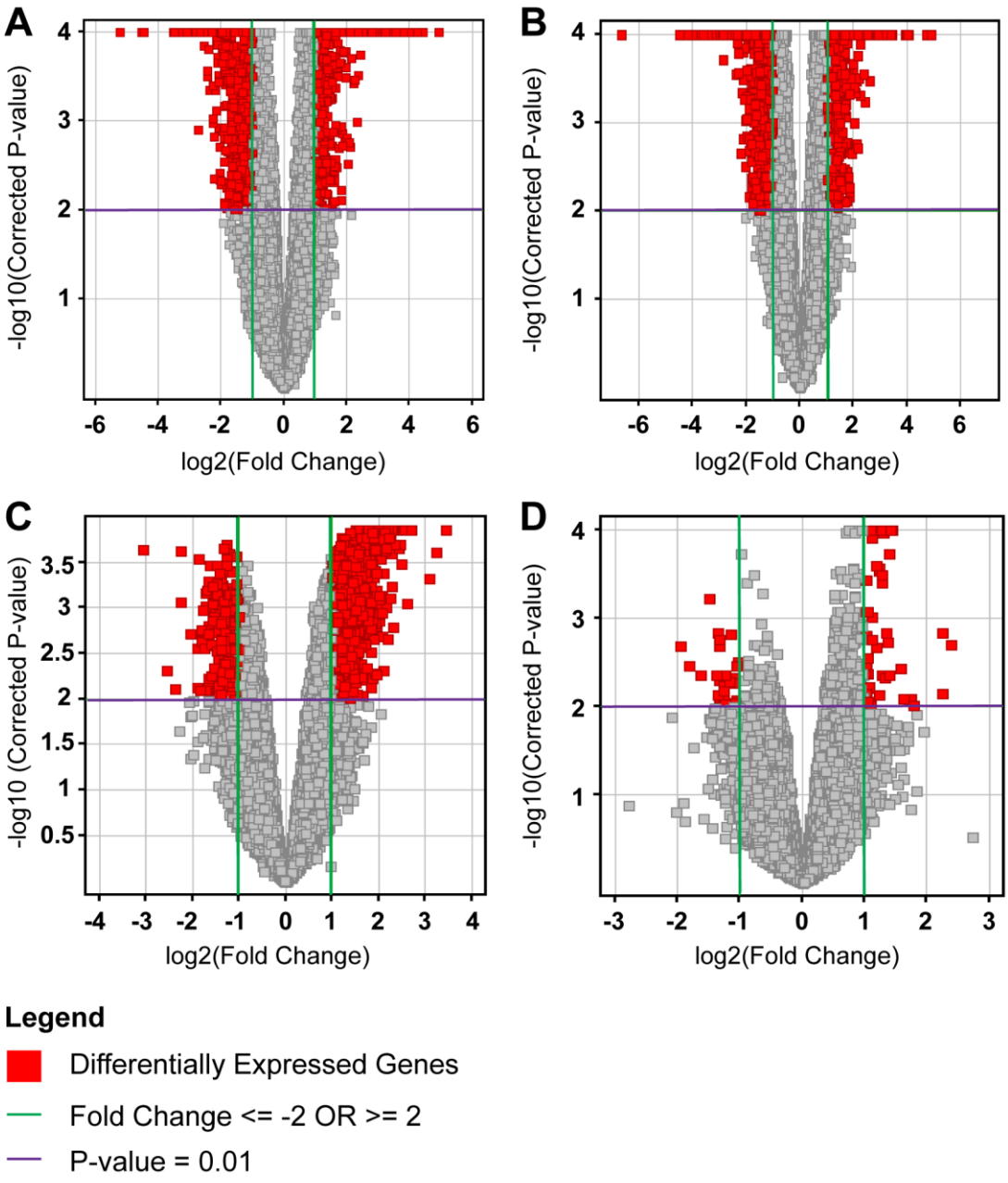


Figure 5: Volcano plots showing differentially expressed genes for the individual pairs of conditions: (A) HG vs C; (B) LG vs C; (C) HG vs LG ; (D) NG vs C.

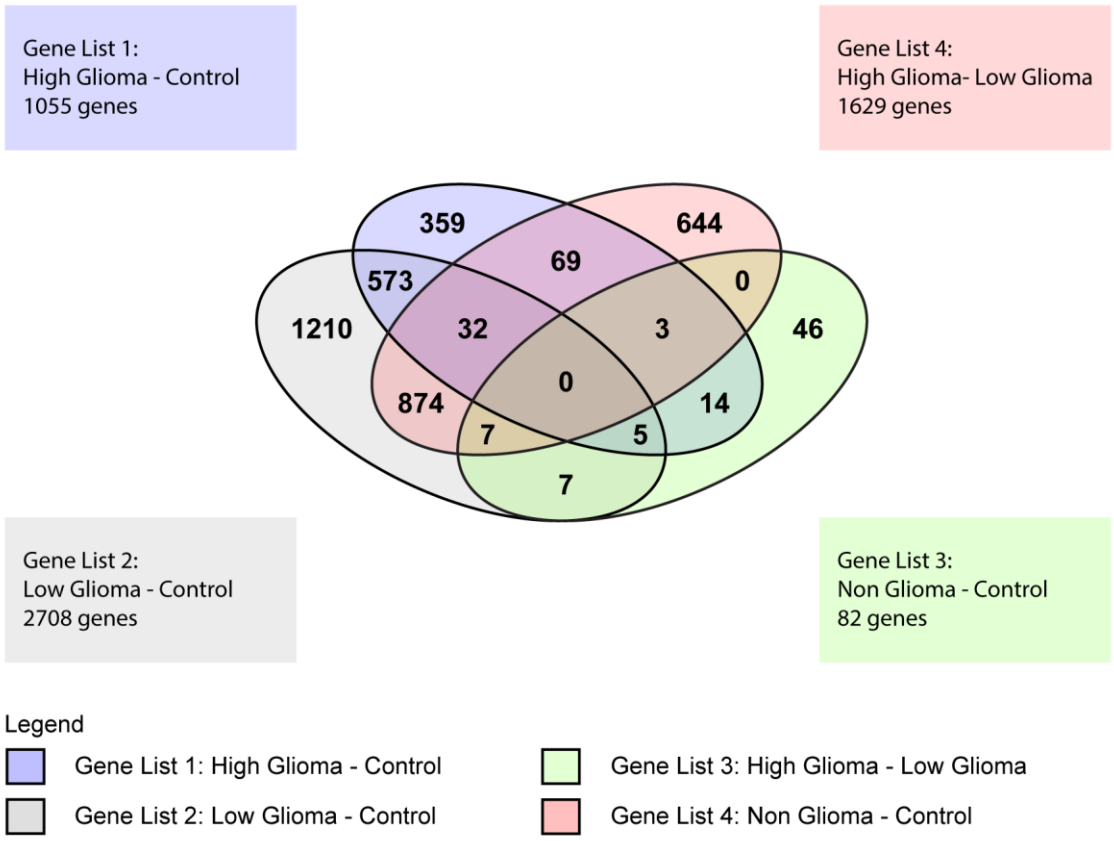


Figure 6: Venn diagram of differentially expressed genes for the different condition pairs. The Venn diagram summarizes the number of distinct and overlapping differentially expressed genes found in the four condition pairs: HG vs C (Gene List 1), LG vs C (Gene List 2), NG vs C (Gene List 3), and HG vs LG (Gene List 4).

Figure 7: Canonical pathways with corrected B-H p-value <0.05 identified by IPA for the four different condition pairs: (A) HG vs C, (B) LG vs C, (C) HG vs LG and (D) NG vs C. The significant pathways are shown along the x-axis of the bar charts. The ratio defined the proportion of differentially expressed genes from a pathway related to the total number of genes that make up that particular pathway (line graph, z-axis). The height of the bar represents the -log (B-H corrected p-value). The orange colour bar indicates a positive z-score (predicted activation), the blue colour bar indicates a negative z-score (predicted inhibition), the grey colour bar indicates no predicted activity and the white bar was given a z-score of zero. Colour intensity indicates prediction confidence (light to dark represents less to more confidence).

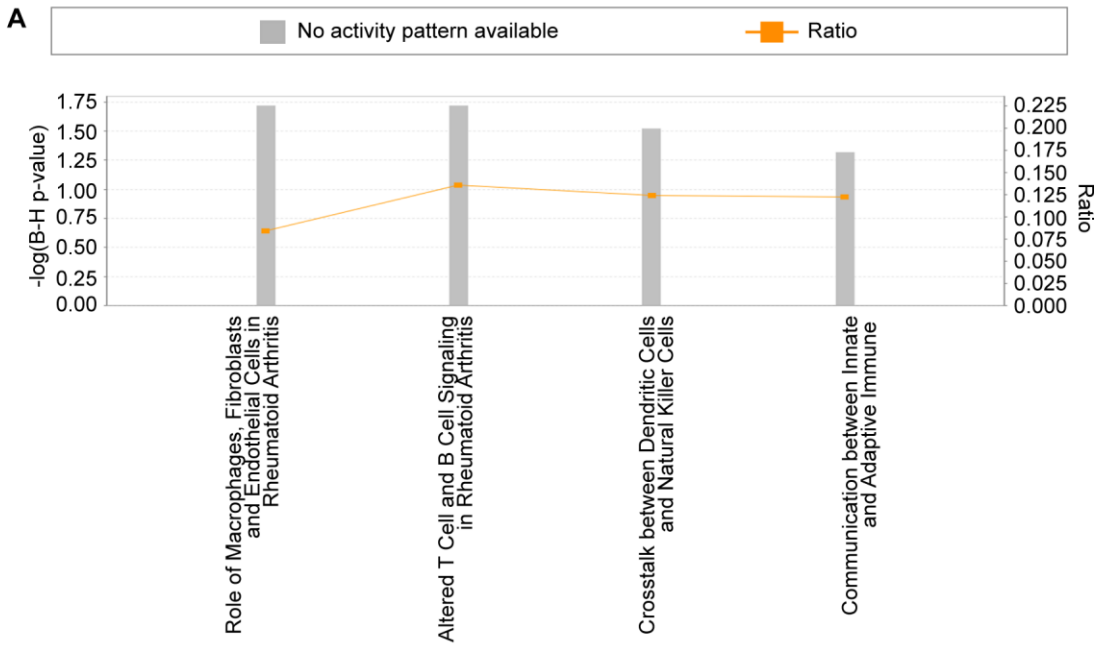


Figure 7(A): Canonical pathways with corrected B-H p-value <0.05 as identified by IPA for the HG vs C pair.

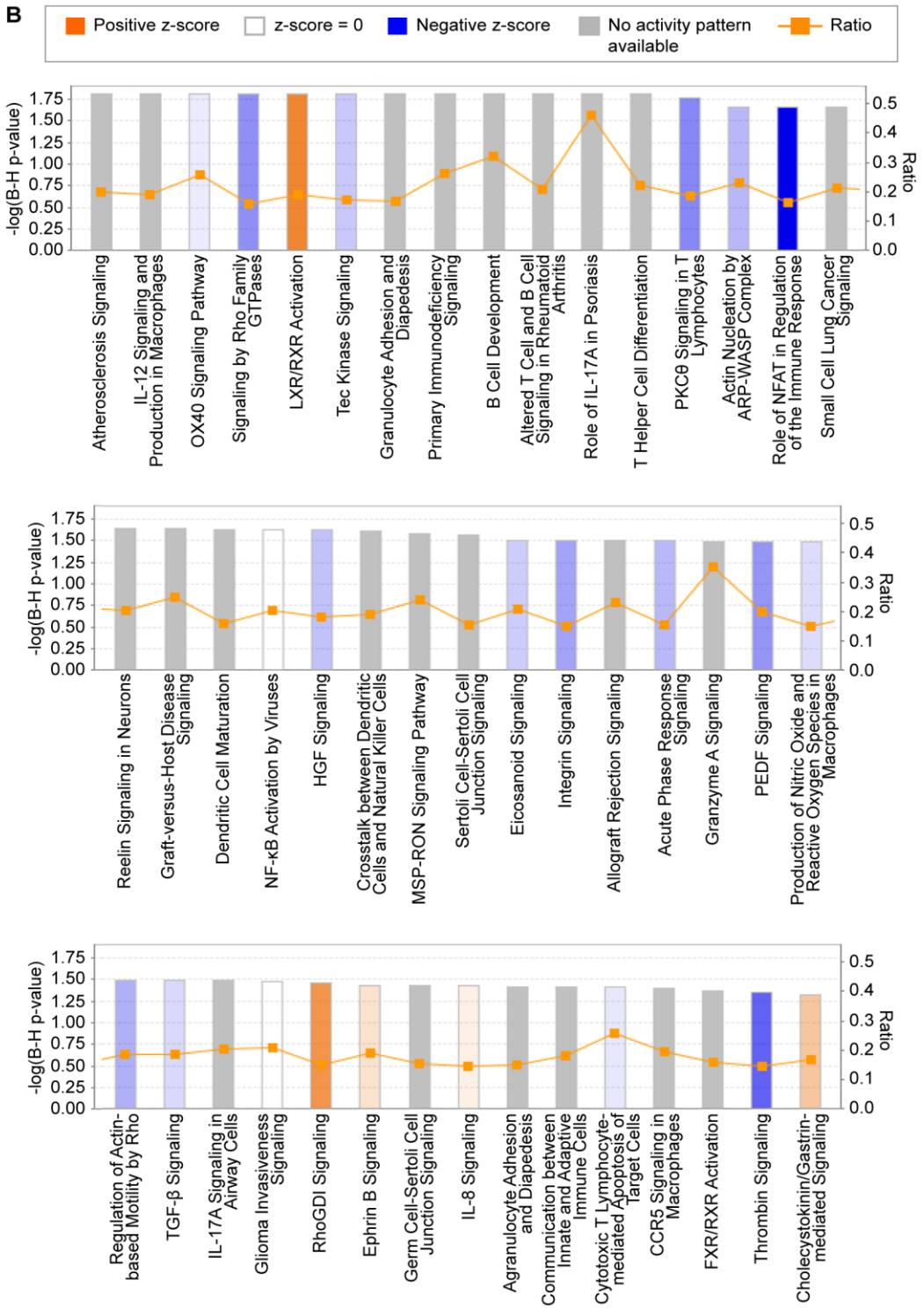


Figure 7(B): Canonical pathways with corrected B-H p-value <0.05 as identified by IPA for the LG vs C pair

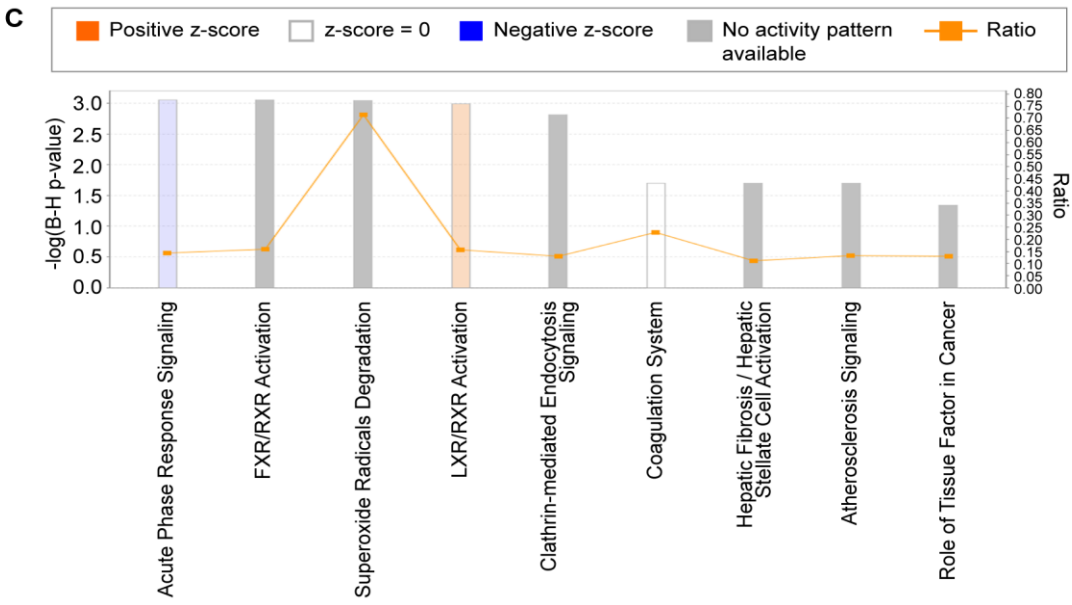


Figure 7(C): Canonical pathways with corrected B-H p-value <0.05 identified by IPA for the HG vs LG pair.

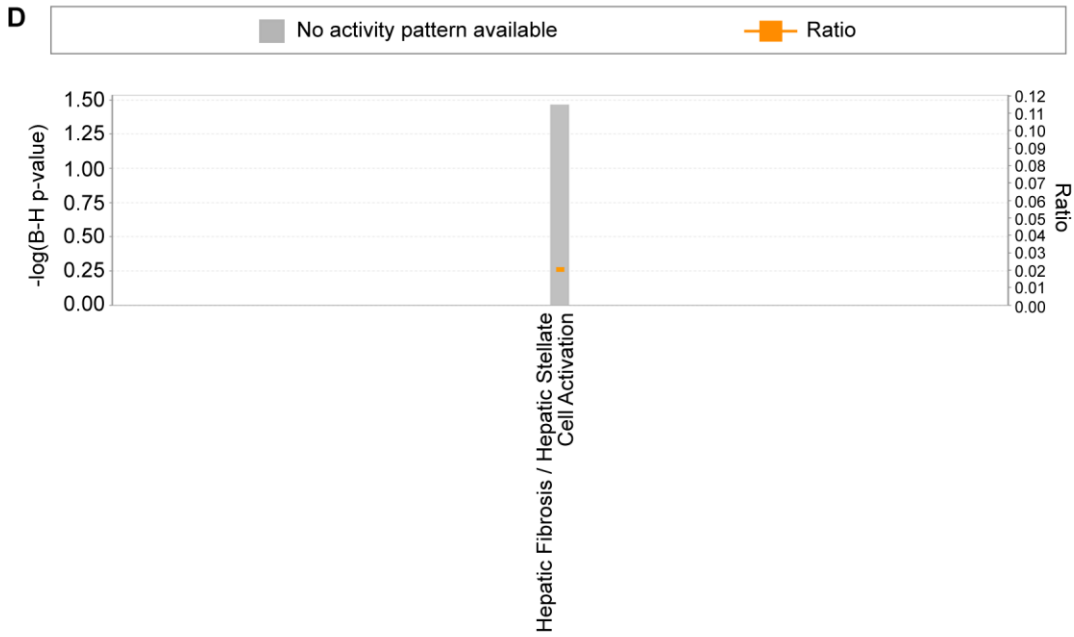


Figure 7 (D): Canonical pathways with corrected B-H p-value <0.05 identified by IPA for the NG vs C pair.

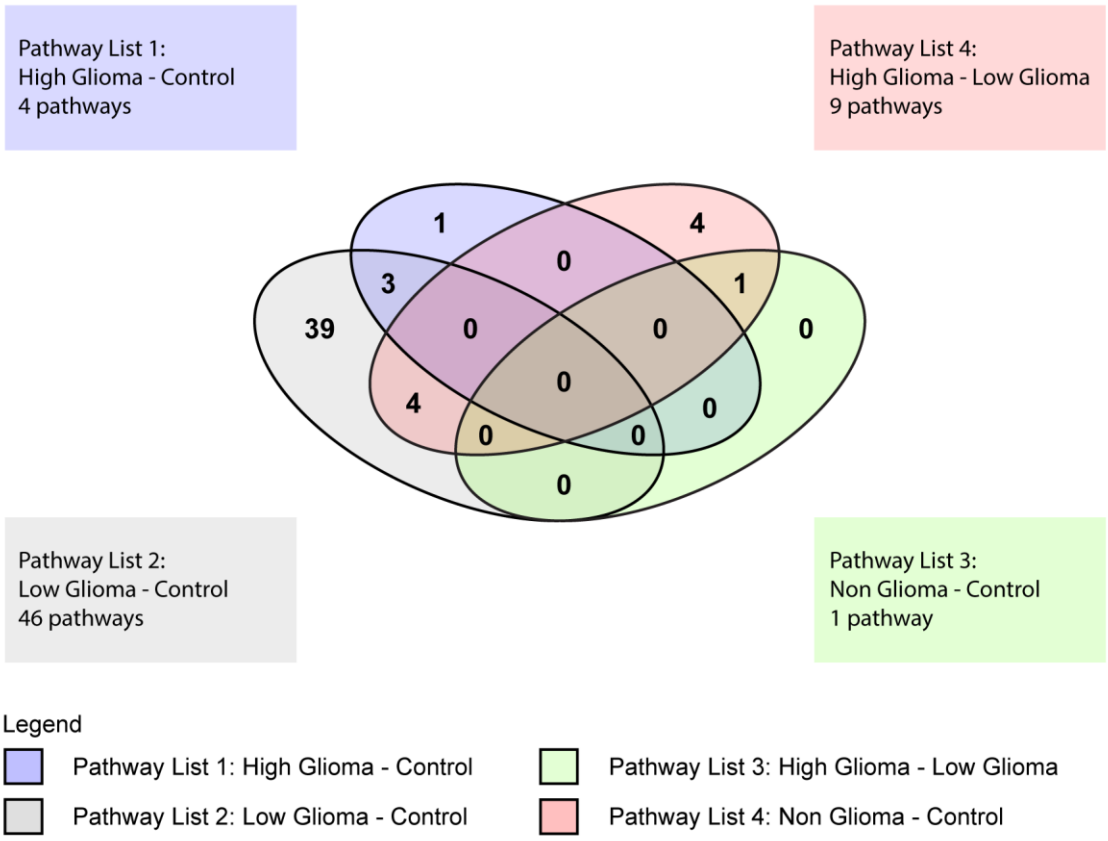


Figure 8: Venn diagram of canonical pathways. The Venn diagram summarizes the number of distinct and overlapping pathways found in the 4 condition pairs: HG vs C (Pathway List 1), LG vs C (Pathway List 2), NG vs C (Pathway List 3), and HG vs LG (Pathway List 4).

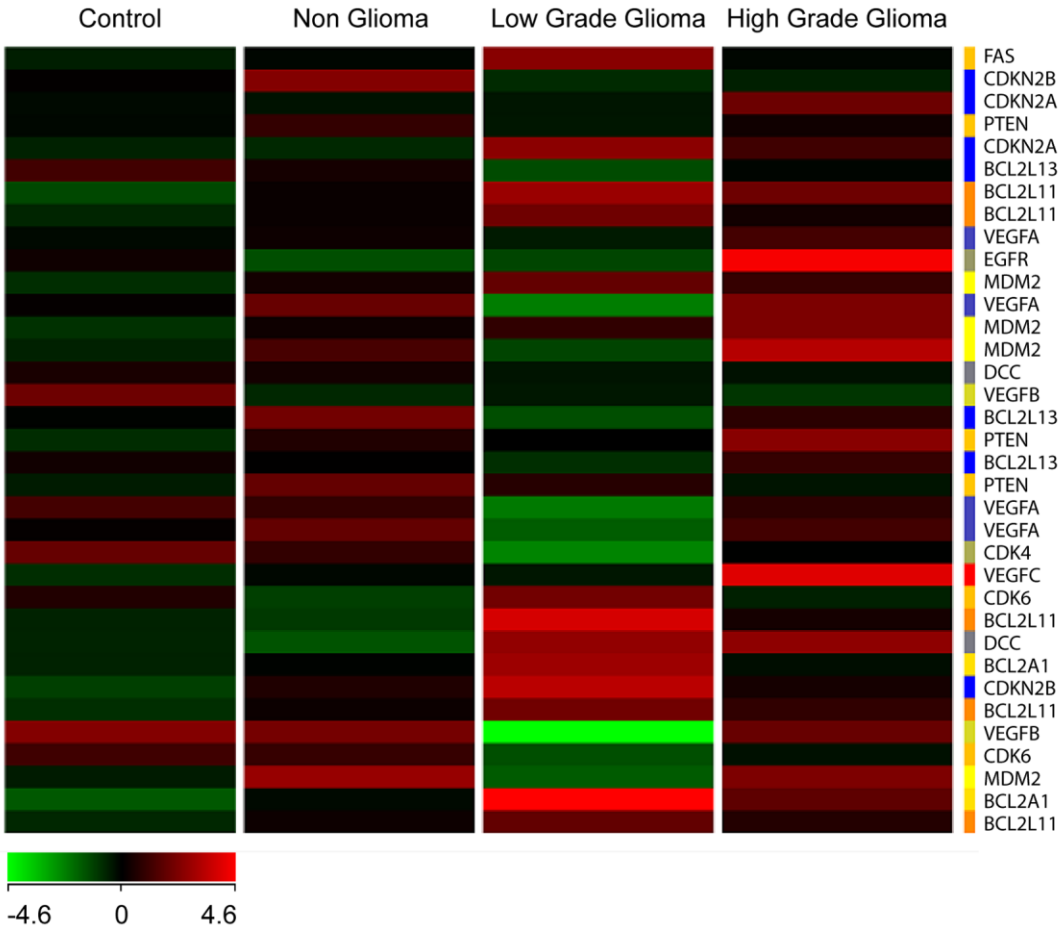


Figure 9: Heat Map of selected differentially expressed genes for the 4 different conditions.

Model	Median	Mean
1	0.11830	0.17220
2	0.110700	0.163200
3	0.11180	0.16360
4	0.1540	0.1882

Table 7: Median and mean adjusted r^2 values for models 1-4

Condition	HG vs C	LG vs C	HG vs LG	NG vs C
Total no. of genes	1055	2708	1629	82
Upregulated genes	479	713	1287	56
Downregulated genes	576	1995	342	26

Table 8: Summary list of significant differentially expressed genes for the 4 pairs of conditions

	HG vs C	HG vs LG	LG vs C	NG vs C
Total number of pathways	4	9	46	1

Table 9. No. of significant pathways for each pair of conditions with a corrected B-H p-value of <0.05

Table 10: List of pathways with a corrected B-H p-value <0.05 for the 4 condition pairs: (A) HG vs C, (B) HG vs LG, (C) LG vs C and (D) NG vs C

	HG vs C	HG vs LG	LG vs C	NG vs C
Actin Nucleation by ARP-WASP Complex			✓	
Acute Phase Response Signaling		✓	✓	
Agranulocyte Adhesion and Diapedesis			✓	
Allograft Rejection Signaling			✓	
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	✓		✓	
Atherosclerosis Signaling		✓	✓	
B Cell Development			✓	
CCR5 Signaling in Macrophages			✓	

Cholecystokinin/Gastrin-mediated Signaling			✓	
Clathrin-mediated Endocytosis Signaling		✓		
Coagulation System		✓		
Communication between Innate and Adaptive Immune Response Cells	✓		✓	
Crosstalk between Dendritic Cells and Natural Killer Cells	✓		✓	
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells			✓	
Dendritic Cell Maturation			✓	
Eicosanoid Signaling			✓	
Ephrin B Signaling			✓	
FXR/RXR Activation		✓	✓	
Germ Cell-Sertoli Cell Junction Signaling			✓	
Glioma Invasiveness Signaling			✓	
Graft-versus-Host Disease Signaling			✓	
Granulocyte Adhesion and Diapedesis			✓	
Granzyme A Signaling			✓	
Hepatic Fibrosis / Hepatic Stellate Cell Activation		✓		✓
HGF Signaling			✓	
IL-8 Signaling			✓	
IL-12 Signaling and Production in Macrophages			✓	
IL-17A Signaling in Airway Cells			✓	

Integrin Signaling			✓	
LXR/RXR Activation		✓	✓	
MSP-RON Signaling Pathway			✓	
NF-κB Activation by Viruses			✓	
OX40 Signaling Pathway			✓	
PEDF Signaling			✓	
PKCθ Signaling in T Lymphocytes			✓	
Primary Immunodeficiency Signaling			✓	
Production of Nitric Oxide and Reactive Oxygen Species in Marcrophages			✓	
Reelin Signaling in Neurons			✓	
Regulation of Actin-based Motility by Rho			✓	
RhoGDI Signaling			✓	
Role of IL-17A in Psoriasis			✓	
Role of Macrophages, Fibroblast and Endothelial Cells in Rheumatoid Arthritis	✓			
Role of NFAT in Regulation of the Immune Response			✓	
Role of Tissue Factor in Cancer		✓		
Sertoli Cell-Sertoli Cell Junction Signaling			✓	
Signaling by Rho Family GTPases			✓	
Small Cell Lung Cancer Signaling			✓	
Superoxide Radicals Degradation		✓		

Tec Kinase Signaling			✓	
TGF-β Signaling			✓	
T Helper Cell Differentiation			✓	
Thrombin Signaling			✓	

Table 11. Detailed list of pathways with corrected B-H p-value <0.05 and with significant differentially expressed genes for the 4 pairs of conditions.

Table 11(A): HG vs C

Ingenuity Canonical Pathways	-log(B-H p-value)	Genes
Role of Macrophages, Fibroblasts and Endothelial Cells in Rheumatoid Arthritis	1.72E00	CAMK4, FZD3, PRSS2, IL6R, CEBPD, PIK3R5TLR8, IL17RC, FZD9, IGHG1, IL6, IRAK3, IL1R1, CEBPE, FOS, TLR10, WNT10A, LEF1, TLR3, TNF, TNFSF13B, PRSS3, WNT5A, TRAF1
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	1.72E00	TLR10, HLA-DOA, SLAMF1, PRTN3, TLR8, CD79A, IL, TLR3, TNFRSF13C, TNF, TNFSF13B
Crosstalk between Dendritic Cells and Natural Killer Cells	1.52E00	HLA-G, KIR3DL1, PRF1, KIR2DL2, FSCN2, NCR3, KIR2DL4, IL6, TLR3,TNF, ITGAL
Communication between Innate and Adaptive Immune Cells	1.32E00	HLA-G, TLR10, CCL4, TLR8, IGHG1, IL6,TLR3,TNFRSF13C, TNF, TNFSF13B

Table 11(B): LG vs C

Ingenuity Canonical Pathways	-log(B-H p-value)	Genes
Atherosclerosis Signaling	1.82E00	APOE, CCR3, ABHD3, COL2A1, ALOX12, CCR2, IL6, NFKB2, IL1F10, SAA4, PLA2G2A, ALOXE3, ITGB2, COL1A1, ALB,IL18, ALOX15B, PLA2G4D, TGFB, SERPINA1, S100A8, TNF, RBP4, APOC3
IL-12 Signaling and Production in Macrophages	1.82E00	APOE, IL12RB1, MAF, PIK3R5, MST1, ALOX12, TGFB1, SERPINA1, S100A8, NFKBIB, PRKD1, PPARG, NFKB2, MAPK12, SAA4, PIK3R3, FOS, ALB, IL18, PIK3R6, MAP3K8, MST1R, TNF, APOC3, RBP4
OX40 Signaling Pathway	1.82E00	B2M, HLA-DOA, TNFRSF4, CD3E, CD4, NFKB2, MAPK12, BCL2, CD3G, HLA-DRB1, HLA-DQB2, FCER1G, HLA-DOB, NFKBIB
Signaling by Rho Family GTPases	1.82E00	CDC42EP5, GNA11, PIK3R5, GNG13, ARHGEF1, LIMK1, MAP3K10, CDH7, ITGA3, GNB3, PPP1R12B, GNA15, RHOU, MRAS, PIP5KL1, CDH13, RND2, SEPT14, CFL1, RHOC, ARPC5L, ARHGEF15, ITGA2, GNAI1, WASF1, NFKB2, GNAZ, MAPK12,MYL9,PIK3R3, FOS, RHOV, PIP5K1C, WAS, NCF2, PIK3R6, SEPT6
LXR/RXR Activation	1.82E00	APOE,ABCG5, NR1H4, C9, SERPINF,1 IL6NFKB2, IL1F10, SAA4, SERPINF2, HPR,IL18, ALB, CD14, S100A8, SERPINA1, NCOR1, PTGS2RXRB, TNF, AGT, RBP4, APOC3
Tec Kinase Signaling	1.82E00	GNA11, PIK3R5, GNG13, ITGA3, GNB3, GNA15, RHOU, MRAS, TNFRSF10A, PRKD1, RND2, STAT5A, RHOC, ITGA2, GNAI1, BMX, GNAZ, NFKB2, MAPK12, PIK3R3, FOS, RHOV, WAS, VAV3,

		FCER1G, PIK3R, TNF
Granulocyte Adhesion and Diapedesis	1.82E00	CLDN11, SELL, IL1F10, CXCL5, CLDN7, ITGB3, ITGA3, CLDN4, CCL3L3, CXCL1, CLDN9, CLDN10, HRH2, C5AR1, MMP28, ITGA2, GNAI1, ITGAL, ITGB2, CXCL16, IL18, CLDN5, CCL4, CXCL2, CX3CL1, TNF, CLDN3, HSPB1
Primary Immunodeficiency Signaling	1.82E00	IL7R, IL2RG, AIRE, CD3E, CD4, IGKC, IGLL1/IGLL5, ICOS, RAG1, CD79A, IGHG1, TNFRSF13C
B Cell Development	1.82E00	IL7R,HLA-DOA,HLADRB1,CD79B, IGKC, CD86,HLA-DOB, RAG1,CD79A
Altered T Cell and B Cell Signaling in Rheumatoid Arthritis	1.82E00	HLA-DOA, SLAMF1, CD79B, CD79A, IL6, IL1F10, NFKB2, TNFRSF13C, IL18, HLA-DRB1, TLR5, TGFB1, PRTN3, FCER1G, HLA-DOB,CD86,TNF
Role of IL-17A in Psoriasis	1.82E00	S100A7,IL17RC,S100A8,CXCL1,CXCL5DEFB4A/DEFB4B
T Helper Cell Differentiation	1.82E00	IL2RG,HLA-DOA, IL12RB1, IL6, BCL6, TBX21, IL18,HLA-DRB1, TGFB1, ICOS, FCER1G,HLA-DOB,CD86,GATA3,TNF
PKCθ Signaling in T Lymphocytes	1.77E00	HLA-DOA, CD3E, CD4, PIK3R5, NFKB2, PIK3R3, FOS, MAP3K10, CD3G, HLA-DRB1, GRAP2, VAV3, SOS1, MRAS, PIK3R6, FCER1G, HLA-DOB, CD86, MAP3K8, NFKBIB, MAP3K3
Actin Nucleation by ARP-WASP Complex	1.67E00	RND2, ARPC5L, RHOC, ITGA2, WASF1, NCK2, RHOV, ITGA3, PPP1R12B, WAS, SOS1, RHOU, MRAS
Role of NFAT in Regulation of the Immune Response	1.66E00	HLA-DOA, PLCB2, CD3E, CD4, GNA11, PIK3R5, GNG13, FCGR2B, HLA-DRB1 GNB3, GNA15, SOS1, MRAS, NFKBIB, CD79B, GNAI1, CD79A, GNAZ, NFKB2 ,PIK3R3, CD3G, FOS, MEF2D, FCER1G, PIK3R6, CD86, HLA-DOB

Small Cell Lung Cancer Signaling	1.66E00	TP53, NOS1, FHIT, PIK3R5, CDK4, NFKB2, BCL2, MYC, PIK3R3, PIK3R6, PTGS2, NFKBIB, RXRB, CDK2, TRAF1
Reelin Signaling in Neurons	1.65E00	APOE,ARHGEF15,ITGA2,PIK3R5,ARHGEF1,MAPK12, DAB1, ITGAL, ITGB3, PIK3R3, ITGB2, MAP3K10, ITGA3, CNR2,PIK3R6,MAP4K1
Graft-versus-Host Disease Signaling	1.65E00	PRF1,IL18,HLA-DOA,HLA-DRB1, GZMB,FCER1G,CD86,HLA-DOB, IL1F10, IL6, TNF
Dendritic Cell Maturation	1.63E00	B2M,HLA-DOA, PLCB2, IL32, PIK3R5, IL6, IL1F10, IGHG1, FCGR2B, CD1D, HLA-DRB1,CD1A,FSCN2,LY75,NFKBIB, COL2A1, NFKB2, MAPK12, PIK3R3, COL1A1, IL18, PIK3R6, FCER1G, HLA-DOB, CD86,TNF, CCR7
NF-κB Activation by Viruses	1.63E00	CCR5, CD4, ITGA2, PIK3R5, NFKB2, ITGAL, ITGB3, PIK3R3, ITGB2, ITGA3, PIK3R6, MRAS, NFKBIB, PRKD1,CR2
HGF Signaling	1.63E00	ITGA2, PIK3R5, IL6, MAPK12, MET, PIK3R3, FOS, MAP3K10, ITGA3, SOS1, PIK3R6, MRAS, MAP3K8, PTGS2, ELF5, MAP3K3, ELK3, PRKD1,CDK2
Crosstalk between Dendritic Cells and Natural Killer Cells	1.61E00	KIR3DL1, IL2RG, KIR2DL2, NCR3, IL6, NFKB2, ITGAL, CSF2RB, PRF1, IL18, HLA-DRB1, FSCN2, CD226, CD86, KIR2DL4, TNF, CCR7
MSP-RON Signaling Pathway	1.59E00	PIK3R3, CSF2RB, ITGB2, F12, KLK3, PIK3R6, PIK3R5 ,MST1 ,MST1R, CCR2, TNF
Sertoli Cell-Sertoli Cell Junction Signaling	1.57E00	NOS1, SPTBN1, CLDN11 ,CLDN7, TUBB2B, MAP3K10, ITGA3 ,CLDN4, CGN, MRAS, TUBB4A, JUP, CLDN9, CLDN10, GUCY1A3, TJP1, ITGA2, YBX3, MAPK12, CLDN5, WAS, MAP3K8, ACTN4, A2M, MAP3K3, TNF, CLDN3

Eicosanoid Signaling	1.5E00	ALOX15B, PLA2G4D, PTGIR, DPEP3, LTB4R2, PTGDR, ABHD3, TBXA2R, ALOX12, PTGS2, PTGDS, TBXAS1, PLA2G2A
Integrin Signaling	1.5E00	MYLK2, PIK3R5, ITGB3, NCK2, PARVB, ITGA3, PPP1R12B, ITGA11, SOS1, RHOU, MRAS, ITGB4, RND2, CAPN6, RHOC, ARPC5L, ITGA2, RALB, BCAR3, ITGAL, PIK3R3, MYL9, CAPN8, ITGB2, RHOV, WAS, PIK3R6, ACTN4, CTTN
Allograft Rejection Signaling	1.5E00	B2M, PRF1, HLA-DOA, HLA-DQB2, HLA-DRB1, GZMB, FCER1G, CD86, HLA-DOB, IGHG1,TNF
Acute Phase Response Signaling	1.5E00	FN1, C9, SERPINA3, IL1F10, IL6, RBP1, C1RSOS1, MRAS, OSMR, SERPINA1, NFKBIB, AGT, SERPINF1, NFKB2, MAPK12, SAA4, SERPINF2, PIK3R3, FOS, ALB, IL18, HP, A2M, TNF, RBP4
Granzyme A Signaling	1.49E00	GZMA, HIST1H1A, PRF1, HIST1H1E, HIST1H1D, H1F0
PEDF Signaling	1.48E00	TP53, PPARG, GDNF, SERPINF1, PIK3R5, NFKB2, MAPK12, BCL2, PIK3R3, PIK3R6, MRAS, DOCK3, NFKBIB, CASP8
Production of Nitric Oxide and Reactive Oxygen Species in Macrophages	1.48E00	APOE PIK3R5, MAP3K10, HOXA10, RHOU, SERPINA1 ,S100A8, NFKBIB, PRKD1, RND2, RHOC, PPP1R14A, NFKB2, MAPK12, SAA4, PIK3R3, FOS, ALB, RHOV, CAT, NCF2, PIK3R6, MAP3K8, MAP3K3,TNF, RBP4APOC3
Regulation of Actin-based Motility by Rho	1.48E00	RND2, PFN1, CFL1, RHOC, ARPC5L, ITGA2, WASF1, LIMK1, MYL9, RHOV, ITGA3, PPP1R12B, WAS, PIP5K1C ,RHOU, PIP5KL1
TGF-β Signaling	1.48E00	FOXH1, SKI, SMAD6, ACVR1 ,INHBC, MAPK12, INHBB, ACVR1B, BCL2, FOS, PIAS4, AMH, TGFB1, SOS1, MRAS,

		MAP4K1
IL-17A Signaling in Airway Cells	1.48E00	PIK3R3, PIK3R6, PIK3R5,MUC5AC, IL17RC, MUC5B, CXCL1, IL6, NFKB2, CXCL5, NFKBIB, MAPK12, DEFB4A/DEFB4B
Glioma Invasiveness Signaling	1.47E00	RND2, PIK3R3, TIMP3, RHOV, TIMP4, RHOC, PIK3R6, MRAS, PIK3R5, RHOU, PLAU, ITGB3
RhoGDI Signaling	1.46E00	GNA11, GNG13, ARHGEF1, LIMK1, ITGA3, CDH7, PPP1R12B, GNB3, GNA15, RHOU, MRAS, PIP5KL1, CDH13, RND2, CFL1, ARPC5L, RHOC, ARHGEF15, ITGA2, GNAI1, WASF1, GNAZ, MYL9, DGKZ, RHOV, PIP5K1C
Ephrin B Signaling	1.44E00	EPHB4, RGS3, CFL1, GNA11, GNAI1, GNG13, GNAZ, LIMK1, NCK2, EFNB2, GNB3, GNA15, VAV3, MRAS
Germ Cell-Sertoli Cell Junction Signaling	1.43E00	RND2, CFL1, RHOC, TJP1, ITGA2, PIK3R5, MAPK12, TUBB2B, LIMK1, PIK3R3, MAP3K10, RHOV, ITGA3, TGFB1, MRAS, PIK3R6, RHOU, MAP3K8, TUBB4A, JUP, ACTN4, TNF, MAP3K3, A2M
IL-8 Signaling	1.42E00	PLCB2, PIK3R5, GNG13, ITGB3, LIMK1, BCL2, GNB3 ,RHOU, MRAS, CXCL1, NFKBIB, PRKD1, CR2, RND2, RHOC ,GNAI1, IRAK3, MAPK12, AZU1, PIK3R3, MYL9, ITGB2, FOS, RHOV, NCF2, PIK3R6, PTGS2
Agranulocyte Adhesion and Diapedesis	1.41E00	CLDN11 ,SELL, FN1, IL1F10, CXCL5, CLDN7ITGA3, CLDN4, CCL3L3, CXCL1, CLDN9, CLDN10, C5AR1, MMP28, ITGA2, GNAI1, MYL9, ITGB2, CXCL16, IL18, CLDN5, CCL4, CXCL2, CX3CL1, TNF, CLDN3
Communication between Innate and Adaptive Immune	1.41E00	B2M, CD4, IL1F10, IGHG1, IL6, TNFRSF13C, IL18,HLA-DRB1, CCL4,

Cells		TLR5, CCL3L3, FCER1G, CD86, TNF, CCR7
Cytotoxic T Lymphocyte-mediated Apoptosis of Target Cells	1.41E00	B2M, CD3G, PRF1, CD3E, GZMB, FCER1G, CASP8, BCL2
CCR5 Signaling in Macrophages	1.4E00	FOS,CD3G,CCR5,CCL4,GNB3,CD3E, CD4,FCER1G, MRAS, GNAI1, GNG13, MAPK12,PRKD1
FXR/RXR Activation	1.36E00	PPARG,APOE,ABCG5,CYP27A1, NR1H4 C9, SERPINF1, IL1F10, SAA4, MAPK12, SERPINF2, HPR, ALB, IL18, FABP6, SERPINA1, TNF, RBP4, AGT, APOC3
Thrombin Signaling	1.35E00	PLCB2, CAMK1, GATA1, GNA11, PIK3R5, GNG13, ARHGEF1, PPP1R12B, GNB3, GNA15, SOS1, MRAS, RHOU, PRKD1, RND2, RHOC, ARHGEF15, GNAI1, GNAZ, NFKB2, MAPK12, PIK3R3, MYL9, RHOV, ADCY1, PIK3R6, GATA3
Cholecystokinin/Gastrin-mediated Signaling	1.33E00	RND2, PLCB2, RHOC, EPHA4, IL1F10, CCKMAPK12, FOS, RHOV, IL18, MEF2D, SOS1, RHO, MRAS, PTGS2, TNF, PRKD1

Table 11(C): HG vs LG

Ingenuity Canonical Pathways	-log(B-H p-value)	Genes
Acute Phase Response Signaling	3.05E00	ITIH3, TTR, FN1, APOH, AMBP, SERPINF1, IL6, SAA4, MAPK12, F2, SERPINF2, SERPIND1, PIK3R3, ALB, APOA1, ITIH2, TF, SOS1, SERPINA1, OSMR, CRABP2, A2M, AGT, RBP4
FXR/RXR Activation	3.05E00	ABCG5, TTR, APOB, CYP27A1, APOH, AMBP, SERPINF1, APOC2, SAA4, MAPK12, SERPINF2, CYP8B1, APOL1, ALB, APOA1, TF, APOC1, SERPINA1, AGT, RBP4

Superoxide Radicals Degradation	3.04E00	TYRP1, CAT, NQO1, CYGB, SOD3
LXR/RXR Activation	2.99E00	ABCG5, TTR, APOB, APOH, AMBP, SERPINF1, APOC2, IL6, SAA4, SERPINF2 ,APOL1, ALB, APOA1, TF, APOC1, SERPINA1, NCOR1, RBP4, AGT
Clathrin-mediated Endocytosis Signaling	2.81E00	MYO6, AP2A1 UBB, APOB, PICALM, ARPC5L,RAB7A,PIK3R5, APOC2, SAA4, F2, VEGFA, APOL1, PIK3R3, ARRB2, ALB, APOA1, TF, PIP5K1C, AAK1, APOC1, SERPINA1, ITGB4, RBP4
Coagulation System	1.7E00	SERPINA5, SERPINA1, F7, A2M, SERPINF2, F2 ,PLAT ,SERPIND1
Hepatic Fibrosis / Hepatic Stellate Cell Activation	1.7E00	VCAM1, FN1, MYH14, IL6, BAX, BCL2, COL1A2, COL16A1, VEGFA, COL5A1, CXCL3, COL1A1, IGF2, COL6A1, COL13A1, TGFB1, MYH3, EDNRA, A2M, COL7A1, EGFR, AGT
Atherosclerosis Signaling	1.7E00	VCAM1, APOB, APOC2, IL6, SAA4, PRDX6, APOL1, COL1A2, COL1A1, ALB, ALOX15B, APOA1, TGFB1, APOC1, SERPINA1, RBP4
Role of Tissue Factor in Cancer	1.34E00	STAT5A, CFL1, PIK3R5, F7, MAPK12, F2, LIMK1, VEGFA, PIK3R3, ARRB2, ITGA3, CXCL1 ,CYR61, EGFR

Table 11(D): NG vs C

Ingenuity Canonical Pathways	-log(B-H p-Value)	Genes
Hepatic Fibrosis / Hepatic Stellate Cell Activation	1.46E00	IL1R2, IL1R1 ,COL9A2, MMP9

Condition	Regulation of expression	Genes	Fold change (from GeneSpring analysis)
NG vs C	Upregulated	MMP9	2.35
LG vs C	Upregulated	MAP3K8	2.46
LG vs C	Downregulated	TP53 SOS1	2.81 2.62
HG vs C	Upregulated	FOS IL6	2.28 4.06
HG vs C	Downregulated	TNF	2.90
HG vs LG	Upregulated	EGFR VEGFA	2.44 2.13
HG vs LG	Downregulated	MAPK12	4.09

Table 12: Final list of genes selected for validation

Condition	Gene	Fold change from GeneSpring	Fold change from ddPCR	p-value	Bonferroni correction: new 0.05 threshold, 10 tests	Result (p-value)
NG vs C	**MMP9**	+2.35	+6.49	0.0068	0.005	**False**
LG vs C	**MAP3K8**	+2.46	+1.61	0.00003	0.005	**True**
	TP53	-2.81	+2.00	0.00007	0.005	**True**
	SOS1	-2.62	-1.69	0.00362	0.005	**True**
HG vs C	**FOS**	+2.28	+3.55	0.00853	0.005	**False**
	IL6	+4.06	+3.05	0.00001	0.005	**True**
	TNF	-2.90	+1.60	0.00620	0.005	**False**
HG vs LG	**EGFR**	+2.44	-1.25	0.43	0.005	**False**
	VEGFA	+2.13	+1.36	0.24	0.005	**False**
	MAPK12	-4.09	+1.19	0.27	0.005	**False**

Table 13: List of significant genes after Bonferroni correction

CHAPTER 4: DISCUSSION

This study has advanced the idea of using blood-based gene expression studies as an indicator of neoplastic changes occuring in brain tissue. This idea was based upon the sentinel principle and extrapolated to the study of brain tumours. However, in the very first study, the sentinel principle was used to identify subjects with an increased risk of colorectal cancer compared to normal subjects (74). In this study, we have used the sentinel principle not only to identify patients with a glioma but also to differentiate between high grade, low grade, non-glioma and control subjects.

The unsupervised hierarchical clustering and principal component analysis clearly showed that the four groups of subjects clustered into 3 statistically significant groups as represented by the ellipses, which showed a distinct directionality in the different groups based on similarities in gene expression (see Results Section, Fig.4). The fact that the non-glioma and control subjects clustered together and were distinct from the high and low grade tumour patients, indicated that the changes in gene expression in blood in these 2 groups were clearly different from that of the glioma patients indicating specificity of expression. This lends further credence to the sentinel principle that substances are released from the tumour into the bloodstream (67,68) and may be distinct for each tumour subtype. Although the blood samples in our study were taken from patients after presentation to the hospital with neurological symptoms, it is highly likely that these substances were released during the early stages of tumour formation (67) and continued to persist in blood even as the tumour enlarged based upon the theory and evidence from the sentinel principle (67,68,74-76).

The brain, as an immunologically privileged site, is protected by the blood-brain barrier which restricts the movement of water soluble molecules by tight junctions (116) and a low level of transcytosis (117). The breakdown of the blood-brain barrier in brain tumours can be visualized by either freeze fracture electron microscopy (118) or contrast enhanced magnetic resonance imaging (MRI) using gadolinium (119). The normal blood-brain barrier is impermeable to contrast medium but there is a gradual increase in the degree of disruption of the blood-brain barrier corresponding to the grade of the tumour. WHO grade II tumours show little or no contrast medium enhancement, WHO grade III tumours enrich more contrast medium than grade II tumours while WHO grade IV tumours (GBM) show the greatest gadolinium enhancement (100). This observation fits well with our postulation that substances from the brain are able to cross the blood-brain barrier and enter the circulation due to the varying degrees of disruption of the blood-brain barrier during glioma formation.

The differentially expressed genes for the four different conditions were unique, but also had some commonality. Most of the unique and common genes in the HG and LG tumour samples were transcription factors, cytokines, proto-oncogenes, oncogenes, growth factors and tumour suppressor genes. These genes are involved in inflammation, tumour signalling pathways, glioma formation, tissue necrosis, apoptosis, homeostasis, cytoskeletal architecture, maintenance of the extracellular matrix and determination of cell fate. Interestingly, there were also a substantial number of genes involved in the innate and adaptive immune system suggesting that modulation of the immune system plays a critical role in tumour response. In addition, genes known to be involved in the pathogenesis of GBM were also upregulated in blood. These genes included *EGFR,*

VEGF and *IL-6*. This evidence implied that some of the changes occurring in the tumour tissue may be reflected in blood, suggesting that these substances may be released into the circulation through disruption of the blood-brain barrier or through complex signalling mechanisms.

The canonical pathways for the 4 sets of conditions mirrored to a certain extent the differential gene expression pattern. These included pathways involved in the innate and adaptive immune response, interleukin, acute phase response, glioma invasiveness, NF-κB activation and TGF-β signalling. The latter 3 pathways are also involved in the pathogenesis of gliomas. Again, we see much commonality between the signalling pathways in tissue and blood taken from glioma patients. One of the reasons for this could be the fact that peripheral blood cells share more than eighty percent of the transcriptome with 9 different tissue types including brain (67). More importantly, is the fact that blood cells express organ specific genes and also genes that are responsive to physiological changes and stimuli that were previously thought to be exclusive to certain tissue types (67,120). In the pathogenesis and formation of gliomas, these interactions between blood and tissue, together with disruption of the blood-brain barrier, could possibly explain some of the similarity observed in gene expression between gliomas and peripheral blood cells.

However, the results have to be viewed with caution too. The canonical pattern is calculated based on the activation state of one or more key genes when the pathway is activated. It also depends on the causal relationships with each other (activation and inhibition edge between the genes based on findings in the literature) to generate activity

patterns for the genes and also the end-point function. In the HG vs C pair (Figure 6A), there were only 4 pathways that passed the corrected B-H p-value <0.05 as identified by IPA. In addition, a z-score could not be assigned to any of these 4 pathways thus no prediction could be made about them. However, they may have biological significance as the majority of these pathways are involved in the innate and adaptive immune responses that are altered in cancer.

Another weakness with the IPA is that one gene may be activating and inhibiting at the same time, albeit more commonly on different targets in the same pathway. This may result in ambiguous data from the 2 conflicting gene expression values. If the target were the same, the effects would cancel out if the magnitude of expression were the same, if not, the effect with the larger expression would predominate. This would result in inaccurate, incomplete and a masked effect on gene expression.

The validation of selected genes was done by ddPCR. As mentioned previously, these genes were selected because they were known to be involved in signalling pathways that played an important role in tumourigenesis including the pathogenesis and formation of gliomas. In the selection of 10 genes for validation, 4 of the 10 genes, namely *TP53*, *TNF*, *MAPK12* and *EGFR* showed fold changes that were reversed to that seen in the microarray experiment. *TP53*, *TNF* and *MAPK12* were downregulated in the microarray experiment but upregulated by validation and *EGFR* was upregulated in the microarray experiment but minimally downregulated by validation. The reason for this could be multifactorial. Firstly and most importantly, the probes used for the microarray experiment are different from the primers used in ddPCR. As genes very commonly have

isoforms, it is likely that the primers in ddPCR may be amplifying an isoform of the gene resulting in alternative transcripts. These transcripts may have expression levels that are different from the parent gene. In addition, there may be a negative feedback loop where one transcript inhibits the expression of the alternative transcript of the same gene or *vice versa*. This could result in reversal of expression as seen during ddPCR validation. Secondly, we selected *GUSB* as the housekeeping gene to normalize our ddPCR data. Although *GUSB* showed the least variation with samples compared to *TBP* and *HuPO,* it might still have shown some variation in gene expression in the tumour samples. This could result in reversal of gene expression after validation. Thirdly, microarray analysis is generally used to screen large numbers of genes and the possibility arises that there may be false positives. Fourthly, human samples have huge technical and biological variability and it is likely that the presence of substances such as activators or inhibitors within the samples could be contributing to the differences observed. This is because ddPCR, being far more sensitive and quantitative, is able to detect the expression of genes affected by either inhibitors or activators, that may not be detected by microarray analysis. Fifthly, not all samples were used for validation by ddPCR. Only 3 to 6 samples were used for each set of conditions and this may have affected the level and pattern of gene expression as well.

The initial p-values obtained showed that 7 of the 10 genes chosen for validation had statistically significant p-values <0.05. The genes with initial p-values >0.05 were *EGFR, VEGFA* and *MAPK12.* After applying the Bonferroni correction for the p-value, only 4 of the 10 genes passed this stringent statistical test. The 4 genes were *MAP3K8, IL6, SOS1* and *TP53*. Although the other genes were not considered to be statistically

significant, they could be clinically significant. In addition, p-values are dependent on many factors including sample size, with a larger sample size giving rise to a more reliable p-value (121). In our case with a limited sample size, the p-value could vary by adding or removing even one value. Thus a larger sample size would definitely add more confidence to the p-values that were obtained in our experiments.

Besides the confidence needed in p-values to take forward a panel of genes to establish a reliable biomarker, other factors are also important. A biomarker for a disease has to be highly sensitive and specific. It must also be reliable and reproducible. A biomarker should be able to screen individuals in the preclinical stage of the illness, in other words it should be able to detect individuals at risk for developing the disease. It should also be diagnostic and prognostic, and able to detect changes in the disease state after treatment or recurrence of the disease in cases of relapse (122-124). Therefore, do gliomas lend themselves to be suitably detected by a blood-based biomarker?

In order to answer this question, we need to look at several variables. Firstly, for a biomarker to be used as a screening test in the general population, the incidence of the disease should not be extremely low as we would need to screen an excessively large number of subjects and this would not be cost effective. Gliomas have a low incidence in the population, in the range of 2-3 new cases per 100,000 population per year (115). However, if we take into account all the genetic syndromes that involve brain tumours in known high risk cancer families such as neurofibromatosis types 1 and 2, von Hippel-Lindau disease, tuberous sclerosis complex, Li-Fraumeni syndrome and Turcot syndrome, although individually rare causes of brain tumours, but collectively could be

important and add to the increased incidence of gliomas. Thus in these cases, a simple, non-invasive blood test which involves RNA profiling in whole blood, can be used as an addition to the more traditional methods of cancer screening and detection, which in these syndromes and disease complexes, involves genetic testing.

Secondly, most patients with gliomas especially those with primary GBM, present to the hospital after a relatively short history of symptoms which commonly includes headache, vomiting, seizures, and motor and sensory loss. This was also the case for all the high and low grade glioma patients involved in our study. This is because the tumour has enlarged to the point that it has compressed on brain structures and affected particular regions of the brain involved in specific neurological function. So, it would be reasonable to assume that detection of a small tumour which would be far more amenable to complete surgical resection, would portend to a better prognosis. No one has shown thus far, that patients with a small tumour have a better prognosis simply because the vast majority of patients with gliomas, including the patients in our study, present to hospital when they have neurological symptoms. At the time of presentation, the tumour is relatively large in size and has diffusely invaded into surrounding brain tissue, making complete surgical resection and a cure virtually impossible. In addition, small tumours are less likely to be diffuse and thus less infiltrative into surrounding brain tissue. In many cases, early symptoms of brain tumours are likely to be mild and very non-specific, so a means of stratifying patients for further investigations such as MRI, could be determined by using a blood-based biomarker. Conversely, if a patient had an MRI for some other reason and a suspicious area was detected suggesting a brain tumour, a watch and wait approach could be instituted. The patient could be monitored using a blood-

based biomarker or followed by MRI, although the frequency of initiating such investigations may be difficult to determine. Besides using the blood-based biomarker panel as a screening tool, we could also observe how the differential expression of genes within the panel change in response to treatment or recurrence of the disease. This is an area that needs to be explored further in our study.

One of the criteria mentioned earlier about a reliable biomarker is the fact that it should be able to diagnose patients at the preclinical stage of the illness, in other words it should be able to detect individuals at risk for developing the disease. Recently, a group from the Ohio State University, Columbus, Ohio, USA have found changes in certain cytokines and other immune system components that are altered within 5 years of a patient developing GBM. There were decreased interactions among cytokines, namely between *IL4* and *sIL4RA* within 5 years of the patient being diagnosed with a glioma or GBM. This weakened signalling indicated that the tumour was beginning to suppress the local immune system, thus predisposing the patient to the development of cancer (125). In our case, the differential gene expression pattern observed in the high and low grade tumour samples, could be used to predict the risk or probability of a person developing a glioma based on the sentinel principle.

There are limitations to the use of the sentinel principle and that of a blood-based biomarker to detect changes in a disease state in another tissue. The main limitation is that the blood transcriptome is susceptible to a vast array of changes such as that due to tobacco smoke, environmental pollutants and toxins, and to diseases such as hypertension, diabetes, cardiovascular disease, ischaemic stroke and asthma (69-72, 126-

128). Many cancer patients, including the patients in our study, have these comorbidities and this could have a confounding effect on the differential gene expression pattern observed. In addition, the drawing of blood, temperature and storage conditions can all have an effect on gene expression levels of peripheral blood cells.

This is a preliminary study to assess the possibility of using a blood-based biomarker to differentiate between high grade, low grade, non-glioma and control samples and therefore the results must be interpreted with caution. The main drawback of this study is the small sample size. In order to take this study forward to a blood-based biomarker panel for gliomas, we would need a much larger sample size to give this study more power and to obtain more reliable p-values for the genes selected.

This study would also need to be validated in an independent data set. The purpose of this would be to estimate how accurately our blood-based biomarker (training set) will perform in practice. Therefore, the inclusion of a test set would be needed to "test" our blood-based biomarker in the training phase to limit problems like overfitting. Ultimately, this would give us an insight on how our biomarker panel will perform in an independent data set.

Finally, the data in this thesis will be freely available. As the sample number, n, in this study is small, this will enable those who are interested to verify the results of this study, to use the data as a starting point. They may wish to replicate this study using a similar or larger sample size.

REFERENCES

1. World Cancer Report 2014

2. Ferlay J, Soerjomataram I, Ervik M, Dikshit R, Eser S, Mathers C, et al. Cancer incidence and mortality worldwide. In: International Agency for Research on Cancer (IARC) CancerBase 2012;11

3. US Mortality Data, 2006. National Centre for Health Statistics. Centres for Disease Control and Prevention 2009.

4. Linet MS, Ries LA, Smith MA, Tarone RE, Devessa SS. Cancer surveillance series: recent trends in childhood cancer incidence and mortality in the United States. J Natl Cancer Inst 1999 Jun 16;91(12):1051-8.

5. Louis DN, Ohgaki H, Wiestler OD, Cavenee WK, Burger PC, Jouvet A et al. The 2007 WHO classification of tumours of the central nervous system. Acta Neuropathol 2007; 114:97-109.

6. Smith JS, Jenkins RB. Genetic alterations in adult diffuse glioma: occurrence, significance, and prognostic implications. Frontiers Biosci 2000;5:213-31.

7. Bello MJ, Leone PE, Vaquero J, de Campos JM, Kusak ME, Sarasa JL, et al. Allelic loss at 1p and 19q frequently occurs in association and may represent early oncogenic events in oligodendroglial tumours. Int J Cancer 1995;64:207-10.

8. Reifenberger J, Reifenberger G, Liu L, James CD, Wechsler W, Collins VP. Molecular genetic analysis of oligodendroglial tumours shows preferential allelic deletions on 19q and 1p. Am J Pathol 1994;145:1175-90.

9. Cairncross JG, Ueki K, Zlatescu MC, Lisle DK, Finkelstein DM, Hammond RR, et al. Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. J Natl Cancer Inst 1998;90:1473-79.

10. Bauman GS, Ino Y, Ueki K, Zlatescu MC, Fisher BJ, Macdonald DR, et al. Allelic loss of chromosome 1p and radiotherapy plus chemotherapy in patients with oligodendrogliomas. Int J Radiat Oncol Biol Phys 2000;48:825-30.

11. Iwamoto FM, Nicolardi L, Demopoulos A, Barbashina V, Salazar P, Rosenblum M, et al.Clinical relevance of 1p and 19q deletion for patients with WHO grade 2 and 3 gliomas.J Neurooncol 2008;88:293-98.

12. Watanabe T, Nobusawa S, Kleihaus P, Ohgaki H. IDH1 mutations are early events in the development of astrocytomas and oligodendrogliomas. Am J Pathol 2009;174:1149-53.

13. Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science 2014;343:189-93.

14. Parsons DW, Jones S, Zhang X, Lin JC, Leary RJ, Angenendt P, et al. An integrated genomic analysis of human glioblastoma multiforme. Science 2008;321:1807-12.

15. Ohgaki H, Kleihues P. Genetic pathways to primary and secondary glioblastoma. Am J Pathol 2007;170:1445-53.

16. Ichimura K, Pearson DM, Kocialowski S, Backlund LM, Chan R, Jones DT, et al. IDH1 mutations are present in the majority of common adult gliomas but rare in primary glioblastomas. Neuro-Oncology 2009;11:341-47.

17. Zhang X, Yang H, Gong B, Jiang C, Yang L. Combined gene expression and protein interaction analysis of dynamic modularity in glioma prognosis. J Neurooncol 2012;107:281-88.

18. Idbaih A, Criniere E, Ligon KL, Delattre O, Delattre JY. Array-based genomics in glioma research. Brain Pathol 2010;20:28-38.

19. Malzkorn B, Wolter M, Liesenberg F, Grzendowski M, Stuhler K, Meyer HE, et al. Identification and functional characterization of microRNAs involved in the malignant progression of gliomas. Brain Pathol 2010;20:539-50.

20. Ekstrand AJ, James CD, Cavenee WK, Seliger B, Pettersson RF, Collins VP. Genes for epidermal growth factor receptor, transforming growth factor alpha, and epidermal growth factor and their expression in human gliomas in vivo. Cancer Res 1991; 51(8):2164-72.

21. Aldape KD, Ballman K, Furth A, Buckner JC, Giannini C, Burger PC, et al. Immunohistochemical detection of EGFRvIII in high malignancy grade astrocytomas and evaluation of prognostic significance. J Neuropathol Exp Neurol 2004;63:700-7.

22. Tohma Y, Gratas C, Biernat W, Peraud A, Fukuda M, Yonekawa Y, et al. PTEN (MMAC1) mutations are frequent in primary glioblastomas (de novo) but not in secondary glioblastomas. J Neuropathol Exp Neurol 1998;57:684-9.

23. Rasheed BK, McLendon RE, Friedman HS, Friedman AH, Fuchs HE, Bigner DD, et al. Chromosome 10 deletion mapping in human gliomas: a common deletion region in 10q25. Oncogene 1995;10(11):2243-6.

24. Hartmann C, Meyer J, Balss J, Capper D, Mueller W, Christians A, et al. Type and frequency of IDH1 and IDH2 mutations are related to astrocytic and oligodendroglial differentiation and age: a study of 1,010 diffuse gliomas. Acta Neuropathol 2009; 118:469-74.

25. Lai A, Kharbanda S, Pope WB, Tran A, Solis OE, Peale F, et al. Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin. J Clin Oncol 2011;29:4482-90.

26. Koivunen P, Lee S, Duncan CG, Lopez G, Lu G, Ramkissoon S, et al. Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. Nature 2012; 483:484-8.

27. Li S, Chou AP, Chen W, Chen R, Deng Y, Phillips HS, et al. Overexpression of isocitrate dehydrogenase mutant proteins renders glioma cells more sensitive to radiation. Neuro Oncol 2013;15:57-68.

28. Mohrenz IV, Antonietti P, Pusch S, Capper D, Balss J, Voigt S, et al. Isocitrate dehydrogenase 1 mutant R132H sensitizes glioma cells to BCNU-induced oxidative stress and cell death. Apoptosis 2013;18:1416-25.

29. Wang JB, Dong DF, Wang MD, Gao K. IDH1 overexpression induced chemotherapy resistance and IDH1 mutation enhanced chemotherapy sensitivity in glioma cells in vitro and in vivo. Asian Pac J Cancer Prev 2014;15:427-32.

30. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, et al. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 2005; 352:997-1003.

31. Berger MS, Rostomily RC. Low grade gliomas: functional mapping resection strategies, extent of resection and outcome. J Neuro-Oncol 1997;34:85-101.

32. Gasser T, Szelenyi A, Senft C, Muragaki Y, Sandalcioglu IE, Sure U, et al. Intraoperative MRI and functional mapping. Acta Neurochir Suppl 2011;109:61-65.

33. Sanai N, Mirzadeh Z, Berger MS. Functional outcome after language mapping for glioma resection. N Eng J Med 2008;358(1):18-27.

34. Nabavi A, Thurm H, Zountsas B, Pietsch T, Lanfermann H, Pichlmeier U, et al. Five-aminolevulinic acid for fluorescence-guided resection of recurrent malignant gliomas: a phase ii study. Neurosurgery 2009;65(6):1070-6.

35. Walker MD, Green SB, Byar DP, Alexander E, Batzdorf U, Brooks WH, et al. Randomized comparisons of radiotherapy and nitrosoureas for the treatment of malignant glioma after surgery. N Eng J Med 1980;303(23):1323-29.

36. Jeremic B, Shibamoto Y, Grujicic D, Milicic B, Stojanovic M, Nikolic N, et al. Hyperfractionated radiation therapy for incompletely resected supratentorial low-grade glioma. A phase II study. Radiother Oncol 1998;49:49-54.

37. Mason WP, Krol GS, DeAngelis LM. Low grade oligodendroglioma responds to chemotherapy. Neurology 1996;46:203-7.

38. Paleologos NA, Macdonald DR, Vick NA, Cairncross JG. Neoadjuvant procarbazine, CCNU, and vincristine for anaplastic and aggressive oligodendroglioma. Neurology 1999;53:1141-3.

39. Stupp R, Reni M, Gatta G, Mazza E, Vecht C. Anaplastic astrocytoma in adults. Crit Rev Oncol Hematol 2007;63(1):72-80.

40. Friedman HS, Prados MD, Wen PY, Mikkelsen T, Schiff D, Abrey LE, et al. Bevacizumab alone and in combination with irinotecan in recurrent glioblastoma. J Clin Oncol 2009;27(28):4733-40.

41. Pavlidis ET, Pavlidis TE. Role of bevacizumab in colorectal cancer growth and its adverse effects: A review. World J Gasttroenterol 2013;19(31):5051-60.

42. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. Science 1987;235(4785):177-82.

43. Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, et al. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 1989;244(4905):707-12.

44. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. N Engl J Med 2001;344(11):783-92.

45. Junttila TT, Parsons K, Olsson C, Lu Y, Xin Y, Theriault J, et al. Superior in vivo efficacy of afucosylated trastuzumab in the treatment of HER2-amplified breast cancer. Cancer Res 2010;70(11):4481-9.

46. Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, et al. Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. N Eng J Med 2005;353(16):1673-84.

47. Gianni L, Eiermann W, Semiglazov V, Manikhas A, Lluch A, Tjulandin S. Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH trial): a randomised controlled superiority trila with a parallel HER2-negative cohort. Lancet 2010;375(9712):377-84.

48. Buzdar AU, Ibrahim NK, Francis D, Booser DJ, Thomas ES, Theriault RL, et al. Significantly higher pathologic complete remission rate after neoadjuvant therapy with trastuzumab, paclitaxel, and epirubicin chemotherapy: results of a randomised trial in human epidermal growth factor receptor 2-positive operable breast cancer. J Clin Oncol 2005;23(16):3676-85.

49. Inamura K, Nimomiya H, Ishikawa Y, Matsubara O. Is the epidermal growth factor receptor status in lung cancers reflected in clinicopathologic features? Arch Pathol Lab Med 2010;134:66-72.

50. Gupta R, Dastane AM, Forozan F, Riley-Portuguez A, Chung F, Lopategui J, et al. Evaluation of EGFR abnormalities in patients with pulmonary adenocarcinoma: the need to test neoplasms with more than one method. Mod Pathol 2009;22:128-33.

51. Ladanyi M, Pao W. Lung adenocarcinoma: guiding EGFR-targeted therapy and beyond. Mod Pathol 2008;21:S16-22.

52. Goldman JM, Melo JV. Chronic myeloid leukemia---advances in biology and new approaches to treatment. N Engl J Med 2003;349:1451-64.

53. Steelman LS, Pohnert SC, Shelton JG, Franklin RA, Bertrand FE, McCubrey JA. JAK/STAT, Raf/MEK/ERK, PI3K/Akt and BCR-ABL in cell cycle progression and leukemogenesis. Leukemia 2004;18:189-218.

54. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med 1996;2:561-6.

55. Kantarjian H, Sawyers C, Hochhaus A, Guilhot F, Schiffer C, Gambacorti-Passerni C, et al. Hematologic and cytogenetic responses to imatinib mesylate in chronic myelogenous leukemia. N Engl J Med 2002;346:645-52.

56. Druker BJ, Guilhot F, O'Brien SG, Gathmann I, Kantarjian H, Gattermann N, et al. Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Eng J Med 2006;355:2408-17.

57. DeMatteo RP, Ballman KV, Antonescu CR, Maki RG, Pisters PWT, Demetri GD, et al. Placebo-controlled randomized trial of adjuvant imatinib mesylate following the resection of localized, primary gastrointestinal stromal tumour (GIST). Lancet 2009;373(9669): 1097-104.

58. Heinrich MC, Corless CL, Demetri GD, Blanke CD, von Mehren M, Joensuu H, et al. Kinase mutations and imatinib response in patients with metastatic gastrointestinal stromal tumour. J Clin Oncol 2003;23:4342-9.

59. Heinrich MC, Corless CL, Duensing A, McGreevey L, Chen CJ, Joseph N, et al. PDGFRA activating mutations in gastrointestinal stromal tumours. Science 2003;299(5607):708-10.

60. Verweij J, Casali PG, Zalcberg J, LeCesne A, Reichardt P, Blay JY, et al. Progression-free survival in gastrointestinal stromal tumours with high-dose imatinib: randomised trial. Lancet 2004;364(9440):1127-34.

61. Stupp R, Mason WP, van den Bent MJ, Weller M, Fischer B, Taphoorn MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 2005;352(10):987-96.

62. Zarychanski R, Chen Y, Bernstein CN, Hebert PC. Frequency of colorectal screening and the impact of family physicians on screening behaviour. CMAJ 2007;177:593-97.

63. Levin TR, Zhao W, Conell C, Seeff LC, Manninen DL, Shapiro JA, Schulman J. Complications of colonoscopy in an integrated health care delivery system. Ann Intern Med 2006;145:880-6.

64. Sawin PD, Hitchon PW, Follett KA, Torner JC. Computed imaging-assisted stereotactic brain biopsy: a risk analysis of 225 consecutive cases. Surg Neurol 1998;49(6):640-9.

65. Samadani U, Stein S, Moonis G, Sonnad SS, Bonura P, Judy KD. Stereotactic biopsy of brain stem masses: decision analysis and literature review. Surg Neurol 2006;66(5):484-90.

66. Chen CC, Hsu PW, Wu TW, Lee ST, Chang CN, Wei KC, et al. Stereotactic brain biopsy: single centre retrospective analysis of complications. Clin Neurol Neurosur 2009;111(10):835-39.

67. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. J Lab Clin Med 2006;147:126-32.

68. Liew CC, Mohr S. The peripheral-blood transcriptome: new insights into disease and risk assessment. Trends Mol Med 2007;13(10):422-32.

69. Galdkevich A, Nelemans SA, Kauffman HF, Kork J. Microarray profiling of lymphocytes in internal diseases with an altered immune response: potential and methodology. Mediators Inflamm 2005;6:317-30.

70. Hansson GK. Inflammation, atherosclerosis, and coronary artery disease. N Engl J Med 2005;352:1685-95.

71. Hotamisligil GS. Inflammation and metabolic disorders. Nature 2006;444(7121):860-7.

72. Coussens LM, Werb Z. Inflammation and cancer. Nature 2002;420(6917):860-7.

73. Liew CC. Methods for the detection of gene transcripts in blood and uses thereof. United States patent US 20040014059. 2004;Jan 22.

74. Marshall KW, Mohr S, Khettabi FE, Nossova N, Chao S, Bao W, et al. A blood-based biomarker panel for stratifying current risk for colorectal cancer. Int J Cancer 2010;126:1177-86.

75. Han M, Liew CT, Zhang HW, Chao S, Zheng R, Yip KT, et al. Novel blood-based, five-gene biomarker set for the detection of colorectal cancer. Clin Cancer Res 2008;14(2):455-60.

76. Yip KT, Das P, Suria D, Lim CR, Ng GH, Liew CC. A case-controlled validation study of a blood-based seven-gene biomarker panel for colorectal cancer in Malaysia. J Exp Clin Canc Res 2010;29:128-34.

77. Tsuang MT, Nossova N, Yager T, Tsuang MM, Guo SC, Shyu KG, et al. Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. Am J Med Genet B Neuropsychiatr Genet 2005;133B(1):1-5.

78. Glatt SJ, Everall IP, Kremen WS, Corbell J, Sasik R, Khanlou N, et al. Comparative gene expression analysis of blood and brain provides concurrent validation of SELENBP1 up-regulation in schizophrenia. Proc Natl Acad Sci USA 2005;102(43): 15533-8.

79. Kaushik N, Fear D, Richards SC, McDermott CR, Nuwaysir EF, Kellam P, et al. Gene expression in peripheral blood mononuclear cells from patients with chronic fatigue syndrome. J Clin Pathol 2005;58(8):826-32.

80. Tang Y, Schapiro MB, Franz DN, Patterson BJ, Hickey FJ, Schorry EK, et al. Blood expression profiles for tuberous sclerosis complex 2, neurofibramatosis type 1, and Down's syndrome. Ann Neurol 2004;56:808-14.

81. Tang Y, Gilbert DL, Glauser TA, Hershey AD, Sharp FR. Blood gene expression profiling of neurologic diseases: A pilot microarray study. Arch Neurol 2005;62:210-15.

82. Du X, Tang Y, Xu H, Lit L, Walker W, Ashwood P, et al. Genomic profiles for human peripheral blood T cells, B cells, natural killer cells, monocytes, and polymorphonuclear cells: Comparisons to ischemic stroke, migraine, and Tourette syndrome. Genomics 2006;87:693-703.

83. Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, Rosas HD, et al. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. Proc Natl Acad Sci USA 2005;102(31):11023-8.

84. Maes OC, Xu S, Yu B, Chertkow HM, Wang E, Schipper HM. Transcriptional profiling of Alzheimer blood mononuclear cells by microarray. Neurobiol Aging 2007;28(12);1795-809.

85. Tang Y, Xu H, Du X, Lit L, Walker W, Lu A, et al. Gene expression in blood changes rapidly in neutrophils and monocytes after ischaemic stroke in humans: a microarray study. J Cereb Blood Flow Metab 2006;26:1089-102.

86. Opdenakkar G, Van den Steen PE, Dubois B, Nelissen I, Van Coillie E, Masure S, et al. Gelatinase B functions as regulator and effector in leukocyte biology. J Leukoc Biol 2001;69:851-9.

87. Montaner J, Rovira A, Molina CA, Arenillas JF, Ribo M, Chacon P, et al. Plasmatic level of neuroinflammatory markers predict the extent of diffusion-weighted image lesions in hyperacute stroke. J Cereb Blood Flow Metab 2003;23:1403-7.

88. Castellanos M, Leira R, Serena J, Pumar JM, Lizasoain I, Castillo J, et al. Plasma metalloproteinase-9 concentration predicts haemorrhagic transformation in acute ischemic stroke. Stroke 2003;34:40-6.

89. Montaner J, Molina CA, Monasterio J, Abilleira S, Arenillas JF, Ribo M, et al. Matrix metalloproteinase-9 pretreatment level predicts intracranial haemorrhagic complications after thrombolysis in human stroke. Circulation 2003;107:598-603.

90. Roth J, Vogl T, Sorg C, Sunderkotter C. Phagocyte-specific S100 proteins: a novel group of proinflammatory molecules. Trends Immunol 2003;24:155-8.

91. Viemann D, Strey A, Janning A, Jurk K, Klimmek K, Vogl T, et al. Myeloid-related protein 8 and 14 induce a specific inflammatory response in human microvascular endothelial cells. Blood 2004;105:2955-62.

92. Hershey AD, Tang Y, Powers SW, Kabbouche MA, Gilbert DL, Glauser TA, et al. Genomic abnormalities in patients with migraine and chronic migraine: preliminary blood gene expression suggests platelet abnormalities. Headache 2004;44(10):994-1004.

93. Liebner S, Fischmann A, Rascher G, Duffner F. Grote EH, Kalbacher H, et al. Claudin-1 and claudin-5 expression and tight junction morphology are altered in blood vessels of human glioblastoma multiforme. Acta Neuropathol 2000;1009(3):323-31.

94. Wolburg H, Wolburg-Buchholz K, Kraus J, Rascher-Eggstein G, Liebner S, Hamm S, et al. Localization of claudin-3 in tight junctions of the blood-brain barrier is selectively lost during experimental autoimmune encephalomyelitis and human glioblastoma multiforme. Acta Neuropathol 2003;105(6):586-92.

95. Noell S, Fallier-Becker P, Beyer C, Kroger S, Mack AF, Wolburg H. Effects of agrin on the expression and distributon of the water channel protein aquaporin-4 and volume regulation in cultured astrocytes. Eur J Neurosci 2007;26(8):2109-18.

96. Wolburg H, Noell S, Wolburg-Buchholz K, Mack A, Fallier-Becker P. Agrin, aquaporin-4, and astrocyte polarity as an important feature of the blood-brain barrier. Neuroscientist 2009;15(2):180-93.

97. Noell S, Fallier-Becker P, Deutsch U, Mack AF, Wolburg H. Agrin defines polarized distribution of orthogonal arrays of particles in astrocytes. Cell Tissue Res 2009;337(2):185-95.

98. Saadoun S, Papadopoulos MC, Davies DC, Krishna S, Bell BA. Aquaporin-4 expression is increased in oedenatous human brain tumours. J Neurol Neurosur PS 2002;72(2):262-5.

99. Warth A, Kroger S, Wolburg H. Redistribution of aquaporin-4 in human glioblatoma correlates with loss of agrin immunoreactivity from brain capillary basal laminae. Acta Neuropathol 2004;107(4):311-8.

100. Larsson HB, Stubgaard M, Fredericksen JL, Jensen M, Henriksen O, Paulson OB. Quantitation of blood-brain barrier defect by magnetic resonance imaging and gadolinium-DTPA in patients with multiple sclerosis and brain tumours. Mag Reson Med 1990;16(1):117-31.

101. Wolburg H, Noell S, Fallier-Becker P, Mack AF, Wolburg-Buchholz K. The disturbed blood-brain barrier in human glioblastoma, Mol Aspects Med 2012;33(5-6):579-89.

102. Pantel K, Alix-Panabieres C. Circulating tumour cells in cancer patients: challenges and perspectives. Trends Mol Med 2010;16(9):398-406.

103. Pantel K, Alix-Panabieres C, Rietdorf S. Cancer micrometastasis. nat Rev Clin Oncol 2009;6:339-51.

104. Pantel K, Brakenhoff RH, Brandt B. Detection, clinical relevance and specific biological properties of disseminating tumour cells. nat Rev Cancer 2008;8(5):329-40.

105. Mani SA, Guo W, Liao MJ, eaton EN, Ayyanan A, Zhou AY, et al. the epithelial-mesenchymal transition generates cells with properties of stem cells. Cell 2008;133(4):704-15.

106.Smirnov DA, Zweitzig DR, Foulk BW, Miller MC, Doyle GV, Pienta KJ, et al. Global gene expression profiling of circulating tumour cells. Cancer Res 2005;65(12):4993-7.

107.MacArthur KM, kao GD, Chandrasekaran S, Alonso-Basanta M, Chapman C, Lustig RA, et al. Detection of brain tumour cells in the peripheral blood by a telomerase promoter-based assay. Cancer Res 2014;74(8):2152-9.

108.Coffey RJ, Lunsford LD, Taylor FH. Survival after stereotactic biopsy of malignant gliomas. Neurosurgery 1988;3:465-73.

109.Korshunov A, Sycheva R, Golanov A. the prognostic relevance of molecular alterations in glioblastomas for patients age <50 years. Cancer 2005;104(4):825-32.

110.Cervoni L, Celli P, Salvati M. Long term survival in a patient with supratentorial glioblastoma: clinical considerations. Ital J Neurol Sci 1998;19(4):221-4.

111.Salford LG, Brun A, Nirfalk S. Ten year survival among patients with supratentorial astrocytomas grade III and IV. J Neurosurg 1988;69(4):5069.

112.Scott JN, Rewcastle NB, Brasher PM, Fulton D, MacKinnon JA, Hamilton M, et al. Which glioblastoma multiforme patient will become a long term survivor? A population based study. Ann Neurol 1999;46(2):183-8.

113.Chandler KL, Prados MD, Malec M, Wilson CB. Long term survival in patients with glioblastoma multiforme. Neurosurgery 1993;32(5):716-20.

114.Marko NF, Toms SA, Barnett GH, Weil R. Genomic expression patterns distinguish long term from short term glioblastoma survivors: a preliminary feasibility study. Genomics 2008;91(5):395-406.

115.Lantos PL, VandenBerg SR, Kleihues P. Tumours of the nervous system. In: Graham DI, Lantos PL, editors. Greenfield's neuropathology. London:Arnold;1996.p583-879.

116.Brightman MW, Reese TS. Junctions between intimately apposed cell membranes in the vertebrate brain. J Cell Biol 1969;40(3):648-77.

117.Peters A, Palay SL,Webster H. The fine structure of the nervous system(second edition). New York: Oxford University Press; 1991.

118.Dinda AK, Sarkar C, Roy S, Kharbanda K, Mathur M, Khosla AK. A transmission and scanning electron microscopic study of tumoural and peritumoural microblood vessels in human gliomas. J Neuro-Oncol 1993;16(2):149-58.

119.Sage MR, Wilson AJ. The blood-brain barrier: an important concept in neuroimaging. Am J Neuroradiol 1994;15(4):601-22.

120.Nicholson AC, Unger ER, Mangalathu R, Ojaniemi H, Vernon SD. Exploration of neuroendocrine and immune gene expression in peripheral blood mononuclear cells. Brain Res Mol Brain Res 2004;129(1-2):193-7.

121.Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Clin Orthop Relat Res 2007;455:3-5.

122.Hulka BS. Overview of biological markers. In: Hulka BS, Griffith JD, Wilcosky TC, editors. Biological markers in epidemiology. New York: Oxford University Press;1990. p3-15.

123.Schulte PA. A conceptual and historical framework for molecular epidemiology. In: Schulte PA, Perera FP, editors. Molecular epidemiology: principles and practices. San Diego:Academic Press;1993.p3-44.

124.Mayeux R. Biomarkers: potential uses and limitations. Neurotherapeutics 2004;1:182-88.

125.Schwartzbaum J, Seweryn M, Holloman C, Harris R, Handelman SK, Rempala GA. Association between prediagnostic allergy-related serum cytokines and glioma. PLoS One 2015;10(9):e0137503.

126.Forrest MS, Lan Q, Hubbard AE, Zhang L, Vermeulen R, Zhao X. Discovery of novel biomarkers by microarray analysis of peripheral blood mononuclear cell gene expression in benzene-exposed workers. Environ Health Perspect 2005;113(6):801-7.

127.Wang Z, Neuburg D, Li C, Su L, Kim JY, Chen JC. Global gene expression profiling in whole-blood samples from individuals exposed to metal fumes. Environ Health Perspect 2005;113(2):233-41.

128.Lampe JW, Stepaniants SB, Mao M, Radich JP, Dai H, Linsley PS. Signatures of environmental exposure using peripheral leukocyte gene expression: tobacco smoke. Cancer Epidemiol Biomarkers Prev 2004;13(3):445-53.