
The identification of microRNAs to predict glioma prognosis

Josie Louise Hayes

Submitted in accordance with the requirements for the degree
of PhD. The University of Leeds, Faculty of Medicine and Health,
Leeds Institute of Cancer and Pathology May 2015

Intellectual property and publication statements

The candidate confirms that the work submitted is her own, except where work which has formed part of jointly authored publications has been included. The contribution of the candidate and the other authors to this work has been explicitly indicated below. The candidate confirms that appropriate credit has been given within the thesis where reference has been made to the work of others.

Publications: 1st authored and used in thesis

Chapter 1: **Josie Hayes**, Pier Paolo Peruzzi, Sean E. Lawler. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med* 2014, 20:460-469.

Chapter 2: **Josie Hayes**, Helene Thygesen, Charlotte Tumilson, Alastair Droop, Marjorie Boissinot, Thomas A. Hughes, David Westhead, Jane E. Alder, Lisa Shaw, Susan C. Short, Sean E. Lawler. Prediction of clinical outcome in glioblastoma using a biologically relevant nine-microRNA signature. *Molecular Oncology* 2014, 9(3):704-14.

Chapter 3: **Josie Hayes**, Helene Thygesen, Alastair Droop, Thomas A. Hughes, David Westhead, Sean E. Lawler, Heiko Wurdak, Susan Short. Prognostic microRNAs in high-grade glioma reveal a link to OP differentiation. *Oncoscience* 2014, 2(3):252-62.

Author Contributions

Josie Hayes was involved in experimental design, creation, analysis and interpretation of results and wrote the first draft of each manuscript. David Westhead, Susan Short and Sean Lawler were involved in experimental design, analysis and interpretation of results and contributed to the manuscript.

This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement. The right of Josie Hayes to be identified as Author of this work has been asserted by her in accordance with the Copyright, Designs and Patents Act 1988. © 2010 The University of Leeds and Josie Hayes.

Acknowledgements

Foremost I would like to express my gratitude to my supervisor, David Westhead for his continued guidance and support in the years leading up to, and during my PhD. I would also like to thank my co-supervisors Susan Short, Thomas Hughes and Alastair Droop for their immense knowledge and motivation throughout my PhD. My sincere thanks also go to my external supervisor and mentor Sean Lawler who has provided great inspiration, support and knowledge over the past three years.

I thank all members of the Bioinformatics and Translational Neuro-oncology groups at the University of Leeds for their help and laughter during the hard times and the celebrations. Special thanks must also go to Marjorie Boissinot for her advice and encouragement throughout this PhD. I must also mention Helene Thygesen who was an oracle of statistical knowledge and kept me on the right track.

I would like to thank the Chiocca group at Harvard Medical School who have been of great guidance and encouragement, particularly Maria Speranza and Choi-Fong Cho.

A mention must go to the Yorkshire Regional Cytogenetics department who provided the motivation for this PhD and helped me maintain direction throughout. I am sincerely grateful to have had them as friends and colleagues for the past ten years.

I would like to thank my family for their understanding, compassion and continued support throughout all of my studies. I would also like to mention my friends, particularly Debbie Green and Nuria Navarro-Coy for providing a vent for my frustration and a sense of perspective when I needed it.

Finally I would like to thank Yorkshire Cancer Research for funding this PhD and making it possible.

Abstract

Until now, personalised medicine for patients in oncology has been focused on the use of DNA-based techniques such as mutation detection and fluorescence *in situ* hybridisation, fluorescence-activated cell sorting and immuno-staining for classifying tumours. MicroRNAs are short non-coding RNAs that are involved in post-translational regulation of gene expression. Their expression levels are often altered in cancer. Due to their functional importance and stability in biological samples, they represent another tool that could be used to aid patient management.

Glioblastoma is a disease that has had little improvement in survival over the past decade in comparison to other cancers. A number of new drugs have been explored but even successful trials have shown limited success.

This thesis is focused on identification of microRNAs as signatures for prognosis prediction in glioblastoma. It is separated into four parts; the identification of a microRNA signature that can be used to predict prognosis in glioblastoma; the alignment of glioblastoma microRNA expression with the microRNA expression of oligodendrocyte precursors and its involvement in patient outcome; the use of the expression pattern of the most abundant and robust prognostic microRNA in glioma (miR-9) to delineate glioblastoma subtype and finally the identification of a microRNA signature to predict prognosis in patients treated with the anti-angiogenic drug bevacizumab. The research aims to create signatures suitable for clinical practice, with a small number of predictors, and where possible the function of the microRNAs has been predicted and reviewed to provide confirmation of their role in glioma biology.

The key findings of this research are the formation of robust signatures using microRNAs in a disease where few markers are available and proof of a technique that can be used in future drug studies to improve performance at clinical trials.

Table of contents

Intellectual property and publication statements	2
Acknowledgements	3
Abstract.....	4
Table of contents	5
List of figures	8
List of Tables	9
Abbreviations	10
1. Introduction	15
1.1. Incidence and survival of glioma.....	15
1.2. Pathological diagnosis of glioma.	19
1.2.1. Karyotyping and fluorescent <i>in situ</i> hybridisation.	19
1.2.2. IDH1 mutations.....	20
1.2.3. MGMT methylation.	22
1.2.4. 1p19q codeletion in oligodendroglioma.....	23
1.2.5. BRAF duplication/fusion in pilocytic astrocytoma.....	23
1.3 Standard treatment of glioma.	24
1.3.1. Radiotherapy and imaging.	24
1.3.2. Chemotherapy.....	25
1.4. Genetic features of grade II glioma.	27
1.5. Genetics of high grade glioma.....	28
1.5.1. Molecular subtypes of glioblastoma.	28
1.5.2. Heterogeneity and clonal evolution.	31
1.5.3. Cell of origin.	32
1.5.4. Altered pathways.....	34
1.5.5. EGFR alteration and EGFR variant III.....	36
1.6. The importance of stem cells in glioma.	37
1.7. Novel chemotherapeutics for glioblastoma.....	39
1.8. The study of glioma cell biology in the laboratory.	41
1.9 MicroRNAs.....	42
1.9.1. MicroRNA processing and mechanism of action.....	43
1.9.2. MicroRNAs as predictors of prognosis.	45
1.9.3. MicroRNAs for classification of disease.	46
1.9.4. MicroRNAs as predictors of drug efficacy.	48
1.9.5. MicroRNA-based therapeutics and companion diagnostics.....	49
1.9.6. Assessment of microRNA alterations in a clinical laboratory.	50
1.9.7. Target prediction of microRNAs.	55
1.9.8. MicroRNAs in cancer; functions, alterations and mechanisms.	57
1.9.9. MicroRNA networks in cancer.	60
1.9.10. MicroRNAs as serum biomarkers.....	62
1.9.11. MicroRNA polymorphisms predisposing cancer.....	63
1.9.12. MicroRNAs in glioblastoma.	63
1.10. Survival analysis.	66
1.11. Publically available data.....	70
1.12. The aims and objectives of this thesis.....	70
2. A 9-microRNA signature predicts prognosis in glioblastoma.....	72
2.1. Introduction.	72
2.2. Methods.....	75

2.2.1. TCGA clinical information and expression data.	75
2.2.2. Statistical analysis of microRNA expression data in glioblastoma.	77
2.2.3. Prognostic validation of the signature in an independent cohort using qRT-PCR.	79
2.2.4. Assessment of the 9-microRNA signature in lower grade glioma.	79
2.2.5. Cell culture, transfection and validation of candidate microRNA targets.	80
2.2.6. Identifying predicted microRNA targets associated with OS.	80
2.3. Results.	81
2.3.1. Identification of a 9-microRNA signature associated with prognosis in glioblastoma.	81
2.3.2. Generation of a risk score combining expression values of the 9 microRNAs to predict survival.	85
2.3.3. Assessment of the risk score in glioblastoma subtypes and in relation to other prognostic factors.	88
2.3.4. Risk score validation in an independent dataset.	93
2.3.5. Risk score assessment in lower grade glioma.	94
2.3.6. Predicted targets of these microRNAs.	95
2.4. Discussion.	98
2.4.1. The nine microRNA signature is a molecular indicator of prognosis.	98
2.4.2. Roles of the microRNAs in the signature in glioma biology.	99
2.4.3. Translational relevance of the signature.	100
3. Prognostic microRNAs in high-grade glioma reveal a link to oligodendrocyte precursor differentiation.	102
3.1. Introduction.	102
3.2. Methods.	104
3.2.1. MicroRNA and mRNA expression analysis.	104
3.2.2. Pathway prediction.	106
3.2.3. Analysis of the differentiation pathway.	106
3.2.4. Correlation of microRNA expression of the OP pathway with malignant glioma tumours.	107
3.3 Results.	107
3.3.1 Identification of a high-grade glioma microRNA signature associated with poor patient survival.	107
3.3.2. Determination of the role of OP gene expression in prognosis of glioma.	113
3.3.3. Correlation of glioma tumour microRNA expression with the OP cell stage.	117
3.4. Discussion.	119
3.4.1. Prognostic glioma microRNAs align with OP pathways.	119
3.4.2. Translational relevance of the OP1 prognostic signature.	121
4. Investigation of microRNA-9 in malignant glioma.	123
4.1. Introduction.	123
4.2. Methods.	124
4.2.1. Cell culture and transfection.	124
4.2.2. RNA extraction and quantitative real-time PCR.	126
4.2.3. RNA extraction and Illumina microRNA Sequencing (sequencing performed by Dr. Sally Harrison, Leeds Teaching Hospitals).	126
4.2.4. Cell adhesion assays.	127
4.2.5. Cell viability assays.	127
4.2.6. Transwell migration assays.	128
4.2.7. Western blotting.	128
4.2.8. Luciferase reporter assays.	129
4.2.9. TCGA data, statistical analysis and target prediction.	129
4.2.10. Sequencing bioinformatics pipeline (performed by Dr. Lucy Stead).	130

4.3. Results.	130
4.3.1 Expression of miR-9 in samples with different prognosis, and with different molecular subtypes.	130
4.3.2. Association of miR-9 precursor expression with prognosis and abundance of mature miR-9 in glioma.	134
4.3.3. Levels of expression of miR-9 in tumours with different percentages of tumour cell content.	136
4.3.4. Expression of miR-9 in mesenchymal glioma stem cells.	138
4.3.5 Association of miR-9 with necrosis.	141
4.3.6. Effect of hypoxia on miR-9 expression levels.	142
4.3.7. Pathway enrichment of predicted miR-9 targets.	143
4.3.8. Effect of miR-9 on adhesion.	144
4.3.9 A target of miR-9 is SHC1, a cytoskeletal remodelling protein.	147
4.4 Discussion.	150
5. MicroRNAs predicting response to the anti-angiogenic drug bevacizumab.	153
5.1. Introduction.	153
5.2. Methods.	155
5.2.1. TCGA clinical information and expression data.	155
5.2.2. Generation of a risk algorithm for OS in bevacizumab-treated glioblastoma patients using microRNAs.	156
5.2.3. Validation of the risk score in the test set.	157
5.2.4. Testing of the algorithm across all treatment types.	157
5.2.5. Characterisation of the two groups defined by the signature.	158
5.2.6. Testing of the signature in a group of National Cancer Institute (NCI) cell lines.	158
5.3. Results.	159
5.3.1. An 8-microRNA signature generated from the training set predicts prognosis in bevacizumab treated patients.	159
5.3.2. A risk score combining expression values of the 8 microRNAs predicts survival in the training set of bevacizumab treated patients.	160
5.3.3. Assessment of the signature in the test group of 37 patients.	164
5.3.4. Testing of the signature across all the patients in the TCGA, treated with different treatment types.	166
5.3.5. Characterisation of the two groups defined by the signature.	167
5.3.6. Ability of the signature to predict progression free survival.	170
5.3.7. Assessment of the signature in NCI cell lines.	170
5.4. Discussion.	173
6. Discussion.	176
6.1 MicroRNA expression and its association with prognosis.	176
6.2. The use of microRNAs to determine the ultimate ‘aim’ of the cell/tumour.	177
6.3. MicroRNAs in the clinical setting.	178
6.4. How I have adapted my work for eventual introduction to the clinic.	179
6.5. The role of prognostic microRNAs in malignant glioma.	183
6.6. Limitations of this work.	188
6.7. Conclusion and future perspectives.	189
Appendices	191
References	219

List of figures

Figure 1.1. Histological criteria of the WHO classification of gliomas.	17
Figure 1.2. Overall survival of patients with glioblastoma from 2000 to 2006 from the SEER (surveillance, epidemiology and end results program) database.	18
Figure 1.3. Molecular and cytogenetic markers used to classify a glioma.	20
Figure 1.4. IDH1 mutations inhibit histone and DNA demethylation.	22
Figure 1.5. Anti-tumorigenic action of TMZ involves methylation of the O6 position of guanine, resulting in G>A mutations.	26
Figure 1.6. The differentiation pathway involving oligodendrocyte precursors.	34
Figure 1.7. MicroRNA biogenesis.	44
Figure 1.8. Seed sequences of microRNAs and their relevance in the determination of the mRNA fate.	45
Figure 1.9. Alterations in the microRNA processing machinery in cancer.	58
Figure 1.10. The probability distribution of a parameter, β , when penalised using ridge and LASSO.	69
Figure 2.1. MGMT promoter methylation predicts a better prognosis in glioblastoma.	73
Figure 2.2. Four prognostic microRNA signatures developed for glioblastoma.	74
Figure 2.3. Workflow for generation of the prediction algorithm	77
Figure 2.4. Cox regression on all nine microRNAs identified by LASSO regression on 475 glioblastoma patients.	84
Figure 2.5. The patient groups assigned to the high- and low-risk groups using the median as a threshold.	86
Figure 2.6. Log-rank of the low-risk and high-risk groups in subgroups of glioblastoma.	90
Figure 2.7. Assessment of risk groups in TMZ treated patients, the validation cohort and lower grade glioma.	92
Figure 2.8. Expression of the predicted targets following transfection of a miR-9 mimic into LN229 cells relative to a scrambled control.	97
Figure 3.1. The cell stages in the oligodendrocyte precursor differentiation pathway. Oligodendrocyte precursors are intermediates between neural progenitors and astrocytes and oligodendrocytes.	104
Figure 3.2. The computational analysis pipeline to identify common prognostic molecular signatures in high-grade astrocytoma.	108
Figure 3.3. Fold changes of the differentially expressed microRNA expression between the good and poor prognosis groups in GliIIA (A) and glioblastoma (B).	110
Figure 3.4. Differentially expressed microRNAs in good and poor prognosis groups of glioma point to OP-related pathways.	112
Figure 3.5. Correlation coefficients comparing the fold change of microRNA expression between each stage in the OP pathway and the GliIIA and glioblastoma good and poor prognosis groups.	115
Figure 3.6. The correlation of microRNA expression between each cell type with glioma tumours in the TCGA.	118
Figure 4.1. The pEZ reporter plasmid with the 3' UTR of SHC1.	129
Figure 4.2. In glioma miR-9 is a prognostic microRNA that decreases with poorer patient outcome.	131

Figure 4.3. miR-9 expression is lower in mesenchymal glioblastoma.	133
Figure 4.4. All precursors of miR-9 are associated with survival but levels of the precursors differ across samples.	135
Figure 4.5. miR-9 sequences are highly represented in the total population of microRNA sequences with brain and glioblastoma samples.	136
Figure 4.6. The patient group with below median expression of miR-9 have fewer tumour cells in the sample.	137
Figure 4.7. Levels of miR-9 are low in mesenchymal glioma stem cell lines; other cell lines have varying expression levels.	139
Figure 4.8. Glioblastoma tumours with lower miR-9 expression have more extensive necrosis.	141
Figure 4.9. When cells are exposed to hypoxia, levels of miR-9 are lower. ...	142
Figure 4.10. The top ten enriched pathways of the predicted targets for miR-9, with representation of proportion of genes in the pathway.	143
Figure 4.11. Cell adhesion assays show that miR-9 increased the ability of cells to adhere to each other.	145
Figure 4.12. Cell adhesion assays show that miR-9 increased the ability of cells to adhere to a substrate.	146
Figure 4.13. miR-9 overexpression does not affect cell viability.	147
Figure 4.14. The predicted target sites in SHC1 3' UTR for miR-9.	148
Figure 4.15. Overexpression of miR-9 decreases SHC1 protein levels.	149
Figure 5.1. Density plot of the risk score in the 50 training set patients.	161
Figure 5.2. The two groups of the training set (n=50) defined by using a cut-off value risk score of 0.	163
Figure 5.3. The two groups of the test set (n=37) defined by using a cut-off value risk score of 0.	165
Figure 5.4. The responder and non-responder groups of the whole of the glioblastoma patients in the TCGA as defined by the signature, regardless of treatment.	167
Figure 5.5. Density plot of the risk score in eight NCI cell lines (miR-124a was not included in this risk score).	171
Figure 5.6. The microRNA expression of the 7 microRNAs (original signature minus miR-124a) in eight NCI cell lines.	172
Appendix 2.1. Patient sample collection at UCLan.	191
Appendix 4.1. Consent and transfer for the samples.	214
Appendix 4.3. Characterisation of the GSC cell lines (performed by Dr. Marco Mineo, Harvard Medical School).	216
Appendix 4.4. Glut1 (target of HIF1A) mRNA levels to show cellular response to hypoxia.	218

List of Tables

Table 2.1. Characteristics of patients used in the generation of the signature.	76
Table 2.2. MicroRNAs associated with survival using the LASSO regression test.	82
Table 2.3. The genomic location of the microRNAs identified by LASSO regression.	85

Table 2.4. Median quantile normalised expression of microarray probe for each microRNA across 475 glioblastoma tumours in the TCGA.	87
Table 2.5. Results of multivariable Cox regression incorporating prognostic factors in glioblastoma.	89
Table 2.6. Patient characteristics in the independent dataset used for validation.	94
Table 2.7. Predicted target interactions of the signature microRNAs with significant correlation in expression.	96
Table 3.1. Characteristics of the grade III astrocytoma and glioblastoma TCGA tumours in poor and good prognosis groups.	105
Table 5.1. Characteristics of the patients included in the study.	156
Table 5.2. Patient information for the whole TCGA dataset of glioblastoma patients.	158
Table 5.3. The eight microRNAs with non-zero coefficients from LASSO regression.	160
Table 5.4. Results of log-rank test on each of the risk score minimum densities.	162
Table 5.5. Assessment of the characteristics in the responder and non-responder groups in the 87 bevacizumab treated patients in the TCGA.	169
Appendix 2.2. The top 100 pathways for predicted targets of the microRNAs from Metacore.	200
Appendix 3.1. 150 microRNAs used in the OP signature.	212
Appendix 4.2. Primer Sequences.	215

Abbreviations

2HG: 2-hydroxyglutarate
ABC transporter: ATP-binding cassette transporter
ADAR: Adenosine deaminase acting on RNA
AGO: Argonaute
AKT: V-Akt murine thymoma viral oncogene
ALDH: Aldehyde dehydrogenase
ALT: Alternative lengthening of telomeres
AML: Acute myeloid leukaemia
ARFGEF1: ADP-Ribosylation Factor Guanine Nucleotide-Exchange Factor 1
ASCO: American Society of Clinical Oncology
ATCC: American Type Culture Collection
ATP: Adenosine triphosphate
ATRX: Alpha thalassemia/mental retardation syndrome X-linked
ATXN3: Ataxin 3
BAK1: BCL2-antagonist/killer 1
BBB: Blood brain barrier
BCAN: Brevican
BCL2A1: BCL2-related protein A1
BCNU: Bis-chloroethylnitrosourea

BCR-ABL: Breakpoint cluster region- C-abl oncogene 1, non-receptor tyrosine Kinase fusion
BIM: BCL2-Like 11 (apoptosis facilitator)
BRAF: V-Raf murine sarcoma viral oncogene homolog B
CA: California
CAB39: Calcium binding protein 39
CD133: PROM1, prominin 1
CD44: CD44 molecule (indian blood group)
CDKN2A: Cyclin-dependent kinase inhibitor 2A
CHI3L1: Chitinase 3-like 1 (cartilage glycoprotein-39)
CI: Confidence interval
CIC: Capicua transcriptional repressor
CLDN2: Claudin 2
CLL: Chronic lymphocytic leukaemia
CML: Chronic myeloid leukaemia
CNS: Central nervous system
CO₂: Carbon dioxide
CpG: —C—phosphate—G—
CRC: Colorectal cancer
CREB: CAMP responsive element binding protein
Ct: Threshold cycle value
CTLA-4: Cytotoxic T-lymphocyte-associated protein 4
DGCR8: DiGeorge syndrome critical region gene 8
DIANA: DNA Intelligent Analysis
DIAPH2: Diaphanous-related formin 2
DLL4: Delta-like 4
DNA: Deoxyribonucleic acid
DSG2: Desmoglein 2
EB: Embryoid body
ECM: Extracellular matrix
EGFR: Epidermal growth factor receptor
EGFRvIII: EGFR variant 3
EMT: Epithelial mesenchymal transition
ESC: Embryonic stem cells
EZH2: Enhancer of zeste homolog 2
F13A1: Coagulation factor XIII, A1 polypeptide
FASLG: Fas ligand (TNF Superfamily, member 6)
FDA: US Food and Drug Administration
FDR: False discovery rate
FFPE: formalin-fixed, paraffin-embedded
FGF: Fibroblast growth factor
FISH: Fluorescence *in situ* hybridisation
FNDC3B: Fibronectin type III domain containing 3B
FUBP1: Far upstream element (FUSE) binding protein 1
G-CIMP: Glioma CpG island methylator phenotype
GABA: Gamma-aminobutyric acid

GABRA: Gamma-aminobutyric acid A receptor
GAPDH: Glyceraldehyde-3-phosphate dehydrogenase
GBM: Glioblastoma
GIIIA: Grade III astrocytoma
GLuc: Gaussia luciferase
GP: Glial progenitor
GSC: Glioma stem cell
Gy: Gray (absorbed radiation unit)
HER2: Human epidermal growth factor receptor 2
HGG: High grade glioma
HIF1A: Hypoxia inducible factor 1, alpha subunit
HITS-CLIP: High-throughput sequencing of RNA isolated by crosslinking immunoprecipitation
HMM: Hidden markov model
HP: Haematological precursor
HR: Hazard ratio
IDH1: Isocitrate Dehydrogenase 1
IGFBP2: Insulin-like growth factor binding protein 2
JAK: Janus kinase 2
KDM: Histone lysine demethylase
KHDRBS2: KH domain containing, RNA binding, signal transduction associated 2
KPS: Karnofsky performance status
KRAS: Kirsten rat sarcoma viral oncogene homolog
LASSO: Least Absolute Shrinkage and Selection Operator
LIMMA: Linear Models for Microarray Data
LMNA: Lamin A
LNA: Locked nucleic acid
LOH: Loss of heterozygosity
LYN: V-src-1 yamaguchi sarcoma viral related oncogene homolog
mAB: Monoclonal antibody
MAPK: Mitogen-activated protein kinase
MDM2: Mouse double minute 2, human homolog of
MDM4: Mouse double minute 4, human homolog of
MGMT: O-6-methylguanine-DNA methyltransferase
MRI: Magnetic resonance imaging
mRNA: Messenger RNA
MRS: Magnetic resonance spectroscopy
MYPN: Myopalladin
NCI: National Cancer Institute
NF1: Neurofibromin 1
NGS: Next generation sequencing
NKIRAS2: NF- κ B inhibitor interacting ras-like 2
NKX2-2: NK2 homeobox 2
NP: Neural progenitor
NSC: Neural stem cell

NSCLC: Non-small cell lung cancer
OCT4: Octamer-binding protein 4
OLIG2: Oligodendrocyte lineage transcription factor 2
OP: Oligodendrocyte precursor
OS: Overall survival
P4HA2: Prolyl 4-hydroxylase, alpha polypeptide II
PAR-CLIP: Photoactivatable-ribonucleoside-enhanced crosslinking and immunoprecipitation
PARP: Poly (ADP-ribose) polymerase
PCNA: Proliferating cell nuclear antigen
PCR: Polymerase chain reaction
PD-1: Programmed cell death 1
PDGFRA: Platelet-derived growth factor receptor, alpha polypeptide
PDL-1: Programmed cell death ligand 1
PFS: Progression-free survival
PI3K: Phosphatidylinositol-4,5-bisphosphate 3-kinase
PIAS3: Protein inhibitor of activated STAT, 3
PKA: Protein kinase A
PTEN: Phosphatase and tensin homolog
qRT-PCR: quantitative real-time PCR
RISC: RNA-induced silencing complex
RNA: Ribonucleic acid
rRNA: Ribosomal RNA
RTK: Receptor tyrosine kinase
SDS: sodium dodecyl sulfate
SE: Standard error
SEER: Surveillance, Epidemiology, and End Results
SHC1: Src homology 2 domain containing transforming protein 1
SI: International system of units
SLC25A24: Solute carrier family 25 (mitochondrial carrier; phosphate carrier), member 24
SLC31A2: Solute carrier family 31 (copper transporter), member 2
SNP: Single nucleotide polymorphism
SOX2: SRY (sex determining region Y)-box 2
STAT3: Signal transducer and activator of transcription 3
TBST: Tris-buffered saline and Tween 20
TCF: Transcription factor
TCGA: The Cancer Genome Atlas
TESK2: Testis-specific kinase 2
TET2: Ten-eleven translocation 2
TGFB: Transforming growth factor, Beta
TMZ: Temozolomide
TNFAIP3: Tumour necrosis factor, alpha-induced protein 3
TOP2A: Topoisomerase (DNA) II alpha 170kDa
TRBP: TAR (HIV-1) RNA binding protein
U6: non-coding small nuclear RNA component of U6 small nuclear ribonucleoprotein

US/ USA: United States of America

UTR: Untranslated region

VEGF: Vascular endothelial growth factor

WHO: World Health Organisation

WNT: Wingless-type MMTV integration site family

WNT4: Wingless-type MMTV integration site family, member 4

WT1: Wilm's tumour 1

XPO5: Exportin-5

YY1: Yin and yang 1

ZEB: Zinc finger E-box binding homeobox

1. Introduction

'The scientist is not a person who gives the right answers, he's one who asks the right questions.' **Claude Lévi-Strauss, Le Cru et le Cuit, 1964.**

Gliomas are central nervous system tumours with particularly poor outcomes. These are the most common tumour originating in the brain, and they represent a very broad diagnostic category that encompasses the most malignant and common form, glioblastoma (Louis et al., 2007a). Gliomas are named due to their morphological similarity to glial cells, which function as a scaffold and insulation for neurons allowing the conduction of electrical impulses important for neurological function. The different types of glioma can be characterised by the glial cell type they most resemble. For example, astrocytomas resemble astrocytes and oligodendrogliomas resemble oligodendrocytes (Louis, 2007a). The outcomes of the different types of glioma are hugely variable and correct classification is paramount to effective patient management (Riemenschneider et al., 2010). Despite efforts to improve the outcome of patients suffering with this devastating disease, the incidence and survival of patients with glioma has changed little over time (Inskip et al., 2010).

Gliomas are graded from I to IV according to the World Health Organisation (WHO) classification system, assessing their increasingly aggressive pathological features (Louis et al., 2007a). Low-grade gliomas (I-II) are benign and well-differentiated and high-grade gliomas (III-IV) are malignant and anaplastic. Stratification based on grade depends more on pathologic features than on the cell type of origin. The assessment procedure includes features such as nuclear pleomorphism, high cellularity and mitotic index, endothelial cell proliferation and necrosis (Louis et al., 2007b; Goodenberger & Jenkins, 2012) (Fig. 1.1).

1.1. Incidence and survival of glioma.

The incidence of primary brain tumours has been increasing since the 1970s and has an incidence rate of 7.1/100,000 in the UK, as reported in 2014 (Sehmer et al., 2014). Higher grades predominate in men and low grade brain tumours are more frequent in women (Goodenberger & Jenkins, 2012). The most common adult brain tumours are gliomas and 86% of these are

glioblastoma (Sehmer et al., 2014; McKinney, 2004).

Age-related incidence differs by subtype of glioma. Astrocytic tumours show a small peak in patients under 10 years of age, which decreases to its lowest incidence between 10 and 20 years, and then steadily increases over the next 50-60 years of life (Inskip et al., 1995; McKinney, 2004). The majority of grade IV glioblastomas occur in adults between 45 and 75 years of age, with a progressive increase starting from age 30 (Louis et al., 2007b).

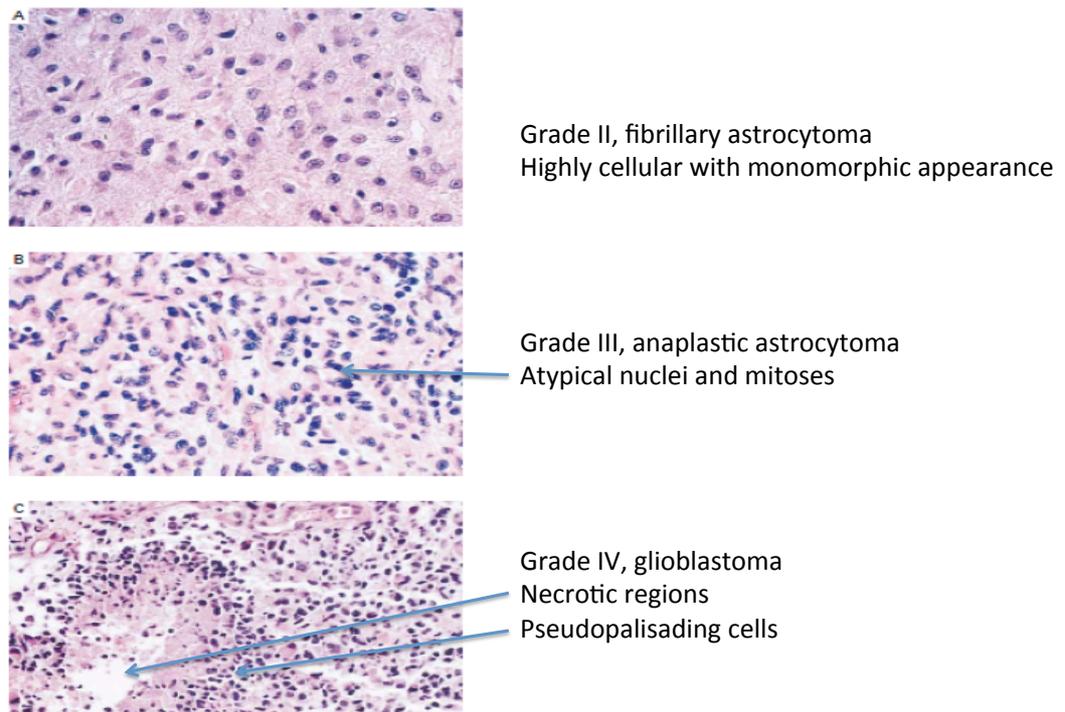


Figure 1.1. Histological criteria of the WHO classification of gliomas.

Adapted from DeAngelis, 2001 (DeAngelis, 2001). The WHO classification grades gliomas from I to IV. Grade I gliomas are non-invasive, and pilocytic astrocytoma is an example of this (not shown). Grade II tumours are infiltrative and have low proliferative capacity, however they frequently transform to a higher grade. Grade III tumours show evidence of histological malignancy with nuclear atypia and a high mitotic index. Grade IV tumours show the features of grade III but with necrotic foci, frequently surrounded by pseudopalisading cells (Louis et al., 2007a). These tumours may also exhibit vascular proliferation (DeAngelis, 2001).

Median overall survival for glioblastoma, the most aggressive and malignant glioma, is 15 months, with only 3-5% of patients surviving more than three years (Koshy et al., 2011; Krex et al., 2007)(Fig. 1.2). This highly aggressive glioma has an incidence of 3.32 per 100,000 people/year in males and 2.24 in females (Ohgaki et al., 2004). Age is a significant prognostic factor, with patients below the age of 50 having a median survival of 8.8 months compared

to a median survival of 7.3 months between the ages of 50-59 years (Ohgaki et al., 2004).

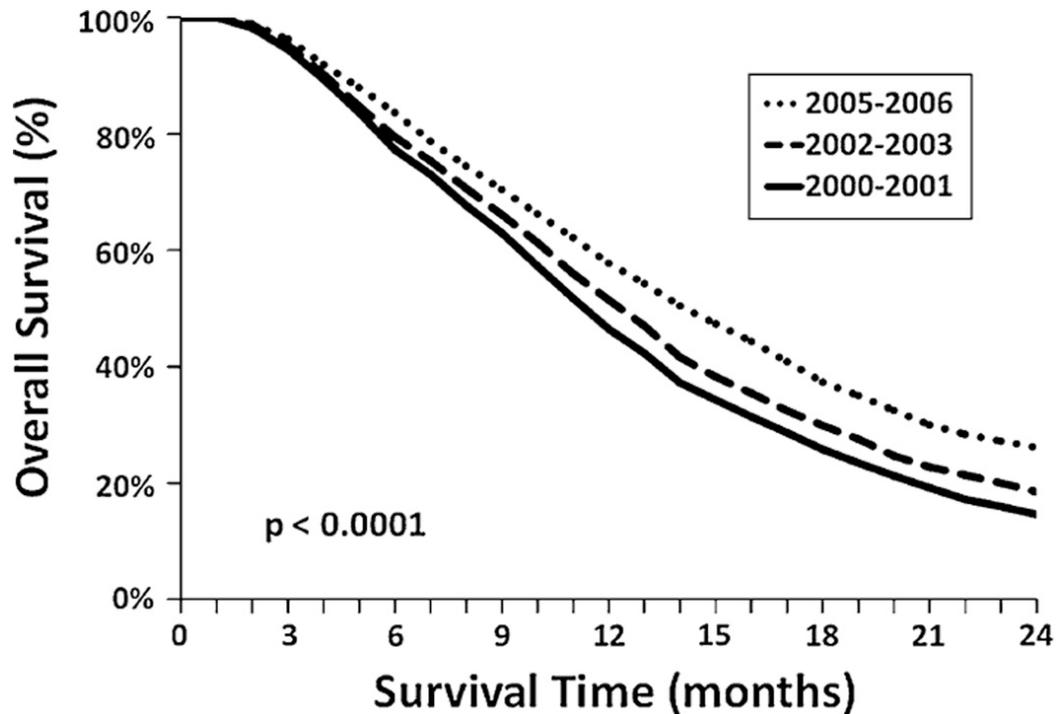


Figure 1.2. Overall survival of patients with glioblastoma from 2000 to 2006 from the SEER (surveillance, epidemiology and end results program) database.

Image from Koshy et al. 2012 (Koshy et al., 2011). Median survival of glioblastoma improved between the years 2000 and 2006 from 12 months to 15 months due to administration of the standard chemotherapeutic temozolomide with post-operative radiotherapy. This data is from the SEER database, which is a National Cancer Institute (NCI) source for cancer statistics in the US. The lines indicate the survival rates of two-year periods between 2000 and 2006.

Patients with a grade III anaplastic astrocytoma have a 2-year survival rate of 58% and, as with glioblastoma, younger patients have a better outcome (Laws et al., 2003). Other prognostic factors for high-grade glioma include extent of resection and Karnofsky performance score (KPS), which assesses patients' functional status on a scale of 1 to 100. A KPS of 80-100 is considered normal function, 50-70 suggests the patient is unable to work and 0-40 suggests the patient is unable to care for himself or herself (Laws et al., 2003).

1.2. Pathological diagnosis of glioma.

The current clinical diagnosis of glioblastoma does not solely rely on the histological features described by the WHO but also includes cytogenetic features based on molecular markers known to be associated with patient outcomes. Patients who develop a glioblastoma *ab initio*, termed a primary glioblastoma, generally have poorer outcomes than those who progress from a lower grade, termed secondary glioblastoma (Bleeker et al., 2012). Therefore, it is important to distinguish primary glioblastoma from secondary glioblastoma to ensure patients are appropriately treated and monitored.

1.2.1. Karyotyping and fluorescent *in situ* hybridisation.

Integration of histological and genetic findings allows a more precise diagnosis of the grade and subtype of the glioma (Fig. 1.3), which defines the variable patients outcomes from these tumours.

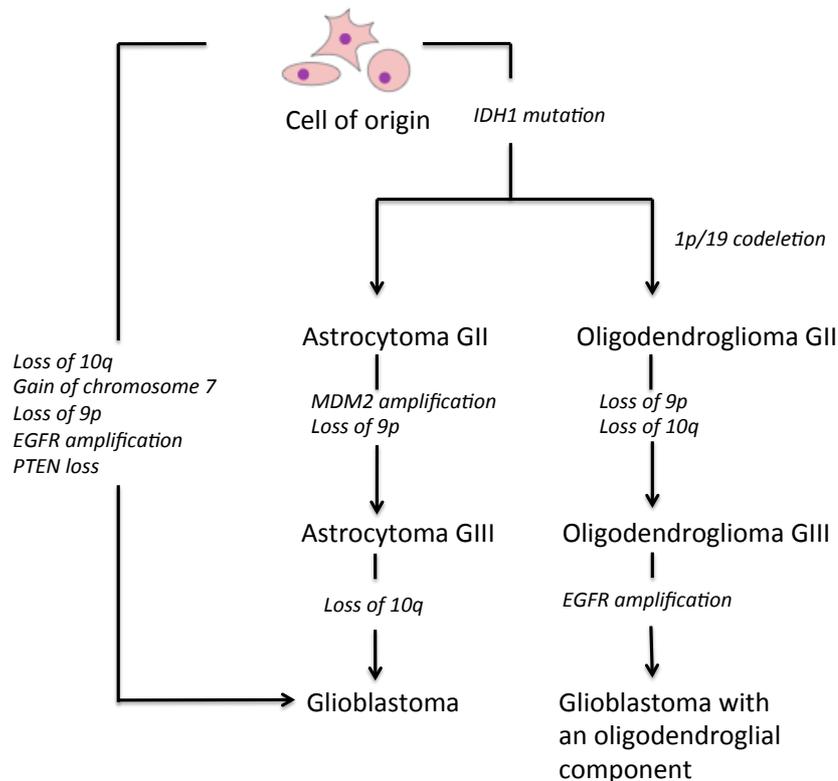


Figure 1.3. Molecular and cytogenetic markers used to classify a glioma.

Adapted from Bleeker et al. (Bleeker et al., 2012). Karyotyping glioma samples provides information on chromosome copy number. In most cases a gain in chromosome 7 will distinguish a glioblastoma from other tumour types (Wiltshire et al., 2000). Oligodendrogliomas, which are associated with a more favourable prognosis, are determined in this way by detection of 1p19q (which can also be performed using molecular testing). In addition to karyotyping, molecular tests such as fluorescent in situ hybridisation (FISH) and sequencing are performed to determine other tumour characteristics such as EGFR and MDM2 amplification.

1.2.2. IDH1 mutations.

Patients with secondary glioblastoma, which accounts for approximately 5% of all glioblastomas, often have mutations in the isocitrate dehydrogenase (IDH) gene (Parsons et al., 2008). The most common mutation is *IDH1* R132H. Parsons et al showed that only 7% of primary glioblastoma but 83% of

secondary glioblastoma had detectable IDH mutations in 2008, which fuelled further analysis in lower grade glioma the following year (Parsons et al., 2008; Yan et al., 2009). IDH mutations are frequent in glioma although extremely rare in other CNS tumours and therefore serve as a method of classification when the histological results are conflicting (Yan et al., 2009). Detection can be performed by immunohistochemistry using mutation specific antibodies or DNA sequencing (Berghoff et al., 2013). Tumours lacking a mutation in *IDH1* can harbour a mutation at amino acid 172 of the *IDH2* gene. In total, *IDH1* and *IDH2* mutations have been identified in 86% of grade II astrocytoma and oligodendroglioma and 82% of grade III disease (Yan et al., 2009). With adjustments for age, grade, MGMT status, treatment and genomic profile, *IDH* mutation is considered a favourable prognostic marker (Ducray et al., 2009). IDH is an enzyme that catalyses the oxidative carboxylation of isocitrate to α -ketoglutarate. There are three forms of IDH; IDH1 is cytosolic and IDH2 and 3 are mitochondrial (Geisbrecht & Gould, 1999). All mutations in *IDH1* are located at amino acid 132 of the protein. The most common R132H mutation is a gain of function mutation that results in an increase in 2-hydroxyglutarate (2HG) in cells (Fig. 1.4) (Zhang et al., 2013). One enzyme inhibited as a consequence of the high 2HG levels is TET2 (ten-eleven translocation 2), which catalyses the conversion of 5-methylcytosine to 5-hydroxymethylcytosine, resulting in DNA demethylation (Zhang et al., 2013). *TET2* promoter methylation has also been reported in gliomas without *IDH1* mutation suggesting that this may be an important mechanism of gliomagenesis (Kim et al., 2011c). These mutations are obvious targets for therapy in gliomas, and small molecule inhibitors targeting the IDH1 protein have already been developed (Popovici-Muller et al., 2012). These inhibitors have been shown to delay the growth of glioma cells and promote differentiation and are currently in phase I clinical trials (study number NCT02073994) (Rohle et al., 2013).

The effects of *IDH* mutation can result in widespread promoter DNA methylation and suppression of gene transcription. In 2010, Noushmehr *et al.* identified a glioma CpG island methylator phenotype (G-CIMP) which defined a subgroup of glioma (Noushmehr et al., 2010). This phenotype is tightly associated with *IDH1* mutation and introduction of an *IDH1* mutation into cells is sufficient to establish extensive DNA hypermethylation (Turcan et al., 2012).

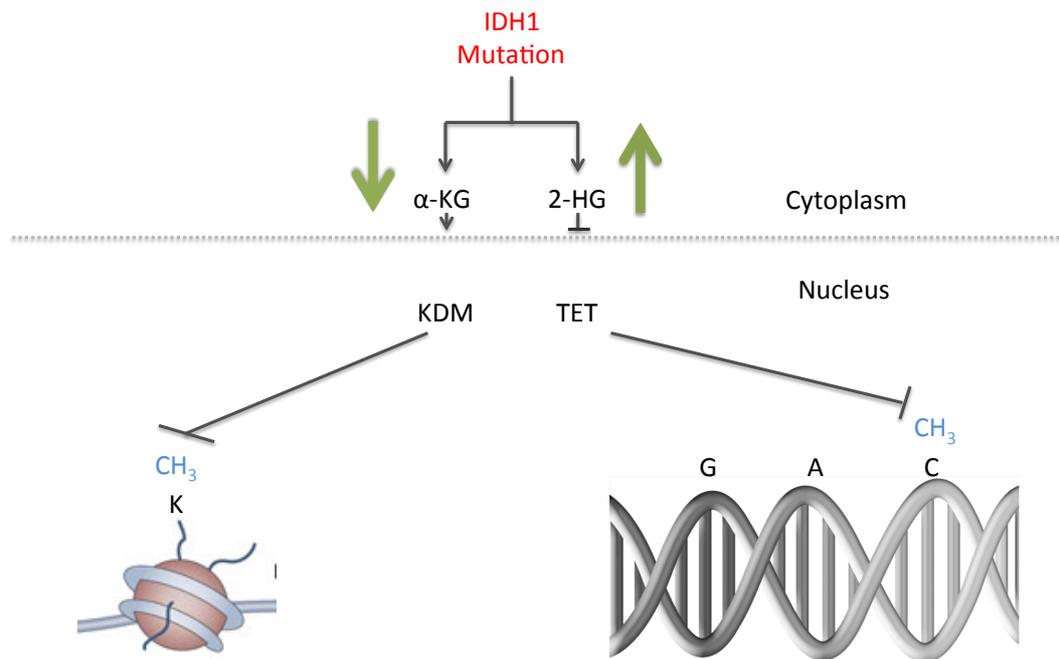


Figure 1.4. IDH1 mutations inhibit histone and DNA demethylation.

Adapted from Yang et al. (Yang et al., 2012a). IDH1 mutations are gain of function mutations that cause an increase in 2-hydroxyglutarate (2HG), which is an antagonist of α -ketoglutarate. Lower α -ketoglutarate and higher 2HG result in inhibition of both histone lysine demethylase (KDM) and the TET family of 5-methylcytosine (5mC) hydroxylases, which act to remove methyl groups (CH_3) from histone lysines (K), and cytosines (C) in the DNA, respectively. This is an epigenetic change in the cell, which causes altered differentiation processes.

1.2.3. MGMT methylation.

Promoter methylation of O⁶-methylguanine-DNA-methyltransferase (MGMT) is another predictive indicator used clinically to indicate responsiveness to alkylating agents, such as temozolomide (TMZ) (Hegi et al., 2005). MGMT is a DNA repair enzyme that can reverse the cytotoxic effect of alkylating agents by removing the O⁶ methyl group caused by TMZ treatment (discussed in detail on page 26). Methylation of the MGMT promoter most likely results in decreased expression of the MGMT gene rendering cells susceptible to these chemotherapeutic agents (Hegi et al., 2005). MGMT promoter methylation is

thought to occur as part of the G-CIMP phenotype, since it is present in almost all cases with G-CIMP and *IDH1* mutation. In non-G-CIMP phenotypes, *MGMT* methylation is still reported in 50% of cases (Bady et al., 2012; Turcan et al., 2012).

1.2.4. 1p19q codeletion in oligodendroglioma.

Oligodendroglioma is a subtype of glioma in which the cells bear some resemblance to oligodendrocytes (Louis, 2007a), which exhibit a branched-like morphology and function to produce myelin. Patients with an oligodendroglial tumour have been shown to have a more favourable prognosis than other glioma types, with current treatment regimes, therefore it is important to distinguish between oligodendroglioma and astrocytoma (Jeuken et al., 2004). Unfortunately, the histological criteria used to discriminate between these subtypes are poorly defined (Jeuken et al., 2004). There is a strong association between oligodendroglial tumours and loss of the short arm of chromosome 1 and the long arm of chromosome 19 in tumour cells. Patients with this aberration have a favourable prognosis (Ducray et al., 2009; Riemenschneider et al., 2010). This aberration is termed 1p19q codeletion; it can be characterised by an unbalanced translocation between these chromosomes or loss of heterozygosity in these chromosomal regions. In the past, detection of the 1p19q codeletion was performed using conventional karyotyping, but more recently molecular methods such as FISH, multiplex ligation dependent probe amplification (MLPA) and loss of heterozygosity (LOH) PCR have been employed (Berghoff et al., 2013).

1.2.5. BRAF duplication/fusion in pilocytic astrocytoma.

Pilocytic astrocytoma is a grade I glioma, and is a well-defined lesion that does not invade the brain (Louis et al., 2007a). Therefore, it is important to distinguish between these and other astrocytomas. Around 50-70% of pilocytic astrocytomas have fusions of the v-RAF murine sarcoma viral oncogene homolog B1 (*BRAF*) gene with the *KIAA1549* gene which results in duplication of the activation domain and a deletion of the N-terminal inhibitory domain of *BRAF*, resulting in expression of the mutant *BRAF* protein (Siegal, 2015). Detection of the *BRAF:KIAA1549* fusion gene is vital in classification of this low grade lesion, especially since glioblastoma and pilocytic astrocytoma share a

proliferative microvascular morphology and therefore can cause diagnostic uncertainty (Siegal, 2015).

1.3 Standard treatment of glioma.

Cancer treatment for solid tumours frequently involves initial resection of the tumour bulk by surgery. Surgery may not always be possible and is dependent on tumour location. When performed, brain surgery is a particularly traumatic procedure as it involves a craniotomy (an operation to open the skull). This may also be performed whilst a patient is awake to reduce the likelihood of neurological deficits. Awake craniotomy is usually performed with assessment of language and motor function at critical points in the procedure, which aids in brain mapping to define the limits of a tumour resection intra-operatively (Tate, 2015). Following maximal surgical resection, patients are treated with adjuvant therapy (therapy following the main treatment of surgery). Adjuvant therapy is designed to eliminate the remainder of the tumour and in brain tumours includes radiotherapy and chemotherapeutic agents.

1.3.1. Radiotherapy and imaging.

Conventional fractionated radiotherapy is a protocol where external beam radiation is delivered to the tumour site over a number of sessions to destroy tumour cells. The patient's head is immobilised and a CT (computed tomography) simulator creates an image of the tumour and surrounding brain using X-rays. This map of the tumour is merged with magnetic resonance imaging (MRI) and can be used to determine the treatment field, angles and energy source for most effective local control and limited toxicity. The unit of measurement of ionising radiation dose in the International System of Units (SI) is Gray (Gy), and 1Gy defines the absorption of one joule of radiation energy per one kilogram of matter (BIPM, 2006). Standard treatment for glioblastoma is 60Gy in 30 fractions of 2Gy daily 5 days a week over 6 weeks post-operatively (Barani & Larson, 2015). The dose is determined by a balance between the tumour sensitivity and tolerance of normal tissue; effective treatment versus toxicity.

Improvement in tumour imaging in recent years has benefited surgery and radiotherapy outcomes. MRI has traditionally been used in glioma imaging since the 1980s (Doyle et al., 1981). This method of imaging uses radio waves

and a magnetic field to create a detailed image of the brain. Advances on this method use diffusion to provide pathological information on a cellular level, and these include diffusion-weighted imaging and diffusion tensor imaging (Svolos et al., 2014). Magnetic resonance spectroscopy (MRS) is a more recent technique that can map the metabolic profiles of the brain using various metabolites (Chronaiou et al., 2014). This is of particular advantage in clinical practice for gliomas because they are infiltrative, and comparison of metabolic markers between tumour and normal brain tissues allows more accurate assessment of the tumour margin. Additionally, glioblastoma is extremely heterogeneous, and metabolic imaging can help to determine particularly aggressive regions by using the high Choline to N-acetyl aspartate ratio, also known as composite nutritional index. This index can be used in glioblastoma and lower grade tumours with potential sites of malignancy (Pirzkall et al., 2002). Imaging can also include other metabolites such as 2-hydroxyglutarate (2HG), which is increased in IDH mutated tumours (Chronaiou et al., 2014). However, none of these modalities are used routinely, all are the subject of research.

1.3.2. Chemotherapy.

Chemotherapy for gliomas has limited efficacy, which is in part due to the inability of many chemotherapeutic agents to cross the blood-brain-barrier (BBB), which separates the blood from the brain extracellular fluid. Tightly spaced endothelial cells, which allow the passage of only water and lipid-soluble molecules, form this barrier, and glucose and amino acids can be transported by selective transport. Lipophilic agents may be actively prevented from crossing the BBB by a membrane protein P-glycoprotein (Deeken & Loscher, 2007).

The standard chemotherapeutic for high-grade glioma, as previously mentioned is TMZ (brand names Temodar, Temodal and Temcad). This is an imidazotetrazine derivative of the alkylating agent dacarbazine, which can cross the BBB, and it alkylates/methylates DNA. Most often this occurs at the N-7 or O-6 positions of guanine residues (Fig. 1.5). Methylation at the O-6 position most likely causes the anti-tumorigenic properties of TMZ, and if unrepaired, can give rise to G>A mutations (Johnson et al., 2014). This is because O6-

methylguanine can base-pair with both cytosine and thymine, with the pairing rate dependent on the sequence context (Dosanjh et al., 1991). Also, if the O6 methylguanine remains in the template strand, DNA repair mechanisms can cause double strand breaks, which are toxic to the cell (Margison and Santibáñez-Koref, 2002).

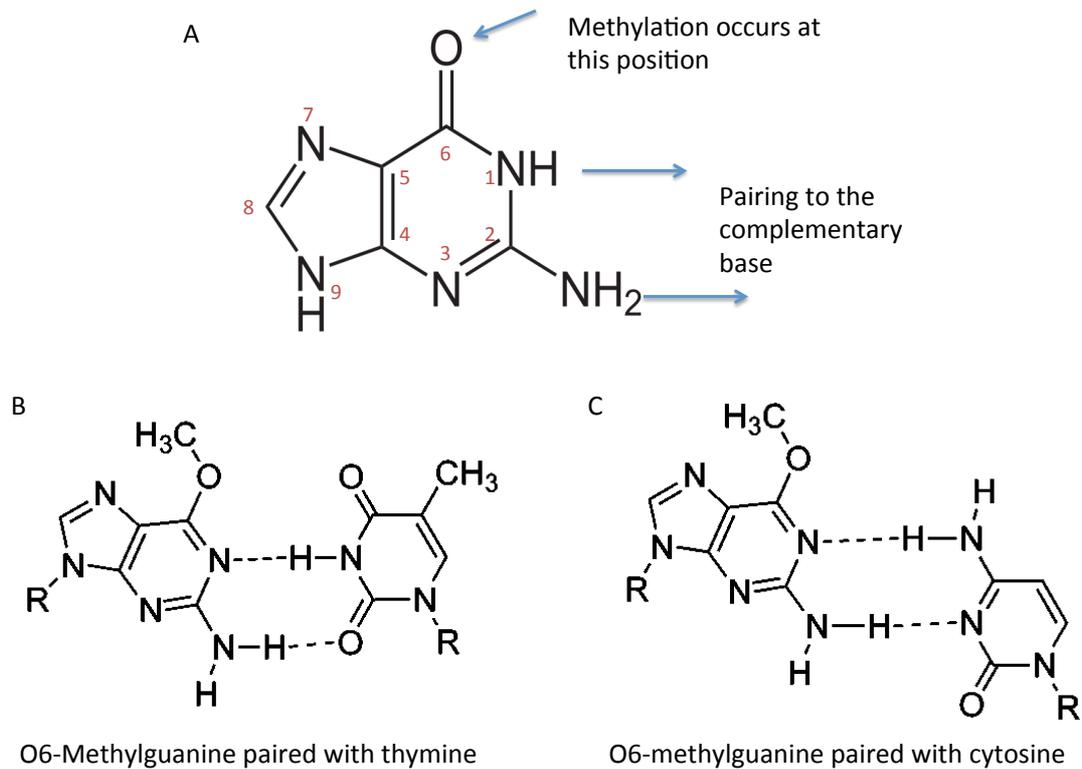


Figure 1.5. Anti-tumorigenic action of TMZ involves methylation of the O6 position of guanine, resulting in G>A mutations.

TMZ can cause methylation of the O3, O6 and N7 positions of guanine. It is the O6 adduct that gives TMZ its tumorigenic properties because it can base pair with both cytosine and thymine. This figure shows how the O6 adduct binds chemically to cytosine and thymine. A) The structure of guanine without methylation. It is at the O6 position that methylation occurs. B) O6-methyl guanine paired with thymine. C) O6-methylguanine paired with cytosine (this may also have another hydrogen bond from the O position of cytosine in solution at neutral pH) (Warren et al., 2006).

Grade II and III gliomas are heterogeneous in terms of response to therapy and have a wide range of survival times. Following maximal safe resection of the

tumour, treatment options include radiation and/or either TMZ or, an older treatment regimen of PCV (Procarbazine, CCNU, Vincristine), or regular surveillance by MRI until tumour progression (Riemenschneider et al., 2010). Unfortunately, although slow-growing in the case of grade II tumours, these usually recur two to twenty years after surgery and may have progressed to higher-grade glioma (Claus et al., 2015). The variability seen in response to current treatment regimens cannot be fully explained by known prognostic factors such as age, extent of tumour resection, molecular features, and histology. It has been shown that a proportion of patients exhibit hypermutation following treatment with TMZ, including TMZ-induced mutations that are known to drive progression to HGG (Johnson et al., 2014). All of these patients with TMZ-induced hypermutation and recurrent HGG bear IDH1 mutations, which would otherwise confer a good prognosis (Johnson et al., 2014). Treatment of glioblastoma involves maximal safe resection, adjuvant radiotherapy and concomitant (alongside the other treatments) TMZ (Stupp et al., 2009). Despite these aggressive treatments, most patients show limited response and survival is still relatively short (median survival 12-15 months (Koshy et al., 2011)). It is thought that the heterogeneity of glioblastoma may allow selection of certain subclones following treatment, which contributes to treatment resistance (Johnson et al., 2014). The presence of glioma stem-like cells; slowly dividing cells that are capable of reseeding the tumour, may also offer the tumour a mechanism to evade treatment strategies (discussed in more detail on page 37); (Ye et al., 2013).

1.4. Genetic features of grade II glioma.

As shown in Figure 1.3 on page 20, the genetic profile at different stages in the progression of glioma reflects the grade and subtype of glioma and therefore can be used in clinical diagnosis. Advances in the study of glioma have shown that there are more specific genetic changes that are not yet used in clinical practice, and these are likely to be included in the clinical guidelines in the fourth edition of the classification of CNS tumours (Louis et al., 2014).

Mutations of the genes ATRX (alpha thalassaemia/mental retardation syndrome X-linked), CIC (homolog of the *Drosophila* gene *capicua*) and FUBP1 (encoding far-upstream element (FUSE)) have been shown to delineate the glioma

subtypes. CIC and FUBP1 are encoded at chromosomes 19q and 1p respectively, regions of the 1p19q codeletion. ATRX is encoded on the long arm of chromosome X. It is a chromatin modifier which, in glioblastoma, is associated with the alternative lengthening of telomeres (ALT) phenotype; a telomere lengthening mechanism that is independent of telomerase (Schwartzentruber et al., 2012). IDH mutations are tightly associated with ATRX mutations and are observed in grade II-III astrocytomas (71%), oligoastrocytomas (68%), and secondary glioblastomas (57%) (Jiao et al., 2012). The IDH/ATRX mutation signature is therefore a marker of astrocytomas and is associated with a poorer outcome (median 51 months). CIC and FUBP1 are seen in less than 10% of astrocytomas or oligoastrocytomas but are frequently observed in oligodendrogliomas (46% and 24%, respectively); these are associated with a more favourable prognosis (median survival 96 months) (Jiao et al., 2012).

1.5. Genetics of high grade glioma.

Glioblastoma was the first disease to be studied by The Cancer Genome Atlas (TCGA), a large program designed to integrate comprehensive data on multiple cancer types, including gene expression, methylation, copy number and microRNA expression (TCGA, NIH). This produced a wealth of information on a large dataset of glioblastoma patients (>500 cases) (Cancer Genome Atlas Research Network, 2008). Since then studies performed by the TCGA and other laboratories have improved our understanding of the genetics of glioblastoma (Verhaak et al., 2010; Brennan et al., 2013). More recently, the TCGA released data for lower grade glioma including grade II and III gliomas (>500) (Gonda et al., 2014).

1.5.1. Molecular subtypes of glioblastoma.

The first study to show that glioblastoma could be separated into subgroups using molecular information was performed at Genentech in 2006 and showed that, based on transcriptional profiling data, three subtypes of glioblastoma exist; named proneural, mesenchymal and proliferative. The proneural subtype tumours bear resemblance to a neuronal lineage, express histological markers including OLIG2 (oligodendrocyte lineage transcription factor 2), DLL3 (*Drosophila* delta homolog) and BCAN (brevican) and have the best outcome of

the three subtypes. The mesenchymal and proliferative subtypes are associated with losses of chromosome 10, gains of chromosome 7, PTEN loss and EGFR amplification. The proliferative subtype tumours express the markers PCNA (Proliferating Cell Nuclear Antigen) and TOP2A (Topoisomerase II Alpha 170kDa) and are able to grow in the absence of EGF (epidermal growth factor) and FGF (fibroblast growth factor). The mesenchymal subtype expresses CHI3L1 (chitinase 3-like 1) and the angiogenic marker VEGF. Recurrent tumours have been shown to shift towards this phenotype (Phillips et al., 2006). Further studies using TCGA data in 2010 reported an additional molecular subtype (Verhaak et al., 2010). Verhaak *et al* clustered mRNA expression patterns of glioblastoma into four subtypes; proneural, classical, neural and mesenchymal based on 840 mRNAs (210 mRNAs defining each subgroup). The classical subtype tumours were associated with chromosome 10 loss and chromosome 7 gain (although trisomy chromosome 7 was seen in other subtypes). This group exhibited high-level *EGFR* amplification along with mutant *EGFR*. Homozygous deletions at chromosomal band 9p21.3 were also evident in this subtype, notably in *CDKN2A* (Cyclin-Dependent Kinase Inhibitor 2A), and these were almost mutually exclusive with other components of the retinoblastoma pathway. The mesenchymal subtype was associated with deletions at chromosomal band 17q11.2, which includes *NF1* (neurofibromin 1). Markers reported to be expressed in this subtype overlapped with that of the earlier Genentech signature; *CD44* and *CHI3L1*. Genes in the tumour necrosis factor super family were expressed highly in this subtype, which may be a reflection of high necrosis and inflammatory infiltration. The proneural subtype showed a high rate of alteration of *PDGFRA* (platelet-derived growth factor receptor, alpha polypeptide) and *IDH1*. There was high expression of oligodendrocyte development genes including *PDGFRA*, *NKX2-2* (NK2 homeobox 2) and *OLIG2*. Proneural development genes were also expressed in this group, such as *DLL4* detected in the Phillips' proneural group (Phillips et al., 2006).

The fourth subtype, neural, was characterised by expression of neuron markers, for example *NEFL* (neurofilament, light polypeptide) and *GABRA* (gamma-aminobutyric acid (GABA) A receptor). Gene ontology categories for

the genes expressed in this subtype pointed to neuron projection and axon and synaptic transmission (Verhaak et al., 2010).

The Verhaak subtypes show mRNA expression that is resonant with various characterised neural cell types; the proneural subtype bore resemblance to an oligodendrocytic signature, the classical subtype was aligned with a murine astrocytic signature, the neural subtype showed high expression for genes expressed by neurons and the mesenchymal subtype showed a similarity to cultured astroglia (Verhaak et al., 2010). Methylation of the *MGMT* promoter was not associated with subtype, but the subtypes did differ in response to treatment. The classical and mesenchymal subtypes were the only subtypes shown to respond to treatment with radiotherapy and TMZ (Verhaak et al., 2010). The longer survival of the patients with a proneural glioblastoma subtype may therefore be due to the younger age of these patients (Phillips et al., 2006). The subtypes determined in 2010 by Verhaak *et al.* were revised in 2013, when further data was made available in the TCGA (Verhaak et al., 2010; Brennan et al., 2013). This confirmed these four subtypes; neural, proneural, classical and mesenchymal. In addition, they further separated the proneural group into G-CIMP positive proneural and G-CIMP negative proneural. Astonishingly, these two groups represented the opposite ends of the spectrum in prognosis for glioblastoma. Patients with a G-CIMP positive proneural subtype were the group with the best prognosis of all glioblastoma patients in the TCGA, and the G-CIMP negative proneural tumours conferred the worse prognosis of all glioblastomas (Brennan et al., 2013).

G-CIMP refers to a subgroup of gliomas, with a more favourable prognosis, that have extensive DNA methylation across over 1500 loci in the genome (Noushmehr et al., 2010). Methylation of DNA occurs predominantly at the dinucleotide CG in vertebrates. Cytosine and guanine are separated by one phosphate, and therefore these sites are often termed CpG sites (cytosine-phosphate-guanine). Cytosines have the ability to become methylated at the 5-position in this context, forming 5-methylcytosine, and in mammals this methylation can serve to turn off gene expression (Cooper, 1983). Regions with a high frequency of CpGs in the genome are called CpG islands, representing 1% of the genome (Vinson & Chatterjee, 2012). In cancer, many tumour suppressor genes are inactivated by this mechanism, including those involved

in DNA repair, apoptosis and cell cycle pathways (Esteller, 2002). The finding that proneural gliomas with the G-CIMP have a better prognosis may reflect the fact that they have different mechanisms of origin. It has recently been shown that initial transforming events in G-CIMP gliomas are different to those that arise in the formation of non-G-CIMP glioblastomas, which result from NF1 loss in a proneural context (Ozawa et al., 2014).

A further signature was developed based on glioma stem cells expanded from glioma specimens, and separated these into proneural and mesenchymal subtypes based on eight mRNAs (Mao et al., 2013). The proneural subtype showed higher expression of the proneural markers *CD133* (CD133 antigen), *OLIG2*, *SOX2* (SRY (sex determining region Y)-box 2) and *NOTCH1* (notch 1) and the mesenchymal subtype showed high expression of *CD44* (CD44 antigen), *LYN* (V-Yes-1 Yamaguchi Sarcoma Viral Related Oncogene Homolog), *WT1* (Wilma's tumour 1) and *BCL2A1* (BCL2-related protein A1). The mesenchymal subtype was shown to be more aggressive in *in vitro* and murine intracranial xenograft assays. The subtypes also had different features *in vitro* with proneural cells forming spherical neurospheres and the mesenchymal cells forming irregular aggregates with some cells having adherent properties when grown in the absence of serum (Mao et al., 2013).

1.5.2. Heterogeneity and clonal evolution.

Glioblastoma Multiforme is so named due to its high degree of heterogeneity. Many studies have highlighted the various levels at which heterogeneity can occur in glioblastoma, including cellular, molecular, metabolic, genetic and epigenetic levels (Vartanian et al., 2014). This is a considerable barrier towards the success of therapies for this disease, not only because each patient's tumour is different (inter-tumour heterogeneity), but also because the tumour contains a wealth of different cell types that may be resistant to a particular therapy (intra-tumour heterogeneity).

Inter-tumour heterogeneity has been described in 1.5.1, where glioblastoma can be clustered into at least 3 subtypes according to molecular genetics. Despite a similar diagnosis based on histology, different tumours may have different mutations and gene expression patterns; they occur in patients of different ages, and are located in different regions of the brain (Larjavaara et al.,

2007; Bozdag et al., 2013). These tumours show differences in their response to current treatments and some are more aggressive than others (Brennan et al., 2013).

It has since been discovered that more than one of these molecular subtypes may be present within one tumour (Sottoriva et al., 2013). It is clear that although a tumour may display the overall transcriptional pattern of one particular subtype, there may be therapy resistant subclones present within that tumour that confer an advantage under such a selection process (Meyer et al., 2015). Clonal evolution refers to the number (how many subclones), hierarchy (time) and importance (size of the subclone) of each cellular population within a tumour that has arisen in response to the environment. For example, when a tumour has been exposed to chemotherapy or radiotherapy certain cell types will be killed; this leaves remaining/resistant cell types the chance to expand, creating a tumour with different populations. This has been shown by the altered subpopulations of cells in recurrent glioblastoma in relation to the initial lower grade tumour (Johnson et al., 2014).

1.5.3. Cell of origin.

Heterogeneity studies of glioblastoma raise issues as to the cell of origin and initial oncogenic events in this disease. Ozawa *et al.* reported that most non-G-CIMP gliomas arise from a common proneural precursor (Ozawa et al., 2014). Gains of chromosome 7 and loss of chromosome 10 were the first events in glioblastoma, and elevated PDGFRA was an initial driver in non-G-CIMP glioblastoma. Subsequent loss of NF1 was then sufficient to induce a mesenchymal gene expression pattern (Ozawa et al., 2014). These alterations are not sufficient to induce gliomagenesis in all neural stem cell types, and specifically oligodendrocyte precursors (OPs) have been suggested as the cell of origin in which gliomagenesis is initiated (Liu et al., 2011a). OPs are the precursors of both oligodendrocytes and astrocytes (Fig. 1.6) and have the ability to form astrocytic and oligodendroglial tumours through deletions of *CDKN2A* (Lindberg et al., 2014). It has also been postulated that somatic mutations may occur prior to tumour initiation by associating the number of somatic mutations in a patient's tumour with their age (Tomasetti & Vogelstein, 2013).

G-CIMP tumours are a separate group of glioblastomas that arise in younger individuals, most frequently in the frontal lobe (Sturm et al., 2012). These are therefore likely to have arisen from a neural precursor population present in this spatial and temporal context of the brain (Lai et al., 2011). The initial event for these tumours is most likely the IDH1 mutation followed by p53 mutation, because the probability for C>T mutations is the highest (Lai et al., 2011). The particular mutations that are acquired in nature suggest that the p53 mutation occurs on the coding strand and the IDH1 mutation occurs on the template strand. IDH1 mutant protein therefore will be expressed immediately, whereas a round of replication must take place before mutant p53 protein is expressed. In a cell proliferating slowly this could be a considerable time (Lai et al., 2011). The IDH1 mutation has been shown to be sufficient, alone, to generate the G-CIMP signature by measuring changes in the methylome when mutant IDH1 is introduced into cells (Turcan et al., 2012).

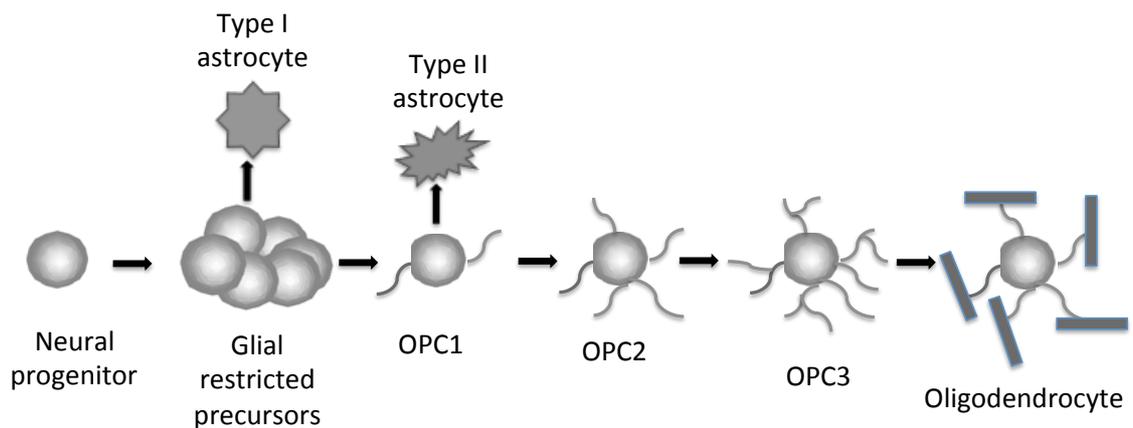


Figure 1.6. The differentiation pathway involving oligodendrocyte precursors.

Neural progenitors first become glial-restricted, and these glial restricted cells have multipotent capacity in that they can differentiate into astrocytes and also oligodendrocytes. A glial-restricted precursor may differentiate directly into an astrocyte, or into an oligodendrocyte precursor, which progresses through various differentiated states before terminally entering the oligodendrocyte lineage (Letzen et al., 2010). Neural progenitors are defined as cells expressing the ganglioside epitope A2B5 and the intermediate filament protein nestin. Glial restricted precursors start to express PDGFRA and OLIG1. OP cells begin to express O1 (oligodendrocyte marker 1) and O4 (oligodendrocyte marker 4) and later express O1, GalC and CNPase markers. Mature oligodendrocytes can be distinguished by their expression of myelin basic protein, which is important for their function.

1.5.4 Altered pathways.

The first study by the TCGA identified the p53 (>31%), retinoblastoma (44%) and receptor tyrosine kinase pathways (>83%) as significantly altered in glioblastoma (Verhaak et al., 2010). The p53 pathway is altered in many cancers and its role is in responding to stress that can cause infidelity of DNA replication, disrupting cell division. Stress signals are conveyed to the p53 protein through post-translational modifications, and this elicits a transcriptional network influencing cell cycle checkpoints leading to senescence and apoptosis (Harris & Levine, 2005). Loss of functional p53 protein is reported in many cancers and is associated with the transition of a cell from an epithelial type to a

mesenchymal type (epithelial-mesenchymal transition; EMT) (Muller et al., 2011).

The p53 pathway is linked to the second network significantly altered in glioblastoma; the retinoblastoma pathway, through p14Arf (Bates et al., 1998). This pathway, discovered from its association with the heritable development of retinoblastoma (an eye tumour) is altered in almost all cancers and is essential for initiation of replication (Nevins, 2001). Loss of p53 and Rb, or other components of the pathway, allows the cell to pass through cell cycle checkpoints, avoiding apoptosis, resulting in daughter cells with genetic aberrations.

The third significantly altered network in glioblastoma involves the receptor tyrosine kinases (RTKs). These are transmembrane proteins that transduce an external signal into the cell. The N-terminus of these proteins is extracellular, and acts as a receptor for ligands such as epidermal growth factor (EGF). The C-terminus has kinase activity; it phosphorylates intracellular substrates to activate downstream signalling cascades (Hubbard & Till, 2000). These signalling cascades are only activated when the receptor has a bound ligand, however in cancer, mutation of these receptors allows them to be constitutively active, without the requirement of ligand binding (Ballotti et al., 1989). More recently it has also been shown that proteolytically cleaved TKIs have the ability to migrate to the nucleus where they can directly exert their effects (Song et al., 2013a). Altered TKIs in cancer are targets for therapies, and the first was Imatinib (brand name Gleevec) for chronic myeloid leukaemia (CML) (Druker et al., 1996). This drug binds to the ATP binding site of the constitutively active TKI formed by the Bcr-Abl fusion blocking its catalytic action. This resulted in a high patient response rate although subsequent resistance through mutation quickly became apparent (Mauro & Druker, 2001; Shannon, 2002). Since then a number of first and second generation RTK inhibitors have been approved for treatment of various cancers, most of which require detection of alteration of the TKI they inhibit in the patient's tumour before administration (Cohen et al., 2003; Kwak et al., 2010). One such pathway influenced by RTKs that was highlighted by the TCGA study in 2010 is the phosphoinositide-3-kinase (PI3K) pathway, which is an important intracellular cell cycle pathway (Verhaak et al., 2010). It is the class IA PI3Ks that are influenced by RTKs, and these are

recruited to the membrane and convert phosphatidylinositol-4,5-bisphosphate (PIP₂) to phosphatidylinositol-3,4,5-triphosphate (PIP₃) which provides docking sites for kinases such as AKT (Liu et al., 2009).

The second study by the TCGA shed more light on the significant aberrations in glioblastoma through sequencing analysis (Brennan et al., 2013). This showed that, in addition to the pathways identified to be significantly altered in glioblastoma previously, more than 40% of tumours have a non-synonymous mutation in a chromatin modifier gene (Brennan et al., 2013). Pathways involved in the different molecular subtypes were determined showing classical glioblastoma had a down-regulation of pro-apoptotic proteins, mesenchymal glioblastoma exhibited an increase in endothelial markers and the MAPK (mitogen activated protein kinase) pathway, proneural glioblastoma showed elevation of the PI3K pathway and G-CIMP positive tumours showed similarity to proneural glioblastoma with an increase in Cox-2, IGFBP2 (Insulin-Like Growth Factor Binding Protein 2) and Annexin-1 (Brennan et al., 2013).

1.5.5. EGFR alteration and EGFR variant III.

The RTK EGFR is frequently altered in cancer, through amplification, rearrangement and mutation (Gan et al., 2009; Li et al., 2014a; Reguart & Remon, 2015). In glioblastoma, EGFR is altered by all these mechanisms and amplification of the receptor is observed in approximately 40% of tumours. Around 50% of tumours with amplified EGFR express a particular rearrangement involving the extracellular binding domain known as EGFR variant III (EGFRvIII) (Sugawa et al., 1990; Ohgaki & Kleihues, 2013). This variant has a deletion of exons 2-7 of the *EGFR* gene resulting in a receptor that cannot bind to a ligand and therefore remains constitutively active (Gan et al., 2009). The variant has been shown to be associated with a better prognosis in glioblastoma, with EGFRvIII negative cell populations in the tumour being radio- and chemo- resistant (Montano et al., 2011). EGFR amplification itself however is associated with a poorer prognosis in some patient groups such as those bearing tumours with gain of chromosome 7, or younger patients treated with radiotherapy (Bienkowski et al., 2013). In other cancers, EGFR inhibitors have been highly successful in improving patient outcome, and the high frequency of EGFR in glioblastoma fuelled evaluation of these inhibitors in this

disease also (Cohen et al., 2003; Lewis et al., 2012; Reardon et al., 2014b). Clinical trials in newly diagnosed and recurrent glioblastoma have generally concluded that first generation EGFR inhibitors offer no improvement in outcome for glioblastoma patients (Reardon et al., 2014b). These results may be due to low brain penetration, or mutations in signalling pathways downstream of EGFR as well as high tumour heterogeneity (Cancer Genome Atlas Research Network, 2008; de Vries et al., 2012; Sottoriva et al., 2013). Biomarkers have been identified for these inhibitors, but have been shown to have no clinical value in the trials to date (Mellinghoff et al., 2005; Verhaak et al., 2010; Reardon & Wen, 2014). A vaccine for EGFRvIII has also been developed by Celldex Therapeutics and has been evaluated in various clinical trials using this in combination with other therapies. Phase II trials for newly diagnosed glioblastoma have shown EGFRvIII was eliminated in 67% patients after three months of therapy and was well tolerated (Del Vecchio & Wong, 2010; Schuster et al., 2015).

1.6. The importance of stem cells in glioma.

Cancer stem cells were first identified in the 1990s in acute myeloid leukaemia. Glioma also has a stem cell niche (Dick, 1991). Glioma stem cells (GSCs) are characterised by their self renewing properties and multipotent abilities allowing them to differentiate into all tissue types and cells that have arisen within a tumour, through precursor stages (Fuchs & Segre, 2000). GSCs bear resemblance to normal stem cells, but lack the ability to tightly regulate the proliferation and differentiation into integrating cell types (Venere et al., 2011). Stem cells that remain following therapy allow the tumour to regrow and therefore they are an important group of cells within the tumour.

It was in 2002 that GSCs were first isolated, and shown to grow as clonogenic spheroids *in vitro* (Ignatova et al., 2002). These cells were shown to differentiate into cell types that form the initial tumour and expressed the neural stem cell surface marker CD133. This marker is often used to isolate these cells from the tumour of a patient by fluorescence-activated cell sorting (Singh et al., 2003). CD133 is not a specific marker for GSCs however, as CD133 negative stem cells have been isolated from glioblastomas. The use of this marker is also complicated by the fact that some antibodies for CD133 recognise only the

glycosylated form which generates false negatives (Bidlingmaier et al., 2008). CD133 negative cells are adherent in their growth pattern whereas CD133 positive cells grow as spheres (Brescia et al., 2012). Combinations of markers have been proven to be most useful and these may include CD44, CD15, L1CAM and integrin $\alpha 6$ (Brescia et al., 2012).

Glioma stem cells have been shown to exhibit high aldehyde dehydrogenase (ALDH) activity, and this has also been exploited for isolation of stem cells of other cancers (Douville et al., 2009). It has been shown however that this activity is more apparent in mesenchymal glioma stem cells compared to proneural stem cells (Mao et al., 2013). Alternatives to the isolation of glioma stem cells using molecular markers include exploiting the auto-fluorescence properties and morphology of glioma cells. For example, cells that simply are able to form neurospheres in the absence of serum represent part of the population of stem cells, and have been shown to have self-renewal capacity (Yuan et al., 2004). Also dye retention over cell divisions can be exploited to identify GSCs, as dye is equally divided between daughter cells at division, the slower the division the less the dye dilution in the cells (Deleyrolle et al., 2011). The origin of GSCs has been of considerable debate. One explanation is that normal neural stem cells transform into glioma. The type of glioma to arise may be as a result of the environment or type of genetic alterations the cell has undergone. This is supported by the fact that deletion of *p53*, *NF1* or *PTEN* in neural stem cells is sufficient to generate glioma, whereas these aberrations in non-stem brain cells does not cause this effect (Alcantara Llaguno et al., 2009). Alternatively, glioma cells may have the ability to reprogram into GSCs under certain conditions. This is supported by the observation that neonatal cortical astrocytes can dedifferentiate into neural stem cells by deletion of *p16Ink4a* and *p19Arf* (Bachoo et al., 2002).

Single cell genomic analysis of glioblastomas has shown that a stem-like compartment of cells does exist within the tumour, but that a continuum exists from stem cell to differentiated cell within a single tumour (Patel et al., 2014). Therefore, *in vitro* models embody extremes of the stemness of a tumour and the full spectrum of stemness is not represented. The genomic stemness signature is strongest in single cells isolated from proneural tumours, which is also supported by the fact that the stem cell signature for proneural GSCs

includes a number of stem-like markers such as *SOX2* and *CD133* (Mao et al., 2013; Patel et al., 2014). Despite tight correlation with a stem cell signature, proneural tumours with the highest heterogeneity had the worst prognosis, suggesting that clinical outcomes are dependent upon the level of heterogeneity, as well as presence of GSCs (Patel et al., 2014). The exposure of the tumour cell population to TMZ can increase the proportion of GSCs within a tumour, exhibited by an increase in the stemness markers such as *CD133*, *SOX2*, *OCT4* and *Nestin* (Auffinger et al., 2014). These dynamic levels of GSCs within the tumour, and possibly their location, contribute to the infiltration and therapy resistant properties of the tumour therefore allowing evasion of current treatments.

1.7. Novel chemotherapeutics for glioblastoma.

As previously mentioned, one major hurdle for the treatment of glioblastoma is the ability to deliver adequate amounts of any chemotherapeutic to the brain. Drugs that have been successful in other cancers have been trialled in glioblastoma but haven't been effective because of this, or because the tumour is highly heterogeneous compared to other tumours. This has been shown by the lack of success of TKIs in glioblastoma compared to other tumours (Reardon et al., 2014b). Inter-tumour heterogeneity in glioblastoma also makes personalised medicine essential, and the biomarkers used to define patient groups may not have been specific enough.

Targeted immunotherapy with monoclonal antibodies (mAb) for highly expressed proteins in glioblastoma has been explored. For glioblastoma with amplified EGFR, the mAb (monoclonal antibody) cetuximab has been shown to have some success (in recurrent glioblastoma) and the type of EGFR mutation may be of importance in patient outcome (Hasselbalch et al., 2010; Lv et al., 2012). AMG 595, which is a mAb for EGFRvIII conjugated to a cytotoxic (maytansinoid DM1) is currently under clinical trial and has been shown to cause disruption and internalisation of microtubules which inhibits the proliferation of glioblastoma cells (Hegde et al., 2014).

Glioblastomas express high levels of VEGF, which contributes to their vascularity, and this has prompted studies into the mAb bevacizumab (brand name Avastin) (Bao et al., 2006). Bevacizumab acts by neutralising the activity

of VEGFA and prevents its binding to the VEGF receptors on endothelial cells, which is important for vasculature formation and angiogenesis (Bao et al., 2006). This drug was approved by the United States Food and Drug Administration (US, FDA) for recurrent glioblastoma in 2009 after phase II studies showed a partial response in at least 19% of patients (Cohen et al., 2009). Since then, two large phase III trials evaluated bevacizumab for newly diagnosed glioblastoma. RTOG 0825 was a trial in Europe and AVAglio a trial in the US (Gilbert et al., 2014; Chinot et al., 2014). These studies were similar in many respects, as they included the standard treatment of TMZ and radiotherapy and evaluated outcome with the same statistical techniques. There were some differences in the studies however. RTOG 0825 excluded poor prognosis patients investigated by biopsy whereas AVAglio allowed all patients who had tumour biopsies to be included. Determination of treatment response was assessed by enhancing tumour only in RTOG 0825 whereas both enhancing and non-enhancing tumours were used for determination of treatment response in AVAglio. Despite these differences, results from the two trials were similar: there was no overall survival benefit from the drug but improved PFS (median PFS in the bevacizumab-treated group was 10.6 months compared to median PFS in the placebo group of 6.2 months) (Chinot et al., 2014; Gilbert et al., 2014).

Mabs have recently been developed to reactivate anti-tumour immunity by blocking immune checkpoint molecules on T-cells. This has shown some remarkable responses in melanoma, and is now being applied to glioblastoma (Cooper et al., 2014; Reardon et al., 2014a). The blockade of the programmed cell death 1 (PD-1) and its ligand programmed death ligand-1 (PDL1) show promise in both leukaemia and solid tumours (Topalian et al., 2012). Nivolumab blocks activation of PD-1 which allows activation of cytotoxic T-lymphocytes against glioblastoma cells and ipilimumab enhances this cytotoxic T-lymphocyte activation by binding to CTLA-4 (cytotoxic T-lymphocyte-associated antigen-4) (Hegde et al., 2014). Patients are now being recruited to test this combination in recurrent glioblastoma (NCT02017717).

1.8. The study of glioma cell biology in the laboratory.

Most research in the glioma field starts in the laboratory using *in vitro* assays. The types of cells and methods of culture used by different laboratories can vary. Established human cell lines for high-grade glioma are available from the American Type Culture Collection (ATCC), and include U251, a pleiomorphic astrocytoid glioblastoma line, U87, an epithelial glioblastoma line and LN229, a glioblastoma line from the right frontal parietal-occipital cortex which is epithelial in morphology (ATCC information, February 2015) (Ponten & Macintyre, 1968). These cell lines are usually assessed annually by short tandem repeat profiling against the ATCC profile to ensure their integrity. Established glioblastoma lines grow as adherent monolayers, in media containing serum and will proliferate indefinitely if monitored and passaged appropriately.

In addition to these established cell lines, some techniques require study of cells more representative of the tumour that also reflect inter-tumour heterogeneity, and therefore primary cell lines are expanded from a patient sample. These are mostly anchorage-dependent, slow-growing lines that will grow for a period of time before entering senescence, and the time is based on nutrient conditions, culture manipulation and the Hayflick limit (Shay & Wright, 2000). The cells from a tumour sample are heterogeneous, and the method of culture will result in selection for populations within the tumour. Due to this, the culture conditions are chosen to reflect the types of cells to be studied. Cells with different anchorage dependencies may be selected for, or cells may be sorted using fluorescence-activated cell sorting for particular cell markers. GSCs are often required for culture, and in this instance cells would be selected for using serum-free media, and those cells with spheroidal growth would be separated from the adherent cells. The spheroids, which represent anchorage independent GSCs can then be grown as neurospheres in low adherent flasks, or as monolayers on laminin-coated flasks. These neurospheres are genetically more representative of the original tumours than established cell lines (Lee et al., 2006).

Cultures are usually incubated at 21% oxygen, since this is the level of atmospheric oxygen, however this does not reflect physiological oxygen levels which often fall below 3%, and even lower in the hypoxic regions of a growing

glioblastoma (Evans et al., 2004). It has been shown that stem cells culture more easily below 7% oxygen (Guo et al., 2013; McCord et al., 2009). This is not practical however without specialist equipment for manipulation of the cultures under these conditions.

1.9 MicroRNAs.

MicroRNAs have attracted a huge amount of interest in the cancer field over the last decade. These short non-coding RNAs play important functional roles in tumour biology, and show promise as biomarkers for patient stratification and treatment monitoring. MicroRNAs were first identified in the nematode *Caenorhabditis elegans* in 1993 (Lee et al., 1993), and by the early 2000s it was recognised that microRNAs have a conserved mechanism and broad functional significance throughout the plant and animal kingdoms. Their function is mainly to regulate protein translation by binding to complementary sequences in the 3' untranslated region (UTR) of target messenger RNAs (mRNAs), which either blocks translation or causes transcript degradation (Krol et al., 2010). At present, there are over 2000 mature human microRNAs recorded in miRBase (August 2013), a searchable annotated database of known microRNA sequences (Griffiths-Jones et al., 2008; Wang et al., 2011; Jacob et al., 2013). The first cancer-associated microRNAs, miR-15 and miR-16, located at 13q14.3 which is a frequently deleted region in chronic lymphocytic leukaemia (CLL), were identified in 2002 (Calin et al., 2002). Subsequent studies have shown that microRNAs play important roles in all recognised cancer hallmarks, and that each tumour type has a distinct microRNA signature that distinguishes it from other cancers and normal tissues. Many cancers can be further sub-classified based on these signatures. Like other cancer-associated genes, microRNA expression can be altered by chromosomal amplification/deletion, methylation and transcription factor activation. Alterations in microRNA processing pathways and target site binding are also common features of cancer cells. Recent years have seen the increased use of molecular diagnostic approaches to refine cancer detection, diagnosis and treatment (Chambers et al., 2012; Chang et al., 2013; Conde et al., 2013; Cushman-Vokoun et al., 2013). MicroRNAs offer an additional genetic component that can be exploited to stratify patients with greater accuracy and may be most useful when integrated

with gene expression and other clinical factors known to robustly predict outcome. In fact, they may be more useful than mRNAs as prognostic indicators due to their stability within clinical samples and their robust expression (Lu et al., 2005). Additionally, the identification of tumour-derived microRNAs in the circulation, and the development of robust assays for sensitive and accurate microRNA detection may allow their use as serum biomarkers. This is an area of intense study, representing a non-invasive method of detection and diagnosis of cancer.

1.9.1. MicroRNA processing and mechanism of action.

The biogenesis of microRNAs occurs through a well-characterised conserved processing mechanism (Fig. 1.7). MicroRNAs are encoded in the genome, and are often expressed as clusters of two or three microRNA hairpins. They also may be encoded in unique transcripts or in introns of protein coding genes. After processing, the mature single stranded microRNAs typically bind to messenger RNA targets in their 3'UTRs, and result in either reduced translation or deadenylation and degradation depending on the degree of base-pairing complementarity with the so-called "seed" region at the 5' end of the microRNA (Filshtein et al., 2012) (Fig. 1.8). Because the microRNA/mRNA binding site is short (6-8 base pairs), each microRNA has the potential to target multiple different mRNAs. It is estimated that collectively microRNAs have roles in regulating up to two thirds of the human genome (Nana-Sinkam & Croce, 2012). Changes in microRNA expression can result in reprogramming of cellular functions, where they play roles in fundamental processes such as development, cellular homeostasis and adaptation to stress. MicroRNA alterations promote a number of pathological conditions as well as cancer (Ebert & Sharp, 2012).

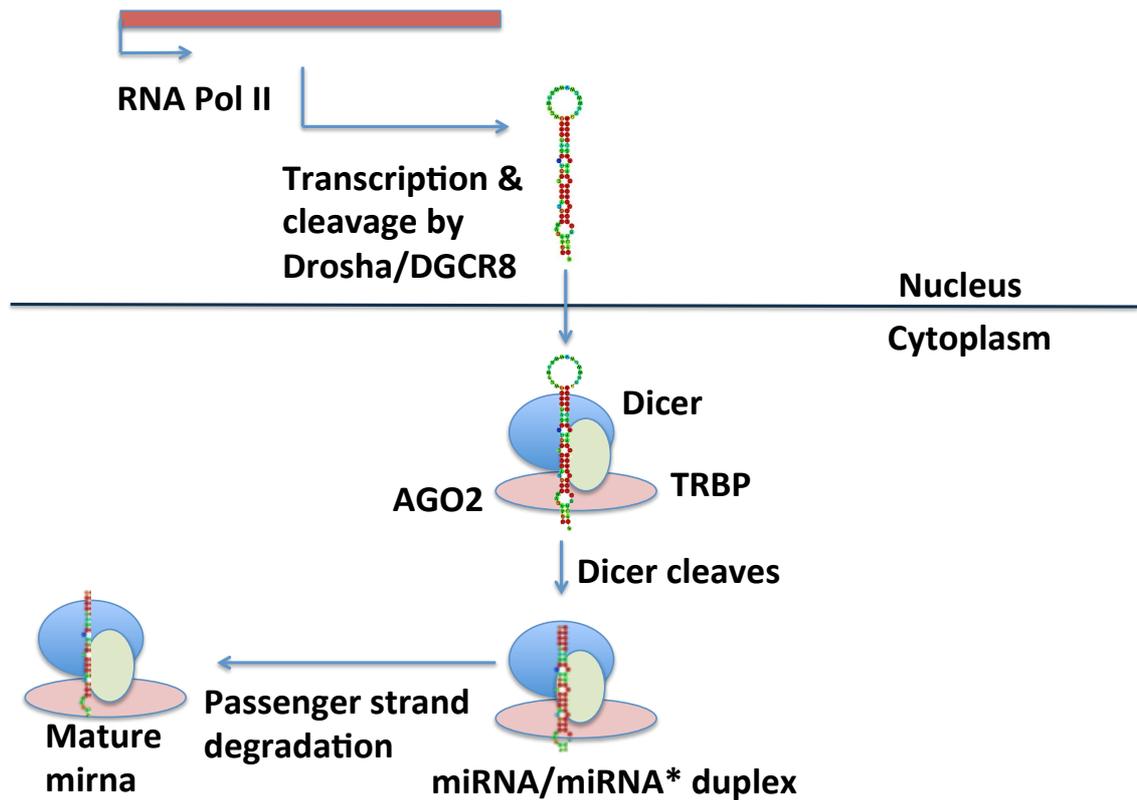


Figure 1.7. MicroRNA biogenesis.

MicroRNAs are transcribed in the nucleus into a primary transcript by RNA polymerase II. These structures have intra-molecular base pairing forming distinct hairpin secondary structures, which are cleaved by Drosha (a type III ribonuclease) and DGCR8 (DiGeorge syndrome critical region 8) in the nucleus to form a 70-nucleotide precursor microRNA molecule (Krol et al., 2010). The precursor is then exported to the cytoplasm by exportin-5 (Melo et al., 2010). Some microRNAs bypass this mechanism, and are produced from very short introns (mirtrons) by splicing and debranching, an activity known as non-canonical processing of microRNAs (Krol et al., 2010). Following exportation, the pre-microRNA is then cleaved by the RNase III Dicer, in conjunction with TRBP (transactivation responsive protein) and AGO2 (Argonaute 2) in the cytoplasm, which yields a microRNA/microRNA duplex. One strand of the duplex (usually the one with the less stable 5' end) is then preferentially incorporated into a microRNA-induced silencing complex (miRISC) whereas the other strand is usually degraded (Krol et al., 2010).*

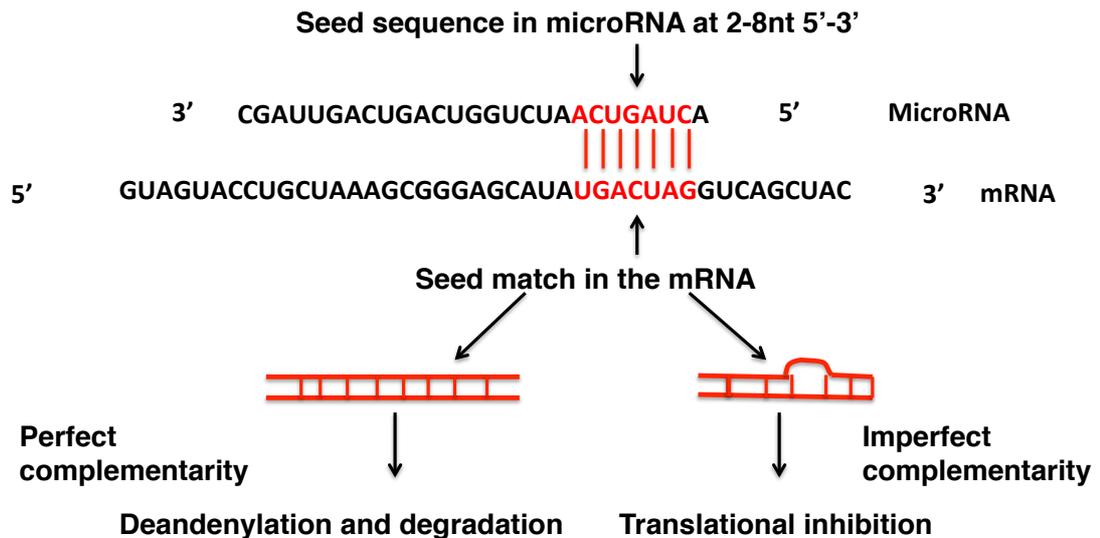


Figure 1.8. Seed sequences of microRNAs and their relevance in the determination of the mRNA fate.

The seed sequence for microRNA binding to the target is between the 2nd to the 8th base pair from the 5' end of the microRNA. This can bind to any region within the 3' UTR (untranslated region) of an mRNA molecule and there can be multiple seed sequences for a microRNA/microRNAs in one 3' UTR. Perfect complementarity of the microRNA and the target leads to deadenylation and degradation. Imperfect complementarity, which is the most common microRNA-target interaction in animals, results in translational inhibition (Krol et al., 2010).

1.9.2. MicroRNAs as predictors of prognosis.

A microRNA expression signature in the six most common human cancers (breast, prostate, lung, stomach, pancreas and thyroid) was identified by Carlo Croce's lab in 2006. This study used microarrays on 540 frozen tumour samples to computationally identify 57 differentially expressed microRNAs in cancer compared to normal tissue (Volinia et al., 2006). The predicted targets of some of these microRNAs were subsequently validated in light of the context of the different tumours. Since then a number of signatures and individual microRNAs have been associated with prognosis in cancer (Hu et al., 2010; S. Srinivasan et al., 2011).

Variations in sample preparation and patient groups can lead to differing conclusions, which is illustrated in attempts to generate signatures in the grade

IV brain tumour glioblastoma multiforme, which has extensive microRNA and mRNA expression data available in TCGA (Cancer Genome Atlas Research Network, 2008). Glioblastoma, as previously mentioned, is a disease where prognosis is particularly difficult to predict, and therefore microRNA expression has been used repeatedly to attempt to stratify patients into good and poor prognosis groups. Of five signatures described in glioblastoma (Lakomy et al., 2011; Niyazi et al., 2011; Srinivasan et al., 2011; W. Zhang et al., 2012a, Sana et al., 2014), only miR-31 and miR-195 were identified as predictors in more than one signature (two signatures associated both with poorer survival). This discordance between studies could be attributed to various factors. The Lakomy and Niyazi studies used formalin-fixed paraffin-embedded (FFPE) samples whereas the others used frozen tissue, although all studies used an optimised microRNA extraction protocol and microarrays or qRT-PCR with locked nucleic acid (LNA) primers for quantification. There were differences in treatment with patients given standard chemotherapy of temozolomide ranging from 35% to 100% of the cohort. There were also differences in cohort size (35-354 patients) and geographical origin. These differences highlight some problems in generating useful microRNA expression signatures for clinical prognosis assessment. This has been addressed to some degree by the use of standardised procedures and large sample sizes, as has been attempted by TCGA. Should a successful signature be generated using correlative data of this type, although not ideal due to the indirect association with tumour biology, it would greatly improve the clinical decisions for patient management, particularly in diseases with wide ranging survival times such as glioblastoma.

1.9.3 MicroRNAs for classification of disease.

Expression of microRNAs can be clustered based on embryonic or developmental origin (Lu et al., 2005) which makes them ideal for classification of cancers arising from different cell lineages. For example, leukaemia is a disorder of hematopoietic stem cells and is currently classified by the WHO according to the lineage of the progenitor cell. MicroRNA expression across subtypes appears to reflect this. Garzon *et al* generated expression signatures by microarray in acute myeloid leukemia (AML) and associated these with specific clinically relevant cytogenetic abnormalities in 122 untreated patients.

These alteration-specific signatures were validated in 60 patients using qRT-PCR (Garzon et al., 2008). Eight microRNAs were over-expressed and 14 under-expressed in AML with 11q23 rearrangements compared to all other AML samples. In AML with trisomy 8, 42 microRNAs were over-expressed, two of which, miR-124a and miR-30d, are located on chromosome 8. In addition, miR-155 was associated with LT3-ITD mutations and miR-181a was decreased in expression in AML with multi-lineage dysplasia. These subtypes are used worldwide for AML diagnosis and this study clearly shows that microRNAs can also be used as delineators of the disease. In CLL, microRNAs were shown to define currently cytogenetically classified tumours with a normal karyotype, those exhibiting deletions of 11q, 13q or 17p and those with trisomy 12 (Visone et al., 2009). Glioblastoma classification using clustered microRNAs based on the cell of origin has also been performed and shown to successfully stratify this high grade brain disease into five groups (Kim et al., 2011b). Distinct microRNA expression patterns have also recently been identified in luminal (epithelial origin), basal-like (myoepithelial origin) and HER2 (human epidermal growth factor receptor 2) breast cancers (Farazi et al., 2014). Although the classification of breast cancer is well-defined compared to some other malignancies, meta-analysis of recent clinical trials have shown incorrect classification of a substantial number of tumours in laboratories with high volume testing (Andorfer et al., 2011) and therefore microRNA analysis may add robustness to current testing. Similarly, in prostate cancer, microRNA patterns are distinct between different cellular subsets when stem/progenitor cells were isolated from prostate tumours indicating microRNA expression patterns are indicative of the cellular populations in a tumour (Liu et al., 2012). These results suggest that when classification of the tumour is dependent on the progenitor cell type, microRNAs are useful in separating these classes. In further support of this, microRNA expression convincingly classified a set of 22 different tumour types according to tissue of origin in a blind study (Rosenfeld et al., 2008) and a study of less well-differentiated tumours showed that microRNAs are better delineators of tumour type than mRNAs (Lu et al., 2005). Subtypes in other cancers have been identified using the huge body of data at TCGA (Koboldt et al., 2012; Kloosterhof et al., 2013). These, although not employed in the clinic as yet, may also be stratified using microRNA expression

patterns. The identification of microRNAs that target current biomarkers may pave the way for microRNA-based tests as an alternative to mRNA/protein expression for prognosis assessment. One such example is the discovery of miR-155 as a target for the biomarker HGAL in diffuse large B-cell lymphoma and Hodgkin lymphoma (Dagan et al., 2012). MicroRNAs have also been shown to play a role in cancer progression and may be useful for the prediction of metastatic outcomes for patient management (Pencheva & Tavazoie, 2013).

1.9.4. MicroRNAs as predictors of drug efficacy.

With the advent of personalised and precision cancer medicine, drugs are increasingly administered to subgroups of patients most likely to respond. MicroRNA signatures can be used, in addition to other predictors, to identify patients likely to benefit from a drug (Rukov et al., 2013). These signatures should be established in large patient groups in the context of clinical trials using quality control criteria (McShane et al., 2013).

Although not yet used in clinical decision-making, several studies have associated microRNAs with well-known biomarkers for treatment therapy decisions. For example, chronic myeloid leukaemia (CML) is treated with the BCR-ABL inhibitor Imatinib. Levels of the BCR-ABL rearrangement, which characterise this disease, decrease over time with Imatinib treatment. It has been discovered that miR-451 levels inversely correlate with BCR-ABL levels (Lopotova et al., 2011) at both the time of diagnosis and upon treatment (Scholl et al., 2012). Likewise, miR-378 has been shown to predict response to anti-angiogenic treatments in ovarian cancer (Chan et al., 2014). Prior to these smaller studies, an *in silico* approach using the NCI-60 human cancer cell line panel showed approximately 30 microRNAs correlated with response to numerous anticancer drugs (Blower et al., 2008) which is evidence that microRNAs play a part in chemo-resistance and could be important in future testing for drug eligibility.

SNPs in microRNA target sites may also be predictors of response; the LCS6 polymorphism in the let-7 binding site in the 3' UTR of KRAS predicted response to anti-epidermal growth factor receptor (EGFR)-based therapy in 100 metastatic colorectal cancer patients (Sebio et al., 2013). Base excision repair genes have been associated with treatment resistance, and variations in the

microRNA binding sites of the 3' UTRs of these genes have been shown to reflect colorectal cancer prognosis and treatment response (Pardini et al., 2013). A notable and interesting example of altered target sites in cancer is the creation of an illegitimate target site for miR-191 in the 3' UTR of MDM4 by the presence of SNP34091, which affects chemosensitivity in ovarian cancer (Wynendaele et al., 2010). Aside from treatment resistance, it is worth noting that SNPs in microRNA binding sites increase the risk of developing cancer, and may be markers for genetic susceptibility studies in some cancers (Ziebarth et al., 2012).

1.9.5. MicroRNA-based therapeutics and companion diagnostics.

Several studies have focused on the use of microRNAs themselves, or anti-microRNA constructs, as therapy for cancer. A considerable hurdle for this has been the delivery of such therapies. Despite the challenges, there are now two clinical trials for microRNA-based therapeutics in cancer among 55 open microRNA clinical trials (ClinicalTrials.gov, accessed September 2014). The most advanced trial involves use of anti-miR-122 (Miravirsen) for hepatitis C therapy (Janssen et al., 2013) which shows reduction in viral RNA with no evidence of resistance. Miravirsen is complementary in sequence to miR-122 but also has a modified LNA structure providing resistance to degradation yet high affinity for its target. The detection of this apparent liver-specific microRNA may become necessary for patient eligibility for Miravirsen in both hepatitis C and other liver disease (Qiu & Dai, 2014). More recent studies have shown that, although the intended target of Miravirsen is mature miR-122, it also has affinity for pri- and pre-miR-122 and this binding results in reduced processing of the miR-122 precursor molecules which enhances its treatment effect (Gebert et al., 2014).

The first microRNA-based therapy in cancer is MRX34: a synthetic miR-34a mimic loaded in liposomal nanoparticles (Bouchie, 2013). Replacement of this tumour suppressor microRNA antagonises essential cancer cell processes such as self-renewal, migratory potential and chemoresistance (Bader, 2012). This therapy is in phase I clinical trial for primary liver cancer and liver metastases from other cancers and should complete by the end of 2015. The delivery is such that the nanoparticles accumulate in liver: the target organ. Quantification

of MRX34 in non-human primates has established a good half-life and exposure in whole blood (Kelnar et al., 2014) and the results to date are encouraging. Systemically delivered miR-34a in preformed lung tumours in mouse models has also been shown to be effective and well tolerated (Trang et al., 2011). MicroRNAs may have a potential use in reducing the drug resistance of tumours as has been shown by the early studies using miR-9* replacement therapy, which reduces levels of SOX2, subsequently reducing levels of ABC transporters in glioblastoma (Jeon et al., 2011). High levels of SOX2 were present in patients less responsive to BCNU in this study and may represent a subgroup of patients who would benefit from this type of therapy.

The advent of microRNA-based treatments may suggest that microRNA detection will be a fundamental part of a clinical laboratory pipeline in the future. Detection of particular microRNAs may be required for the initial eligibility of a drug, patient monitoring and determination of relapse. As with many targeted therapies, resistance is often a result of long-term administration of these drugs (Chong & Janne, 2013) and mutation detection in the sequence of the target microRNA may require monitoring. Combinatorial microRNA-based therapies may ensue in an attempt to reduce tumour resistance. With similar effect, certain anti-miR therapies have the potential to target whole families of microRNAs, reducing the likelihood of resistance (Obad et al., 2011). The study of microRNA-based therapies is still in its infancy, and side effects of these therapies need to be evaluated. MicroRNAs have been shown to be exported from cells in exosomes (Manterola et al., 2014) and therefore they have the potential to become systemic; effects from this may only be apparent in clinical trials. Also, the processing of other microRNAs is likely dampened by overloading the microRNA processing machinery with replacement microRNAs, and the effects of this are uncertain (Choudhury et al, 2012b).

1.9.6. Assessment of microRNA alterations in a clinical laboratory.

Sample Preparation and Processing

For accurate measurement of microRNAs in patient samples, fresh or snap-frozen tissues should ideally be used. Nonetheless, many groups have successfully profiled microRNAs and classified tumours using archived FFPE material (Hu et al., 2010; Niyazi et al., 2011; Lu et al., 2012) some of which

were up to ten years old (Li et al., 2007). This is of value for diagnostic testing since frequently the only clinical specimen available is in the form of a fixed paraffin block. In this respect, microRNAs are considered more optimal than mRNA as they are less prone to degradation during the fixation process (Hall et al., 2012).

The proportion of microRNA in a sample is approximately 0.01% of total RNA, but varies widely (Liang et al., 2007). Isolation of microRNAs can be performed using phenol-based RNA extraction methods, and purified further using commercially available columns optimised to increase microRNA yield. A comparison of three extraction methods (phenol/guanidinium (TRIzol, Invitrogen) followed by isopropanol precipitation, miRNeasy (QIAGEN) and mirVana (Applied Biosystems) column-based kits) showed that while each method produced high quality purified RNA, a selective method-dependent loss of specific microRNAs occurs (Ach et al., 2008). TRIzol extraction led to lower levels of miR-29b, miR-33, and miR-219, and mirVana preparations showed consistently increased levels of miR-149, miR-328, miR-574, and miR-766 compared with other methods in the breast cancer and HeLa cells examined. Although microRNA extraction is generally straightforward, these studies emphasise the importance of using a consistent method and similar concentration of input RNA in control and test samples, and also the possibility of method-dependent pitfalls with certain microRNAs.

Extraction of microRNAs from biological fluids is similar to that from tissues. The major challenge is obtaining a sufficient amount of microRNA for reliable quantification. MicroRNA extraction protocols are not optimised for serum extraction and Li and Kowdley noted that using the QIAGEN miRNeasy kit, ratio of QIAzol to serum should be altered to greater than 7:1 however this amount cannot be accommodated in a standard 1.5ml eppendorf tube, which may decrease RNA yield and increase transfer steps (Li & Kowdley, 2012). The adoption of circulating microRNA assays may require the availability of custom equipment for use in a diagnostic laboratory, which may include robotics, to extract from multiple samples at one time. These novel protocols are rapidly developing.

In blood, circulating microRNAs are often in complexes with proteins including argonaute RISC catalytic component 2 (AGO2), which protect them from

degradation (Wang et al., 2010; Vickers et al., 2011). In addition to microRNA/protein complexes, some microRNAs are protected in cell-derived vesicles, including exosomes and microvesicles which can be relatively easily isolated from plasma (Skog et al., 2008). However, circulating microRNAs more frequently co-localise with AGO2 than vesicles (Arroyo et al., 2011; Fabbri et al., 2012) thus the choice of an initial blood fractionation method is of importance.

MicroRNA detection and quantification methods

Quantification of microRNA levels can be performed using quantitative real-time PCR (qRT-PCR), NGS (next generation sequencing) or hybridisation-based methods such as microarrays and bead-based technologies (Table 1.1). These techniques are not without complications, but are relatively straightforward. The comparative stability of microRNAs in biological samples, and the robustness of microRNA expression is an advantage for clinical testing, however, short sequence length and the similarity of related microRNAs have required some modifications to approaches initially established for mRNA detection. Until fairly recently, microRNA analysis has been performed using qRT-PCR and microarray-based approaches. Now though, NGS is emerging as a cost-effective option.

The number of microRNAs under study determines the method chosen for quantification. Microarray and NGS are global microRNA profiling methods whereas qRT-PCR has mainly been used for fewer microRNAs. A number of inventive qRT-PCR protocols have been developed that can sensitively detect and accurately quantify specific microRNAs (Andreasen et al., 2010). These methods are technically similar to those involving mRNAs, but use a microRNA-specific stem loop oligonucleotide primer for reverse transcription to extend the mature microRNA prior to qRT-PCR. This technique may be adapted for global profiling using a system with spatial separation of the samples and primer mixes: a technique known as 'digital PCR'. The advantages of PCR-based methods for clinical testing are the quick turnaround time and low cost.

Technique	Amount of RNA required	Cost and labour	Benefits and limitations
Microarray	120ng	Moderate	<ul style="list-style-type: none"> • Only profiles mature sequences • No SNP or editing information
qRT-PCR	<500ng for an array	Low	<ul style="list-style-type: none"> • An array cannot cover the whole mirna complement • More beneficial for smaller numbers of microRNAs • Short turnaround time
NGS	~1ug	Moderate/high	<ul style="list-style-type: none"> • Can be expensive if not in a high throughput facility • Must be batched to be cost effective • Complex bioinformatics • Laborious preparation procedures

Table 1.1. Comparison of methods for microRNA quantification. Next generation sequencing (NGS), qRT-PCR (quantitative real-time PCR) and microarray technology are employed currently for the quantification of microRNA. Both microarray and NGS are used for the quantification of the whole complement of microRNAs. More useful for diagnostics is qRT-PCR, which can be quickly performed for the set of microRNAs identified in the signature. Results for qRT-PCR can be obtained in a 24-hour period, which is highly beneficial in cases where treatment decisions may be based on the outcome of the test. In all methods, the percentage of tumour present in the sample should be estimated and the integrity of the microRNA in the sample assessed prior to analysis.

Prior to NGS, microarrays were the most widely used method of global microRNA quantification and several commercial platforms are available (Wang et al., 2007; Wu et al., 2013). With rapid improvements in sequencing methodology NGS has become a method of choice for microRNA profiling (Quail et al., 2008) and provides quantitative analysis of both mature and precursor microRNAs as well as base pair resolution for SNP and mutation detection. NGS methods are already employed for DNA-based sequencing currently in diagnostic laboratories (Morgan et al., 2010; Hayes et al., 2013) and the use of robotics for library preparation improves efficiency of the pipeline. The majority of sequencing protocols for microRNAs have been generated for

the Illumina Genome Analyzer (Illumina Inc.) (Morin et al., 2010; Luo, 2012). A useful pipeline for microRNA sequencing, including bioinformatics, has been developed by Tuschl's group (Farazi et al., 2012; Hafner et al., 2012). NGS microRNA data correlates well with qRT-PCR (Pradervand et al., 2010), however discordance has been observed in a small proportion of microRNAs, probably due to sequence-specific method-dependent issues (Leshkowitz et al., 2013). Multiplexing using sample barcodes allows sufficient read depth for at least 16 samples to be sequenced in a single Illumina HiSeq lane and enable NGS to be a cost-effective microRNA analysis method. Improved indexing protocols, for example employing combinations of both forward and reverse barcodes in order to maximise the number of samples on one lane of the flow cell, has further reduced costs (Tu et al., 2012). If urgent results are required on a sample-by-sample basis the cost and turnaround time of sequencing is inappropriate for diagnostics in most facilities at present. Additionally, the requirement for storage of clinical data for the appropriate time period is also expensive, as with any NGS service, and may influence the platform of choice for a diagnostic laboratory or may suggest centralisation of testing is more appropriate.

For comprehensive analysis of NGS data, the bioinformatics is complex and is a developing area. The short length and sequence similarity of microRNAs can make alignment of sequence reads against the genome difficult. To combat this, many protocols align to a precursor microRNA library prior to aligning to the mature sequence (Auvinen et al., 2012). Multiple tools and protocols are available for analysis of microRNA sequencing data (Farazi et al., 2012; Li et al., 2012) and it is likely that these will be standardised in the near future to be suitable for a clinical pipeline.

To add to the more conventional tests so far described, a novel workflow which employs hybridisation techniques reminiscent of microarray, but with improved high-throughput capacity, is Nanostring® nCounter. This approach uses digital color-coded barcodes and single-molecule imaging for detection and counting of multiple transcripts in a single reaction tube. Nanostring nCounter has been reported to be a high sensitivity assay of easy manipulation with a detection rate of less than one copy per cell for over 800 regions, using as little as 100ng total RNA (Geiss et al., 2008) as well as being cost-effective and suitable for

FFPE samples (Waggott et al., 2012). It has good reproducibility when compared to other microRNA quantification methods (Kolbert et al., 2013) and is a suitable platform for measurement of microRNAs in the clinical setting.

Sample normalisation

Comparative expression analysis requires the normalisation of microRNA expression between samples. This may take the form of comparison to a 'spiked in' non-endogenous microRNA of known concentration, a stably expressed endogenous sequence across samples, or comparison to the total expression of all microRNAs in the sample. For endogenous normalisers, a stably-expressed sequence across the particular sample set should be identified and kits are available to assess this across a number of samples (Jacob et al., 2013). Transfer RNAs, ribosomal RNAs, small nucleolar RNAs and small nuclear RNAs have been employed as endogenous controls for microRNA qRT-PCR. Additionally, "housekeeping" microRNAs have been proposed as reference controls (Peltier & Latham, 2008; J. Hu et al., 2014) but are not appropriate if expression differences are caused by alterations in general microRNA processing. Reference choice for these tests will depend on the microRNAs to be measured and the samples to be tested. Suitable controls should be rigorously validated prior to their introduction into clinical microRNA testing.

Despite the various technical challenges described here, overall the methodology for microRNA measurements in tumour tissue is now well established and, with appropriate in-house knowledge, is readily translatable into the clinical testing arena. The choice of platform will depend upon the individual requirements of the test, equipment availability and local expertise of each laboratory.

1.9.7. Target prediction of microRNAs.

As previously mentioned, the seed sequence of a microRNA is usually from the 2nd nucleotide from the 5' end of the microRNA. These seed sequences bind to the 3' UTR of mRNAs. The binding site can have perfect or imperfect complementarity and the nature of this complementarity determines the fate of the mRNA. The binding sequence can be 6 to 8 nucleotides in length and it can

be present anywhere in the sequence of the 3' UTR (and sometimes in the coding region (Brummer & Hausser, 2014)). These factors complicate the prediction of microRNA targets so in addition to sequence matching, prediction tools often use other parameters to provide a score for how likely the target is. The best known target prediction site is Targetscan, which scores each predicted target sequence based on the following: 3' pairing contribution of the seed, proximity to residues pairing to microRNA nucleotides 13-16, the local AU-rich nucleotide composition, proximity to seed sequences for microRNAs that are co-expressed and therefore cooperate in function, positioning within the 3'UTR at least 15 nucleotides from the stop codon, and positioning away from the centre of long UTRs (Grimson et al., 2007; Garcia et al., 2011). In addition to these scores the database provides a P_{CT} score, which is the probability of preferentially conserved targeting. As selectively maintained seed sequences are more likely to have a relevant biological function, sites with a high score are more likely to be effective (Friedman et al., 2009).

Miranda is another database used for predicting target sites (Miranda et al., 2006). This uses position-specific rules and conservation to predict targets, in a similar way to Targetscan (John et al., 2004). Another prediction tool, Pictar, is a database with a slightly different method of predicting target sites, based on the fact that microRNAs are co-expressed in different cell types to exert their effects in a coordinated fashion (Krek et al., 2005). The target sites are first identified by position information, co-expression and alignment and then are filtered based on their optimal free energy when bound. Each predicted site is scored by a Hidden Markov Model (HMM) Maximum likelihood fit procedure, and the score for each individual site is combined into a total score which provides a ranked list of transcripts (Krek et al., 2005). A more recent target prediction algorithm is that of DIANA micro-T (Paraskevopoulou et al., 2013). This predicts target sites in 3' UTRs and coding sequences using machine learning on a positive and negative set of microRNA target sites, which were taken from PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation) data from Hafner *et al* (Hafner et al., 2010). In addition to these purely bioinformatics prediction tools, TarBase includes experimental data of target sites including what experiments were conducted to test the targeting of a microRNA to these sites (Sethupathy et al., 2006;

Papadopoulos et al., 2009). Experimental methods used for this may include overexpression or knockdown of the microRNA followed by quantification of the levels of the target mRNA and protein. Another method is the use of luciferase reporter assays, which involve the co-transfection of a plasmid containing the 3' UTR of the target upstream of a luciferase reporter and the microRNA mimic or inhibitor (Jin et al., 2013). More high throughput methods use biotin probes to pull down targets using biotinylated microRNA mimics or immunoprecipitation with AGO to pull out the cross-linked RNA present within the complexes (Hafner et al., 2010; Subramanian et al., 2015).

1.9.8. MicroRNAs in cancer; functions, alterations and mechanisms.

The reported global decrease in expression of microRNAs compared to adjacent normal tissue in many cancer types suggests that altered processing may play a role in the pathology of some cancers (Lu et al., 2005). A number of components are involved in the processing of microRNAs and mutations rendering these less effective will inevitably reduce the overall levels of mature microRNAs in the cell (Fig. 1.9).

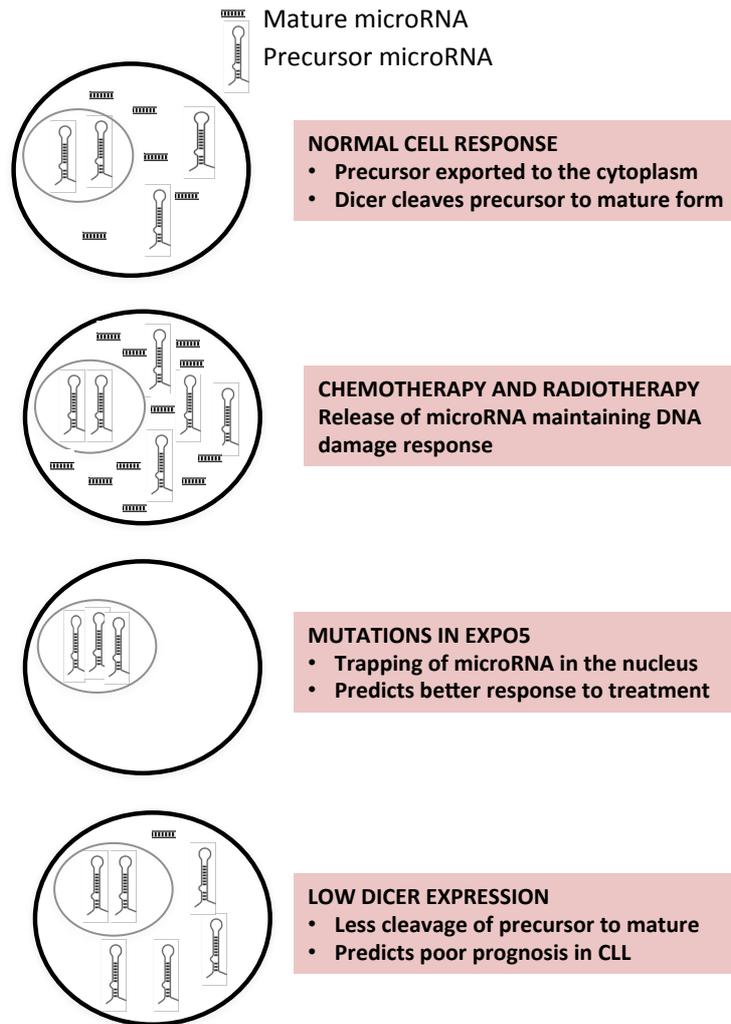


Figure 1.9. Alterations in the microRNA processing machinery in cancer.

The normal processing of a microRNA requires transcription of a primary transcript, cleavage, exportation to the cytoplasm, and further cleavage to generate a mature transcript. If any of the machinery performing these steps is altered in cancer there is global dys-regulation of microRNAs in the cell. It has been shown that chemo- and radiotherapy cause an increase in microRNAs in the cell, which acts to maintain the DNA damage response (Wan et al., 2013). Mutations in the exportation machinery, such as exportin-5, lead to a build up of precursor microRNA in the nucleus and a lack of mature microRNAs in the cytoplasm (Melo et al., 2010), where they usually exert their function. Low expression of processing components, such as Dicer, can dramatically reduce the numbers of microRNAs cleaved from precursor to mature form, again reducing their downstream effects on the cell (Kuang et al., 2013).

The involvement of microRNA processing in the treatment of cancer is suggested by the apparent acceleration in nuclear export of pre-microRNA following radiation- or chemotherapeutic- induced DNA damage (Wan et al., 2013). Hence, the resulting increase in microRNA processing may maintain the DNA damage response in the cell. Specific aberrations of the microRNA processing machinery have been directly associated with various cancers. For example, in some cancers with microsatellite instability, mutations in Exportin-5 (XPO5), which exports pre-microRNA from the nucleus to the cytoplasm (Yi, 2003), lead to trapping of pre-microRNAs in the nucleus preventing further microRNA processing and function (Melo et al., 2010). A SNP in XPO5 has been associated with lung cancer and multiple myeloma, and tumours with this SNP have a better response to chemotherapy (de Larrea et al., 2012; Ding et al., 2013). Low expression of Dicer, the RNase that cleaves precursor microRNAs into their mature form, predicts poorer outcome in CLL and ovarian cancer (Merritt et al., 2008). Furthermore, Dicer expression is lower in CLL patients with unfavourable cytogenetic aberrations (Zhu et al., 2012a) which is consistent with the general assumption that global reduction of mature microRNAs is associated with poorer outcome (Lu et al., 2005). To add to this, recognition of primary microRNA hairpins among a background of other hairpin RNAs is required in order to target them for cleavage and ultimate export. Bartel's group identified sequence determinants within the primary transcripts that license the hairpin for processing by allowing binding of certain proteins, such as SRp20 (Auyeung et al., 2013). Mutations of these regions may also have an implication in cancer, although there have been no reports of processor binding site alterations in cancer to date.

MicroRNA editing is another layer of regulation that is altered in some cancer types. This process is catalysed by 'adenosine deaminases that act on RNA' (ADARs) which convert adenine to inosine (Paul & Bass, 1998). Inosine acts like guanosine and base-pairs with cytidine therefore inducing changes in target recognition (Habig et al., 2007). Editing of miR-376* (the passenger strand of miR-376) has been reported in high-grade glioma and may effect patient outcome (Choudhury et al, 2012a; Seton-Rogers, 2012). Increasing use of NGS and the efforts of TCGA in providing large volumes of data will likely highlight more examples such as this in the future.

1.9.9. MicroRNA networks in cancer.

The presence of ceRNAs, and multiple sequences in the cell at any given time that may “compete” for microRNA binding, emphasises the extensive networks that are involved in microRNA function. The imperfect match and relatively short 6-8 base pair “seed” sequence characteristic of microRNA-mRNA interactions allow for a multitude of potential targets for each microRNA. Additionally, a single mRNA may have target sites for multiple microRNAs, creating redundant molecular networks for the control of gene expression. Because of the potential to predict microRNA-binding sites based on base pairing, they are highly amenable to systems biology approaches. However, many studies have focused narrowly on the specific effect of a given microRNA upon a specific mRNA, defined by bioinformatic prediction algorithms, rather than exploring the extended network of gene expression (Sumazin et al., 2011). One of the reasons for the focus on single bioinformatically predicted targets is experimental tractability, as it is not trivial to identify microRNA targets experimentally in mammalian cells, and many important interactions have been identified using this method. However, the microRNA/single mRNA target approach is limited and may not accurately reflect the most physiologically significant microRNA/target interactions. This can be overcome by screening multiple targets, or using global approaches to identify microRNA/target interactions inside the cell. These include proteomics, gene expression arrays and RNA cross-linking/Ago2 pull-down approaches such as HITS-CLIP, to allow assessment of microRNA-target binding in the cell (Chi et al., 2009; Boudreau et al., 2014; Hu et al., 2013). Future studies in the field would benefit from the application of these techniques as well as assessment of microRNA functions in the context of networks, including sponge interactions and feedback loops which take into account the competitive nature of interactions between microRNAs and their targets.

The influence of groups of microRNAs is exemplified by microRNAs in clusters, which are expressed together and show functional cooperation. For example, the polycistronic oncogenic miR-17~92 cluster of microRNAs specifically induce lymphomagenesis in a B-cell-specific transgenic mouse model (Sandhu et al., 2013) and miR-19b, miR-20a and miR-92 from this cluster along with miR-26a

and miR-223, promote T-cell acute lymphocytic leukaemia development in mouse models (Mavrakis et al., 2011).

Inter-cellular network interactions should also be addressed in the context of microRNAs as shown by the symmetry of cell division in colorectal cancer (CRC). In late-stage CRC and CRC stem cells, divisions are symmetric producing two self-renewing daughter cells. In early-stage CRC, cell fate determinants are localised to opposite poles during division resulting in one self-renewing and one differentiating daughter cell (asymmetric cell division). This has been shown to be controlled by a Snail/mir-146a/Numb/ β -catenin axis (Lerner & Petritsch, 2014). This proposes that miR-146a, induced through Snail dependent β -catenin and TCF (T-cell factor), down-regulates Numb, relieving Numb-mediated degradation of β -catenin and subsequently enhancing WNT (wingless-type MMTV integration site) signalling (Hwang et al., 2014). This effect maintains self-renewing divisions, partly independent of the EMT.

As part of their role in shaping the fate of a cell, microRNAs are fundamental in the control of EMT. Some microRNAs, such as the miR-200 family and miR-34a, are protectors of the epithelial phenotype, and their down-regulation during EMT enhances mesenchymal specifying targets such as ZEB1 and ZEB2 (Hao et al., 2014). The miR-34a family can also be inhibited by ZEB1 (Kim et al., 2011a), establishing a robust feedback loop to ensure the cell is driven towards a more mesenchymal fate. Oncogenic microRNA miR-22 has been shown to inhibit anti-metastatic miR-200 in breast cancer by targeting the TET family of methylcytosine dioxygenases, which results in silencing of miR-200 (Song et al., 2013b). Positive correlation of miR-138 and EMT has uncovered its role in driving the process through multiple targets including Vimentin, transcriptional repressors such as ZEB2 and epigenetic regulators such as EZH2 (Liu et al., 2011b). Similarly, miR-155 has been shown to repress TGF- β -induced EMT and depletion of this microRNA can suppress EMT in a mouse model (Kong et al., 2008).

The study of microRNAs under stress conditions has uncovered some important findings including the EGFR-mediated phosphorylation of AGO2 in response to hypoxia in breast cancer, resulting in suppression of specific microRNAs that depend on AGO2 for their maturation (Shen et al., 2014). This is of huge importance in cancers where EGFR is constitutively active. Understanding of

microRNA regulatory networks has underlined the importance of microRNA control over tumour cell biology, and has highlighted novel therapeutic targets and processes involved in tumour growth. One such example of this is the recent discovery that miR-542-3p weakens the interaction of p53 with its negative regulator MDM2, thus stabilising the protein (Wang et al., 2014c).

1.9.10. MicroRNAs as serum biomarkers.

The use of circulating microRNAs as markers in different cancer types is a rapidly developing area (Schwarzenbach et al., 2014). Tumour cells can release microRNAs, stabilised by their incorporation into microvesicles, which have shown stability in the circulation following multiple freeze-thaw cycles and prolonged exposure to room temperature (Mitchell et al., 2008). A study of 391 patients with non-small cell lung cancer (NSCLC) identified 35 highly expressed microRNAs with predicted binding sites for at least one of 11 genes of the TGF- β pathway, which were significantly differentially expressed at the extremes of survival. Of these, 17 were associated with patient survival and were combined into a risk score significantly predicting survival in advanced NSCLC (Wang et al., 2013). Also, isolation of exosomes from serum showed a signature involving two microRNAs and one small non-coding RNA can be used for non-invasive diagnosis of glioblastoma (Manterola et al., 2014).

The detection of microRNAs in the blood presents some challenges and there is overwhelming discordance between reports in well-studied cancers (Jarry et al., 2014). Appropriate endogenous controls for microRNA quantification in serum are under debate as many mRNA and rRNA species are absent in blood due to circulating RNases (Li & Kowdley, 2012). Clinically, fluctuations of circulating microRNAs can occur as a result of treatment, diet and other factors, increasing noise in these assays. The presence of myeloid and lymphoid cells can alter the levels of certain microRNAs and viral infections of the patient may also effect endogenous microRNA expression (Pritchard et al., 2012). Expression changes of microRNAs are rapid in blood and even a traumatic venepuncture may have the potential to influence this. Despite these hurdles, it is clear that further study is warranted for detection of the presence of microRNAs in the blood for future non-invasive biomarker development and the field is moving rapidly towards that goal.

1.9.11. MicroRNA polymorphisms predisposing cancer.

Aside from treatment resistance, it is worth noting that a number of SNPs in microRNA binding sites are involved in cancer risk, and may be markers for genetic susceptibility studies in some cancers (Ziebarth et al., 2012). They can be used as markers to predict subsets of patients at risk of poor outcome or lack of treatment response. These SNPs may be present in microRNA target sites, in the processing machinery or of the microRNA sequence themselves, altering their targets and ability to be processed (Ding et al., 2013; Wu et al., 2011; Chin et al., 2008).

1.9.12. MicroRNAs in glioblastoma.

MicroRNAs have oncogenic, tumour-suppressive, ECM (extracellular matrix) - responsive and treatment-related roles in glioblastoma (Godlewski et al., 2010b; Yin et al., 2012; Munoz et al., 2014; Kim et al., 2014). Their expression levels can be exploited to classify glioblastoma according to WHO classifications and by prognosis (Srinivasan et al., 2011; Sana et al., 2014; Rivera-Diaz et al., 2015). New molecular classifications have also been generated by clustering microRNA expression and have been shown to delineate glioblastoma into 5 groups with expression patterns reminiscent of different neural cell precursors (Kim et al., 2011b).

One of the most studied microRNAs in cancer is miR-21, and this microRNA has been shown to have an important role in regulating apoptosis in glioblastoma (Quintavalle et al., 2012a). MicroRNA-21 is highly expressed in glioblastoma, and reduces apoptosis in cells through targeting of p53 and FASLG (Papagiannakopoulos et al., 2008; Shang et al., 2015). Through this role, its presence in glioblastoma cells confers drug resistance and therefore oligonucleotide inhibitors of miR-21 have been explored to reduce the effect (Giunti et al., 2015). Additionally, miR-21 has been shown to increase glioblastoma migration by repressing matrix metalloprotease inhibitors (Gabriely et al., 2008).

MicroRNAs have a prominent role in EGFR signalling, and this has been studied extensively in glioblastoma (Wang et al., 2011a; Serna et al., 2014; Wang et al., 2014b). A tumour suppressor microRNA, miR-34a, is decreased in glioblastoma, and this, in combination with amplified EGFR confers a poor

outcome in glioblastoma patients (Yin et al., 2013). When miR-34a was overexpressed in glioblastoma cells, EGFR protein expression decreased and this was shown to be mediated through Yin Yang-1 (YY1) (Yin et al., 2013). Additionally, miR-9 is suppressed by EGFRvIII through the Ras/PI3K/AKT pathway in glioblastoma conferring a growth advantage (Gomez et al., 2014). The oncomiR miR-148a has also been shown to have a role in regulating EGFR and apoptosis in glioblastoma through its targeting of MIG6 and BIM (Kim et al., 2014).

As previously mentioned, microRNAs have prominent roles in development, and their levels fluctuate throughout differentiation (Letzen et al., 2010). They have been shown to have a strong regulation influence over the Notch and WNT pathways in glioblastoma (Chen et al., 2012b; Zoni et al., 2014; Liu et al., 2014). Glioblastoma stem cells have been shown to have a dependence on the TGF- β signalling pathways and NF- κ B, and these are also regulated through microRNA mediated mechanisms involving miR-9, miR-34a and miR10b (Bazzoni et al., 2009; Genovese et al., 2012; Lin et al., 2012). Stem cell function is maintained by the polycomb repressor complexes and components of these complexes, for example BMI-1 (B Lymphoma Mo-MLV Insertion Region 1) and EZH2, which are regulated by miR-128 and miR-101 respectively (Smits et al., 2010, Peruzzi et al., 2013;).

The effects of expression of a microRNA are relatively fast, and these expression levels can be altered very quickly. Some microRNAs have a half-life of just 30 minutes, such as miR-9 (Sethi & Lukiw, 2009). This makes them excellent candidates for mediators of signalling pathways in response to environmental stimulators. In a disease such as glioblastoma, which has extreme environments; microRNAs have been shown to regulate cellular response. For example, low glucose availability reduces the expression of miR-451, which suppresses proliferation and stimulates migration through CAB39 (Calcium Binding Protein 39) (Godlewski et al., 2010a). This contributes to the 'go or grow' behaviour of glioma cells; where cells upregulate one or the other in response to external stimulatory factors (Godlewski et al., 2010a). Similarly, microRNAs have been shown to stimulate cellular response in the presence of hypoxia. A well known hypoxia responsive microRNA is miR-210, and knockdown of this in hypoxic glioblastoma stem cells caused cell cycle arrest,

decreased neurosphere formation and stem cell marker expression (Yang et al., 2014). This microRNA may therefore be a potential candidate for targeting GSCs in the hypoxic niche of glioblastomas. Another mechanism demonstrating hypoxic response is the blocking of the effects of tumour suppressive microRNA miR-297 when cells are grown in hypoxia, conferring survival advantage to glioblastoma cells in this environment (Kefas et al., 2013). On a global level, and previously mentioned but notable in this context, it has also been shown that EGFR has the ability to modulate the maturation of a microRNA in hypoxia by its phosphorylation of AGO, a component of the RISC complex which serves to reduce the overall levels of mature microRNAs in the cell (Shen et al., 2014). MicroRNAs have also been explored as potential therapeutics for glioblastoma. This is a complicated field because RNA-based therapies require delivery strategies to provide adequate dosage to the tumour site. In glioblastoma this is difficult due to the BBB. It has been shown that mesenchymal stem cells can deliver miR-145 and miR-124 to glioma cells when administered intracranially and systemic administration of miR-7 using novel integrin-targeted biodegradable polymeric nanoparticles in mice with human glioblastoma xenografts has also been successful in inhibiting angiogenesis and growth (Lee et al., 2013; Babae et al., 2014). MicroRNA screens have become commonplace for identifying microRNAs associated with a particular function, and one such screen has shown that four microRNAs; miR-1, miR-125a, miR-150 and miR-425, induce resistance to radiotherapy. These all showed correlation with TGF- β expression and miR-1 and miR-125 were shown to be regulated by this pathway directly (Moskwa et al., 2014). Furthermore, miR-125b has been shown to provide resistance to TMZ through PIAS3 (Protein Inhibitor Of Activated STAT, 3), BAK1 (BCL2-Antagonist/Killer 1), TNFAIP3 (Tumour Necrosis Factor, Alpha-Induced Protein 3) and NKIRAS2 (NF- κ B Inhibitor Interacting Ras-Like 2) in glioblastoma (Shi et al., 2012; Shi et al., 2014a; Chen et al., 2014; Haemmig et al., 2014). Three of these targets are implicated in NF- κ B signalling, which is required for sensitivity to TMZ (Liu et al., 2012; Wagner et al., 2013; Haemmig et al., 2014). Similarly, miR-9, which directly targets NF- κ B 1, has TMZ-resistance properties (Bazzoni et al., 2009; Lee et al., 2013).

The most advanced translational study for microRNAs in glioblastoma is an evaluation of miR-10b in gliomas as a whole (Clinical trials identifier NCT01849952). This microRNA has previously been shown to be higher in expression in glioblastomas compared to normal brain tissue, and is involved in cell migration and invasion (Guessous et al., 2013). The trial is currently recruiting and the investigation includes the *in vitro* assessment of primary tumours to anti-miR-10b treatment.

1.10. Survival analysis.

Survival analysis is a type of statistical investigation that determines the probability of a patient surviving up to a certain point in time. This is measured using the hazard function which computes the number of patients who died during a time interval divided by the number of patients alive at the beginning of the interval weighted by the length of the time interval. Time to the event can be measured in days, weeks or years, whichever is most appropriate for the time to event data. In many cases time to death data is incomplete, as some patients may still be alive after the study ends, and other patients may have dropped out of the study. In these cases the observations are censored, to represent the missing data. This includes right censoring, where the individual did not encounter the event during the study period, left censoring (which is not appropriate for overall survival data) where the patient joined the study after the event occurred, and interval censoring, when the event occurred at some time during the study in between two measurements. The most common type of censoring for survival analysis in time to death data is right censoring (Altman & Bland, 1998).

Models used to analyse the survival time of a set of predictor variables include both parametric and non-parametric methodologies (Cantor & Shuster, 1992). Parametric methodology assumes that the survival data has an underlying distribution similar to the known probability distributions, which include exponential, Weibull and lognormal distributions. These models make inferences about the parameters of the distribution and they assume a homogeneous variance of the data. Non-parametric approaches have no assumed distribution and they include the Kaplan-Meier and log-rank test. The Kaplan-Meier method estimates survival probability as a function of time, and is

often interpreted in the form of a graph. Where data are censored this can be represented on the graph as a vertical line at that timepoint, indicating further information from this patient has not been collected. The Kaplan-Meier method is a maximum likelihood estimate of the probability that a person will have a lifetime that exceeds the time of the event and takes the following form for the event time t_i and the number of deaths d_i :

$$\hat{S}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

This is therefore the product of the fraction of survivors at time t_i . If there is no censoring, then n_i is simply the number of survivors just before time t_i . If there is right censoring, n_i is the number of survivors minus the number of losses (Cantor & Shuster, 1992). This means that at the time of the event it is only the number patients still alive that are at risk of death. This is calculated at each time point that an event occurs.

To test the survival distributions of two or more groups, the log-rank test can be used which compares the Kaplan-Meier curve for each group. This tests the null hypothesis that there are no differences between two or more populations if there is an event (i.e. a death) at any time point. Log-rank calculates probabilities as in the Kaplan-Meier method, and then determines the total numbers of expected deaths and the number of observed deaths in each group. From this, a χ^2 test can be performed to test the null hypothesis, which is $(\text{observed-expected})^2/\text{expected}$, producing a significance result (Altman & Bland, 1998).

In certain cases it may be necessary to determine what multiplicative effect an increase in a covariate has on the hazard rate. The Cox proportional hazards model is a semi-parametric model designed to test this (Cox, 1972). This model assesses the proportional change a covariate has on the baseline survival curve. The baseline hazard curve is the curve when all the coefficients are equal to zero. The hazard ratio is the hazard divided by the baseline hazard. This is also the exponent of the coefficient. For example a hazard ratio of two represents a population that has twice the risk of dying compared to the other,

or with a continuous variable represents a 2x increase in risk per unit of the variable.

In some cases it may be necessary to assess a large number of predictors, to determine which may be prognostic. LASSO (least absolute shrinkage and selection operator) has the ability to select variables, based on their prognostic ability, taking into consideration other variables in the model (Tibshirani, 1996). This estimator involves penalisation of the regression coefficients, which results in a number of covariates with coefficients shrunk to zero. LASSO minimises the following (where y is a dependent variable, x is an independent variable, β is an unknown parameter and λ is a tuning parameter controlling the amount of shrinkage):

$$\frac{1}{2} \sum_i (y_i - \mathbf{x}_i^T \boldsymbol{\beta})^2 + \lambda \sum_{j=1}^p |\beta_j|,$$

This estimator uses regular least squares (the part of the equation to the left of the + sign), with a penalty that is determined by λ (the second part of the equation to the right of the + sign). Least squares works by starting with all coefficients at zero, then finding the predictor most correlated with Y and increasing the coefficient of the predictor in the direction of the correlation. This happens until another predictor has an equal correlation, then the coefficient for that predictor is increased in the direction of the least squares. This is continued until all the predictors are included in the model.

The absolute values of the coefficients are used in the penalty function, which results in some coefficients being shrunk all the way to zero. This is in contrast to ridge regression, which penalises on the squares of the values, and therefore all covariates have a non-zero coefficient (Fig. 1.10). This therefore performs a model selection and predictors do not have to be chosen as they would from ridge regression.

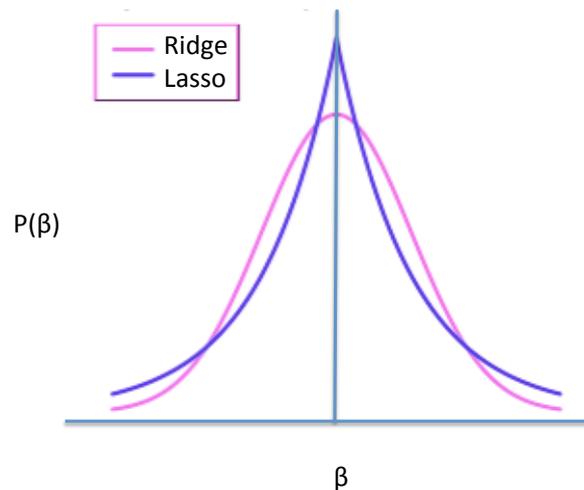


Figure 1.10. The probability distribution of a parameter, β , when penalised using ridge and LASSO.

Using Bayesian probability, penalisation performed by ridge, which uses the squares of the coefficients in the penalty function, produces a smooth probability distribution, which never reaches zero. This means that as the penalty increases, all of the coefficients shrink, but remain non-zero.

Penalisation using LASSO however, uses absolute values for the coefficients in the penalty function, and produces a 'pointy' probability distribution, which results in some coefficients being shrunk to exactly 0. This has the effect that, as the penalty increases, more coefficients are shrunk to exactly zero.

The tuning parameter, λ , which is required for the penalisation is obviously important and the optimal λ can be determined using likelihood cross validation. Cross validation involves separation of the data and generating and validating the model on different subsets, to avoid overfitting of the model. This is performed multiple times with different λ values over many different subsets and the results are averaged over the rounds to reduce variability. The λ with the lowest cross-validated likelihood is then used in the LASSO estimator (Goeman, 2010).

1.11. Publically available data.

The Cancer Genome Atlas (TCGA) is a large program to integrate comprehensive data for multiple cancer types, including gene expression, methylation, copy number and microRNA expression for public use. The first disease to be studied by the TCGA was glioblastoma (Cancer Genome Atlas Research Network, 2008). Glioblastoma data includes over 500 surgically resected pathologically identified glioblastoma samples with information on copy number, methylation, gene expression and microRNA expression. The data was initially based on microarray microRNA and mRNA expression measurements but now sequencing data is also available including RNA-seq and exome-seq. This dataset was extensively analysed in 2013 for glioblastoma (Brennan et al., 2013). In addition to the glioblastoma data, there are also over 500 samples for lower grade glioma (grade II and grade III) with exome, SNP methylation, mRNA and microRNA data available, although data is not complete for every platform. Comprehensive clinical data and sample handling is also provided which makes these datasets extremely powerful, not only for generating hypotheses but also validating wet-lab findings in patient samples.

1.12. The aims and objectives of this thesis.

In many cancers, personalised medicine, or the 'right drug for the right patient, at the right time', has shaped patient management over the past five to ten years. In glioblastoma, although molecular classifications have been created, they have never reached the clinic and the only markers used for patient treatment are *MGMT* promoter methylation and *IDH* mutation. The primary aim of this thesis is to improve this situation, providing molecular biomarkers using microRNAs, which are highly appropriate for clinical assessment.

My background is as a Clinical Scientist in the National Health Service and I have seen how service has changed with the era of personalised medicine and learned how best to implement these tests for patient management. The microRNA signatures generated here have been designed in such a way as to allow easy translation to a clinical laboratory.

MicroRNA signatures for prognosis prediction in glioblastoma have previously been generated (Lakomy et al., 2011; Niyazi et al., 2011; Srinivasan et al.,

2011; Zhang et al., 2012a, Sana et al., 2014) but the microRNAs used are different between signatures. In order to confirm the role of the microRNAs used in the signatures generated here, every effort has been made to validate the role of these microRNAs in glioma biology.

In summary, the work in this thesis can be described as:

1. Generation of a signature that predicts prognosis in glioblastoma and prediction of the pathways targeted by the microRNAs included in the signature.
2. Assessment of all prognostic microRNAs in malignant glioma and prediction and validation of the pathways targeted by these microRNAs.
3. Analysis of a prognostic microRNA in glioma, miR-9, and an attempt to resolve why a phenotypically oncogenic microRNA appears to be expressed at higher levels in patients with a better outcome.
4. Generation of a signature for prediction of response to the anti-angiogenic drug bevacizumab and assessment of the signature using cell line drug sensitivity data.

This work attempts to validate the hypotheses generated using the huge volumes of data in the TCGA using further patient subsets and experimental data.

2. A 9-microRNA signature predicts prognosis in glioblastoma

'Science predicts that many different kinds of universe will be spontaneously created out of nothing. It is a matter of chance that we are in.' **Stephen Hawking, May 2011.**

2.1. Introduction.

In many cancers, a wealth of molecular markers are available for use in diagnosis and patient management (Marchio et al., 2015; Reguart & Remon, 2015). Unfortunately, the development efforts in personalised medicine for glioblastoma have been less successful, and there are few molecular indicators and companion diagnostics for patient management. As previously mentioned, the prognosis of patients with glioblastoma can range from a few weeks to a number of years (Bleeker et al., 2012) and the best marker to predict this is *MGMT* promoter methylation (Fig. 2.1), (Hegi et al., 2005). However, this marker is only of use in patients treated with temozolomide (Riemenschneider et al., 2010). The *IDH* mutation also predicts a good prognosis, although the majority of these patients have progressed from lower grade disease which would already be an indicator of good prognosis (Riemenschneider et al., 2010). Determination of patients with particularly aggressive tumours is paramount to patient management to allow more extensive post-treatment surveillance and possible targeting for more aggressive drugs. Therefore, appropriate molecular markers predicting prognosis in glioblastoma is an unmet need.

MicroRNAs represent an attractive tool for stratifying patients into prognostic categories because of their stability in clinical samples and their ease of quantification compared to mRNAs (Hall et al., 2012). Directly associated markers, such as *EGFR* mutations predicting response to an EGFR inhibitor, are the most ideal companion diagnostics when assessing which treatment regime is best for a patient. In the absence of these however, microRNAs discovered using correlation analysis can be exploited to provide information on a patient's tumour that otherwise would not be available. As a result, there have been multiple attempts to create signatures using microRNAs in different cancers for a variety of applications (Calin & Croce, 2006; Yu et al., 2008;

Garzon et al., 2008; Gokhale et al., 2010; Visone et al., 2009; Díaz-Martín et al., 2014). To date though, there are no microRNA signatures in clinical use.

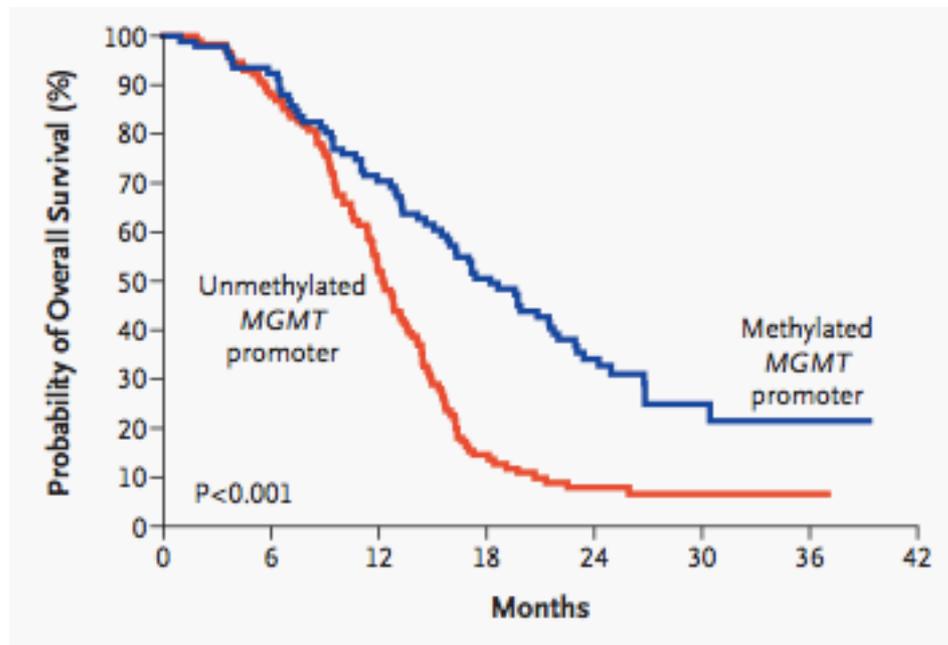


Figure 2.1. MGMT promoter methylation predicts a better prognosis in glioblastoma.

MGMT promoter methylation is currently the only marker that predicts response to treatment for the disease. Image from Hegi et al., 2005 (Hegi et al., 2005).

MicroRNA expression signatures can define tumour types and molecular subgroups (Kim et al., 2011b). It is also possible to cluster the expression of microRNAs into groups based on their embryonic origin (Lu et al., 2005). For this reason, microRNA expression can be used to identify cancers arising from different cell lineages. The discovery of microRNAs as predictors of prognosis became apparent as early as 2004, when let-7 was shown to predict post-operative survival in lung cancer (Takamizawa et al., 2004). Several individual microRNAs have been associated with glioblastoma prognosis (Mizoguchi et al., 2012), but it is likely that multiple microRNAs will provide a more statistically robust approach. Previous prognostic signatures for glioblastoma have been designed (Lakomy et al., 2011; Srinivasan et al., 2011; Zhang et al., 2012a; Sana et al., 2014), although the microRNAs employed are not consistent between studies (Fig. 2.2). Potential reasons for this, as already stated, are due

to the different sample sizes, population differences in microRNA expression, sample types and diverse extraction and quantification methods.

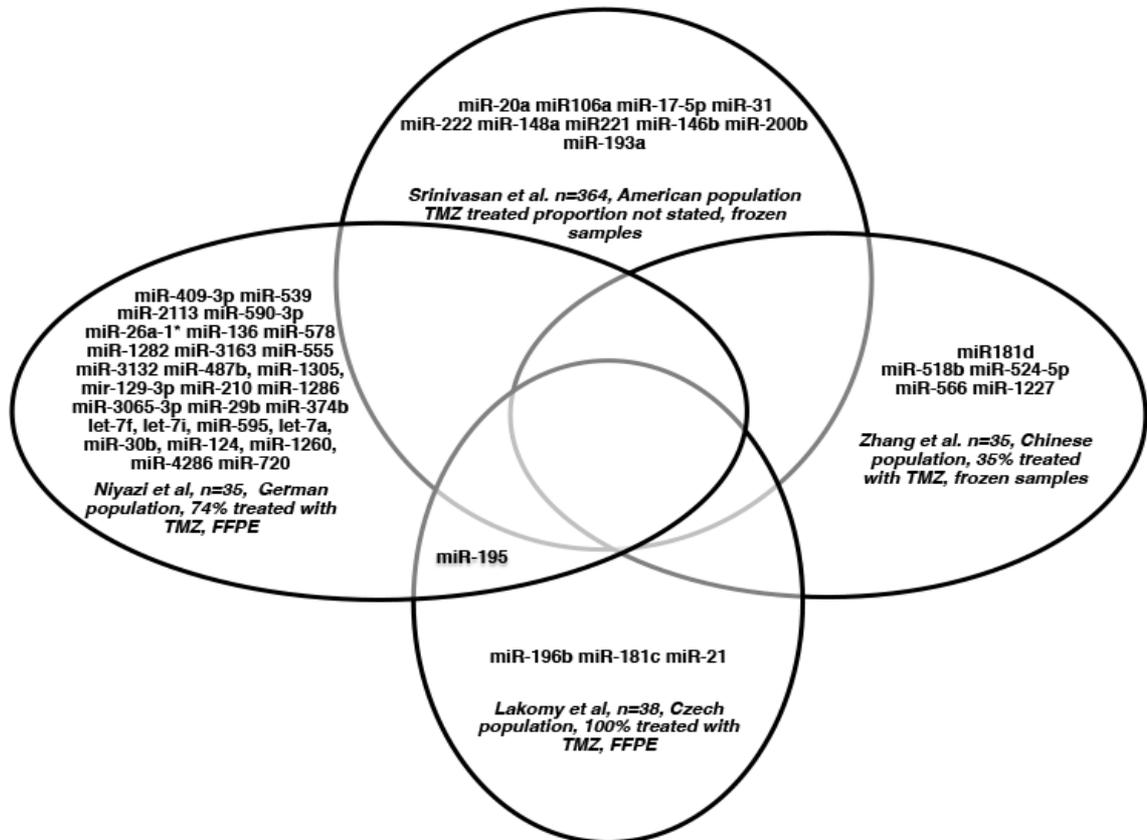


Figure 2.2. Four prognostic microRNA signatures developed for glioblastoma.

There is little concordance in the microRNAs used in the signatures. This may be due to a number of study design differences including sample type and size (Lakomy et al., 2011; Niyazi et al., 2011; Srinivasan et al., 2011; Zhang et al., 2012a).

In the study described here I used a novel methodology, known as LASSO (least absolute shrinkage and selection operator (Tibshirani, 1996)), with glioblastoma data from The Cancer Genome Atlas (TCGA, NIH), to identify a 9-microRNA prognostic signature. The 9 microRNAs were then used to generate a risk score algorithm suitable for clinical prognostic stratification. The signature separated patients according to outcome, was relevant in patients treated with temozolomide and was validated in an independent dataset. Although other microRNA prognostic signatures have been identified in glioblastoma, this is the first to use the whole TCGA dataset; it is relevant across molecular subtypes of

glioblastoma, and is the first to be validated in an independent dataset from a different geographical population. Moreover, the signature microRNAs have been previously implicated in glioblastoma biology, with known functional roles, further supporting the relevance of the signature. Thus I have identified a functionally relevant, robust microRNA-based prognostic signature in glioblastoma.

2.2. Methods.

2.2.1. TCGA clinical information and expression data.

Level 2 Agilent microRNA 8x15k microarray and G4520A microarray gene expression data plus clinical information for 475 glioblastoma and 10 unmatched non-tumour samples were downloaded from TCGA (TCGA, NIH) (accessed October 2012). I chose to use level 2 data, which provides quantile normalised expression data for each probe. This was to aid translation of the signature into a clinical setting, as an exact oligonucleotide sequence can be used as a prognostic tool, rather than the expression of a microRNA, which may have multiple isoforms. Only patients treated with radiotherapy and some form of chemotherapy were selected as the aim in creation of a prognostic signature is to benefit patient management, which most likely will include treatment in some form (Table 2.1). Illumina HiSeq sequencing data (level 3, reads per million of total reads mapping to a mature microRNA) for microRNAs were downloaded for all samples with grade II or III glioma from TCGA ($n=178$; 55 astrocytoma, 47 oligoastrocytoma, 75 oligodendroglioma, 1 not stated; 95 grade II, 112 grade III, 1 not stated).

Characteristic	Number of patients (n=475)
Age (median=59)	
<60 years	248
≥60 years	227
Gender	
Male	293
Female	182
Karnofsky Performance Status	
≤70	141
>70	220
Not available	114
Days to death/ last follow-up (median 430 days)	
<450 days	301
≥450 days	174
≤30 days	20
Therapy	
TMZ	3
TMZ and radiation	187
Other	285

Table 2.1. Characteristics of patients used in the generation of the signature.

The characteristics of the 475 patients included in the generation and testing of the model are shown in the table. There are more males in the study (62%), which is expected for a glioblastoma cohort. KPS (Karnofsky performance status) was calculated prior to surgery and at least 30% had what is considered to be a low KPS (<70). There were 26 IDH mutations recorded in this cohort although 117 did not have IDH mutation information.

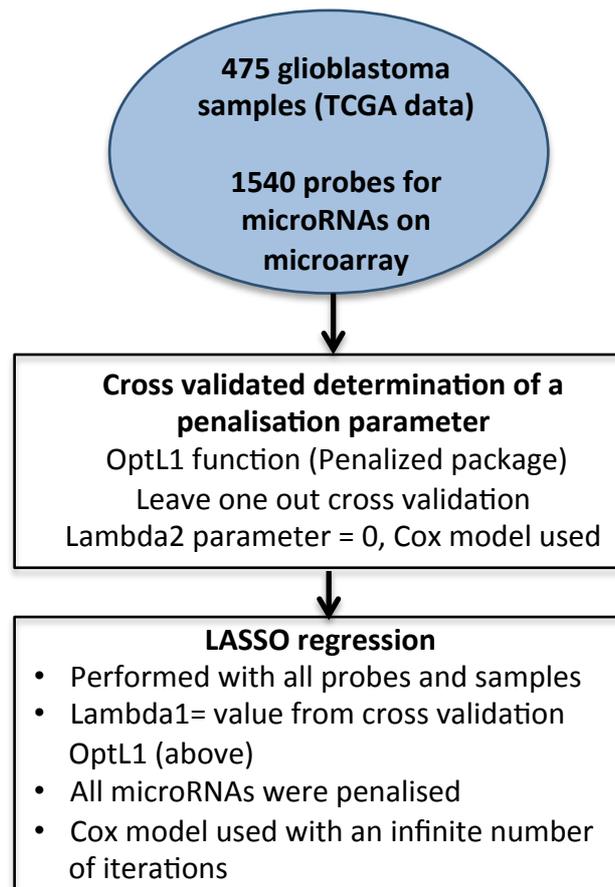


Figure 2.3. Workflow for generation of the prediction algorithm

The prediction algorithm was generated using the Penalized package in R (Goeman, 2010). Firstly an appropriate penalization parameter was determined using the OptL1 function and then LASSO regression was performed using the Penalized function.

2.2.2. Statistical analysis of microRNA expression data in glioblastoma.

Glioblastoma samples were assessed using a LASSO penalised regression analysis (see explanation on page 68) to predict survival using microRNA expression (Tibshirani, 1996) with leave-one-out cross-validation using R software (v2.15.1) and the Penalized package (Goeman, 2010) (Figure 2.3). A risk score was generated using the sum of microRNA expression values weighted by the coefficients from the LASSO regression, as described in Alencar *et al* (Alencar et al., 2011).

This was: $E_{\text{miR-n}}$ = expression of microRNA n.

$$\text{Risk score} = -0.044E_{\text{miR-370}} + 0.062E_{\text{miR-124a}} + -0.066E_{\text{miR-145}} + 0.005E_{\text{miR-34a}} + 0.015E_{\text{miR-10b}} + 0.092E_{\text{miR-148a}} + 0.162E_{\text{miR-222}} + -0.032E_{\text{miR-9}} + -0.021E_{\text{miR-182}}$$

The risk score was applied to all glioblastoma samples in the dataset and the samples separated into low- and high-risk groups using the median as a cut-off. The median was used as a cut-off here for demonstration, but the cut-off may be defined based on treatment availability if introduced as a clinical test. A Cox regression model incorporating age and the log-rank test were used to assess overall survival (OS) of the two groups in the whole dataset, the molecular subtypes of glioblastoma (defined using published classification information from Brennan *et al* (Brennan et al., 2013)) and temozolomide treated patients (obtained from TCGA clinical data). The score was also assessed as a predictor for progression-free survival (PFS). A statistical significance threshold of $p=0.05$ was used throughout. Pearson's correlation coefficient was calculated for correlation of age with risk score. Pearson's correlation was chosen (as opposed to Spearman's correlation) because it assesses for a linear relationship, which is what would be expected from the relationship between age and a factor such as risk score. Multivariable Cox regression models were used for the risk groups and each of the following factors (separately); *MGMT* promoter methylation, gender, *IDH* mutation, subtypes, extent of resection and KPS (at diagnosis) to compare the two predictors using TCGA data (Brennan et al., 2013).

Each microRNA in the signature was assessed for their individual role in survival using univariable Cox regression. The expression levels of all microRNAs were also tested for correlation with each other using Spearman's correlation (Spearman's correlation was chosen as the relationship between two microRNAs may not necessarily be linear).

2.2.3. Prognostic validation of the signature in an independent cohort using qRT-PCR.

(extraction and qPCR performed by Charlotte Tumilson at University of Central Lancashire).

Frozen glioblastoma tissue was obtained from the Brain Tumour North West tissue bank, Royal Preston Hospital, UK. This was collected under local permissions and qPCR was performed at the University of Central Lancashire (Appendix 2.1). Total RNA was extracted using TRIzol® (Life Technologies, UK) according to the manufacturer's guidelines. 1µg of total RNA was reverse transcribed using the NCode miRNA First-strand cDNA synthesis Kit (Life Technologies). Real-time PCR was performed using GoTaq qPCR Master Mix (Promega, Madison, WI) on an Applied Biosystems 7500 PCR Machine with U6 snRNA as an endogenous control (chosen because of its stable expression across normal and cancerous human solid tissues (Peltier & Latham, 2008)). Average Ct values were calculated for each microRNA, and then normalised to U6 average Ct values (dCt). These dCt values were used in the signature algorithm to create risk scores for each patient. One-tailed Cox regression was performed using these scores. The patients were separated according to the 60th percentile (used because there were 10% more patients in this dataset compared to the TCGA dataset with shorter survival than the conventional median of 450 days) and the high- and low-risk groups assessed for association with survival using a one-tailed log-rank test. A one-tailed log-rank test was considered appropriate because the hypothesis is that the high-risk patients have a poorer survival than the low-risk patients, and not the *vice versa*.

2.2.4. Assessment of the 9-microRNA signature in lower grade glioma.

MicroRNA expression for WHO (World Health Organisation) Grade II and Grade III astrocytoma was based on sequencing reads per million mapping to a mature microRNA (as defined in the microRNA database miRbase (Griffiths-Jones et al., 2008)). Risk scores were calculated and significance assessed as above. The median of the lower grade dataset was recalculated and used to separate the samples into two groups.

2.2.5. Cell culture, transfection and validation of candidate microRNA targets.

LN229 glioblastoma cells (ATCC) were cultured in DMEM containing 10% fetal bovine serum at 37°C in 5% CO₂. Cells were transfected with 100nM miR-9 mimic or scrambled control oligonucleotides (ThermoScientific, Waltham, USA), using 10µl of lipofectamine RNAiMAX (Life Technologies, Carlsbad, CA) per 2.5ml of transfection mix in six-well plates containing 150,000 cells/well. RNA was extracted 48 hours post-transfection (miRNeasy, Qiagen, Gaithersburg, MD) and first-strand synthesis catalysed using SuperScript® II Reverse Transcriptase (Life Technologies). Quantitative PCR (qPCR) analyses were performed in triplicate with Taqman assays (Life Technologies) using primers designed by Primer Design Ltd, Southampton, UK. (sequences in Appendix 4.2).

2.2.6. Identifying predicted microRNA targets associated with OS.

Gene expression was compared between two groups of patients with extremely poor prognosis and extremely good prognosis in the TCGA dataset. These were; poor prognosis (survival time < 115 days, *n*=14, minimum KPS at diagnosis=80) and good prognosis groups (survival time >1825 days, *n*=14). These definitions were chosen because they represent the extremes of survival times in the TCGA dataset. The LIMMA (linear models for microarray data) package was used to perform differential expression analysis (Smyth, 2005). This is a modified T-test, where the standard errors are moderated across the genes according to information generated from a linear model of the whole ensemble of genes. The genes with a p-value of less than 0.05 and greater than 1.5-fold change in expression were used as input to RmiR version 1.14, an R-based program for assessment of microRNA targets (RmiR, Favero). Gene ontology analysis was performed using Metacore v6.16 (Thomson Reuters) modified exact Fisher's test and pathways determined using DIANA miRpath (Vlachos et al., 2012) (one-tailed Fisher's exact test for enrichment of predicted microRNA targets). RmiR v1.14 was used to identify targets of the 9 microRNAs amongst the genes which were present in all databases including; Miranda (Miranda et al., 2006), Pictar (Krek et al., 2005) and Targetscan (Lewis et al., 2005) (as loaded by the RmiR vignette). Correlation of microRNA and gene

expression was performed using Spearman's correlation on all 475 glioblastoma samples.

2.3. Results.

2.3.1. Identification of a 9-microRNA signature associated with prognosis in glioblastoma.

In order to identify microRNAs associated with differences in OS in glioblastoma, LASSO regression (Tibshirani, 1996) was performed using microRNA expression data (534 microRNAs, 1510 probes) for 475 glioblastomas. This method is optimised for hi-dimensional data (in which there are more potential predictors than samples) allowing valid inclusion of the 9 microRNAs in the model. The method performs a sub-selection of microRNAs involved in OS by shrinkage of the regression coefficient through imposing a penalty proportional to their size. This results in most potential predictors being shrunk to zero leaving a relatively small number with a weight of non-zero. These microRNAs may not be the only potential predictors in the set, because, if two predictors exhibit co-linearity, LASSO will choose the one that has the strongest association with response (which is not necessarily the only causal one, especially if the difference between the two predictors' degree of association with response is not significant) and the other will be given zero weight.

Using the LASSO method, 12 microRNA probes were identified with non-zero regression coefficients. This included two probes for miR-182, which differed in length by one nucleotide. The longer probe was used for the remainder of the study as a representation for miR-182. This was done so as to provide a single sequence in the signature that can be used for microRNA quantification across platforms. The longer sequence is likely to be more robust as longer sequences align better. Also a probe for miR-565 was identified that has since been excluded from miRBase (Griffiths-Jones et al., 2008) as it is classified as a tRNA fragment; this was not studied further. The LASSO model was refitted without these two probes resulting in a 9-microRNA signature (Table 2.2). MicroRNAs given a negative LASSO coefficient are positive predictors of survival and *vice versa*. Seven of the microRNAs were significantly differentially expressed in non-tumour tissue compared to glioblastoma (Table 2.2). miR-10b,

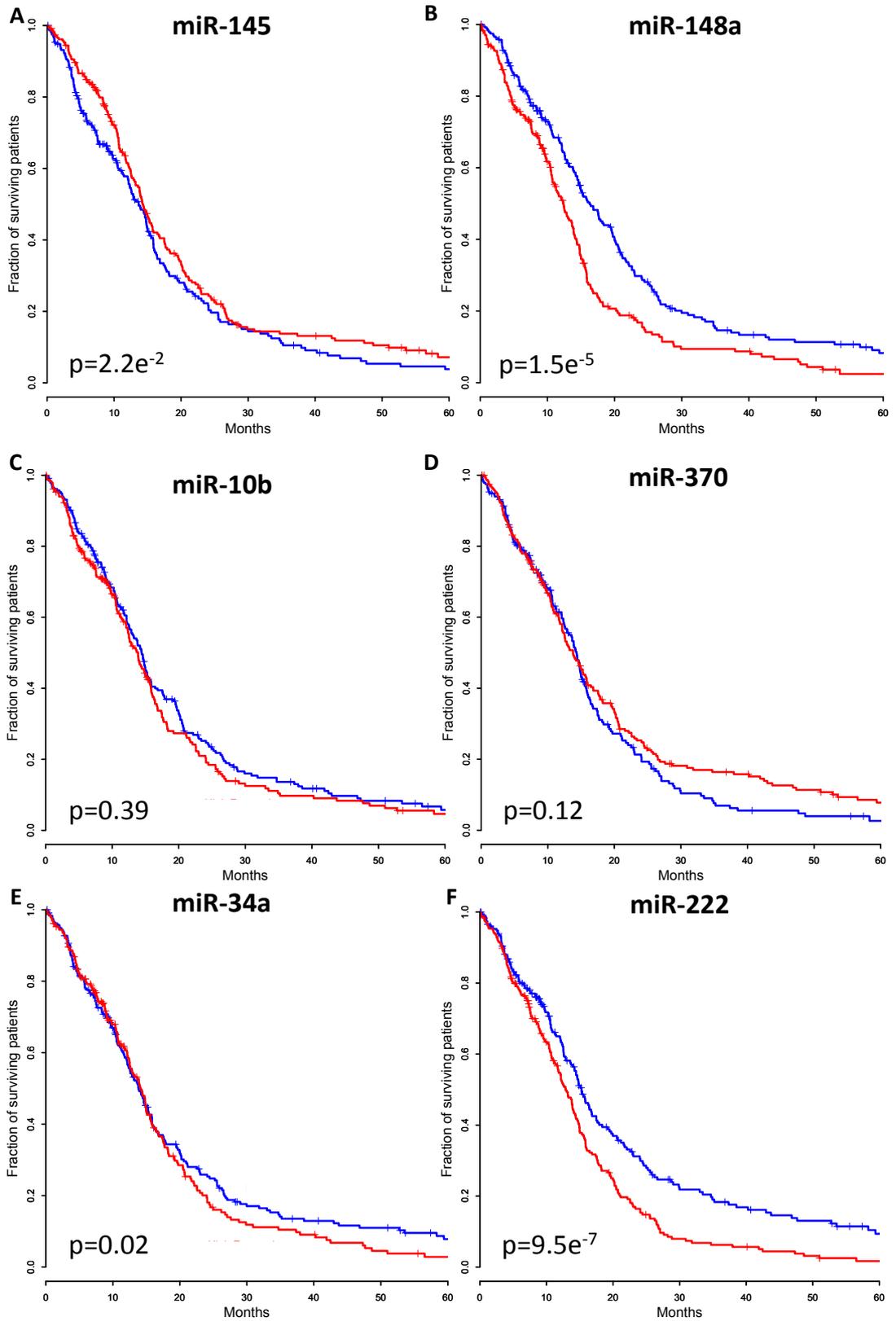
miR-34a, miR-148a and miR-182 were greater in glioblastoma tissue in comparison to non-tumour tissue, whereas miR-124a, miR-145 and miR-222 were less.

MicroRNA	LASSO Penalized coefficient for risk score (log2)	Fold change in GBM compared to non-tumour
miR-124a	0.062	0.032
miR-10b	0.015	10.005
miR-222	0.162	0.278
miR-34a	0.005	3.121
miR-182	-0.021	3.708
miR-148a	0.092	2.752
miR-145	-0.066	0.541
miR-370	-0.044	1.274
miR-9	-0.032	0.863

Table 2.2. MicroRNAs associated with survival using the LASSO regression test.

Significant ($p < 0.05$) results are shown in bold. Nine microRNAs were reported to have non-zero coefficients from LASSO regression with 475 patients and of these, five were negatively associated with patient survival and four were positively associated with survival. Seven were differentially expressed in unmatched non-tumour samples compared to glioblastoma samples. The expressions of these nine microRNAs were used to generate a signature of prognosis in glioblastoma (GBM).

Each of the microRNAs were assessed for their prognostic predictive capacity using Cox regression (Fig. 2.4) and characterised according to their genomic loci (Table 2.3). Four microRNAs; miR-10b ($p=0.39$), miR-124a ($p=0.07$), miR-370 ($p=0.12$) and miR-182 ($p=0.07$) are not individually associated with prognosis using Cox regression. Three of the microRNAs are cleaved from precursors that are encoded in regions of the genome that have been reported to be altered in glioblastoma previously (Riemenschneider et al., 2010).



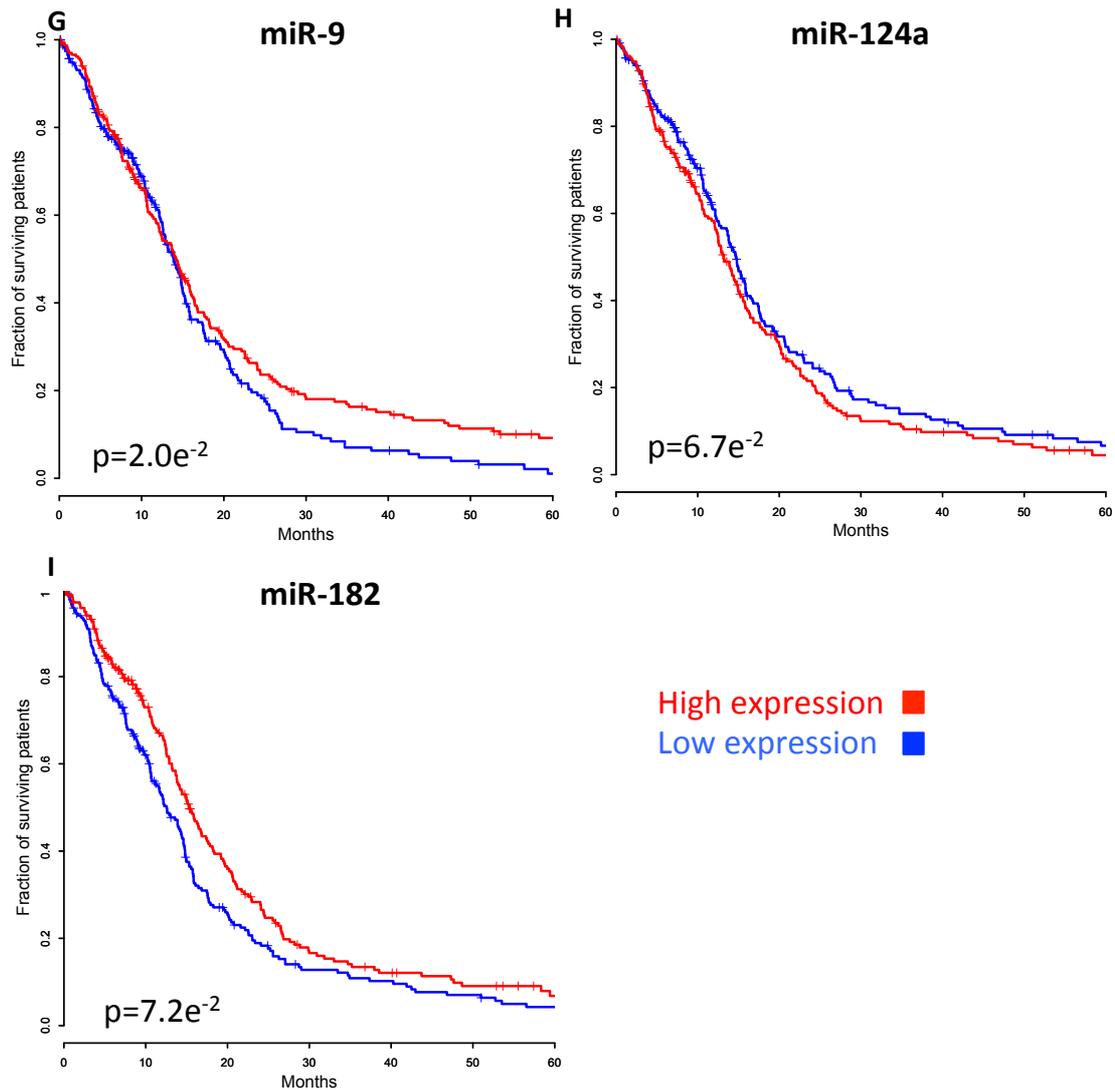


Figure 2.4. Cox regression on all nine microRNAs identified by LASSO regression on 475 glioblastoma patients.

These Kaplan-Meier curves represent log-rank analysis of expression of the microRNA above and below the median microRNA expression of that microRNA. In each case the median microRNA expression for that microRNA was used to dichotomise the data and perform log-rank analysis. The Y-axis represents the fraction of patients alive at each point in time, designated in months on the X-axis. Not all of the microRNAs identified by LASSO regression were significant when tested across the TCGA using univariable Cox regression, which is as expected because LASSO regression takes the expression of other microRNAs into account.

MicroRNA	Chromosome location	Band	Expression in poorer prognosis	Amp/del in Glioma
miR-9	chr1: 156390133-156390221 chr5: 87962671-87962757 chr15: 89911248-89911337	1q22, 5q14.3, 15q26.1	Decreases	NA
miR-148a	chr7: 25989539-25989606	7p15.2	Increases	Amplified
miR-145	chr5: 148810209-148810296	5q32	Decreases	NA
miR-10b	chr2: 177015031-177015140	2q31.1	Increases	NA
miR-222	chrX: 45606421-45606530	Xp11.3	Increases	NA
miR-182	chr7: 129410223-129410332	7q32.2	Decreases	Amplified
miR-370	chr14: 101377476-101377550	14q32.2	Increases	NA
miR-34a	chr1: 9211727-9211836	1p36.22	Increases	1p19q codeletion
miR-124a	chr8: 9760898-9760982, chr8: 65291706-65291814, chr20: 61809852-61809938	8p23.1, 8q12.3, 20q13.33	Increases	NA

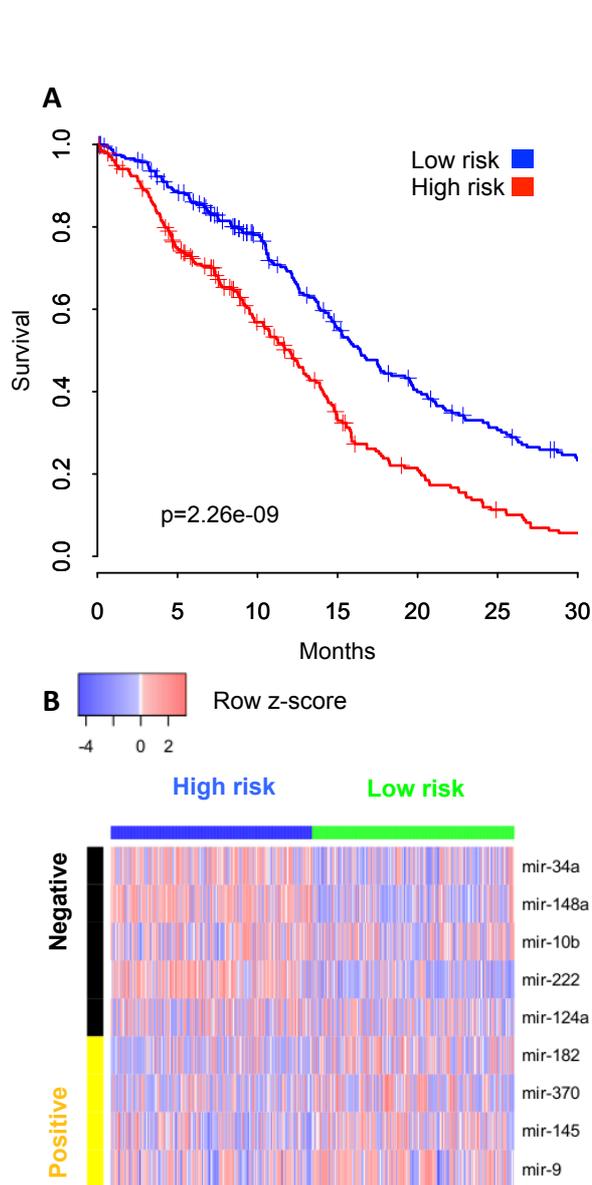
Table 2.3. The genomic location of the microRNAs identified by LASSO regression.

The genomic coordinates for the precursors of each of the nine microRNAs in the signature. Seven of the microRNAs identified have one precursor sequence, but miR-9 and miR-124a both have three. Three microRNAs; miR-148a, miR-182 and miR-34a, reside in regions of the genome that have been reported to be cytogenetically altered in glioblastoma; the short arm of chromosome 1 and chromosome 7 (Riemenschneider et al., 2010). miR-182 is decreased in poorer prognosis yet the chromosome it is encoded on (chromosome 7) is gained in tumours with a poorer prognosis (Goodenberger & Jenkins, 2012).

2.3.2. Generation of a risk score combining expression values of the 9 microRNAs to predict survival.

A risk score was created using the regression coefficients from the LASSO analysis (see methods, page 68) to weight the expression value of each of the 9 microRNAs. The risk score was then separated on the median (1.48 quantile normalised probe expression) to create high and low risk groups. The median survival time of the low-risk group was 13.1 months and the median of the high-risk group was 9.5 months. Risk score was associated with survival using log-rank test (Fig. 2.4, $p=2.26e^{-09}$). Median expression of each signature microRNA in both groups is shown in Fig. 2.5.

Figure 2.5. The patient groups assigned to the high- and low-risk groups using the median as a threshold.



A score for each patient was calculated using the microRNA expression signature and patients were separated into high and low risk groups using the median (1.48 normalised expression) as a cut-off. A) The low-risk group has significantly longer survival times than those in the high-risk group by log-rank test ($p=2.26e-09$). The Y-axis represents the fraction of patients alive at each point in time, designated in months on the X-axis. B) Expression patterns of the significant microRNAs in the high- and low-risk groups, as defined by the risk score, shown in a heatmap. The positive microRNAs (yellow) represent microRNAs that are lower in poorer prognosis and negative (black) represent the opposite. Blue in the heatmap indicates low expression and red indicates high expression in a sample.

Mirna	Median quantile normalised expression of microarray probe	
	Low risk	High risk
miR-124a	8.62	8.90
miR-10b	9.25	9.45
miR-222	9.36	10.58
miR-34a	10.68	11.49
miR-148a	7.94	9.19
miR-182	8.09	7.52
miR-145	9.57	9.21
miR-370	10.04	9.74
miR-9	13.34	12.86

Table 2.4. Median quantile normalised expression of microarray probe for each microRNA across 475 glioblastoma tumours in the TCGA.

Differences between the two groups are often not large, but combined together into a risk score algorithm these microRNAs have prognostic potential. miR-124a which is a microRNA involved in neuronal differentiation, is one of the lowest expressed microRNAs, and miR-9 which is an abundant microRNA is expressed at the highest levels of all microRNAs studied (Krichevsky et al., 2006; Fowler et al., 2011). The levels between the risk groups are not as markedly different as the microRNAs are from each other in the same group.

Pearson's correlation of age with risk score showed a significant direct correlation ($R=0.248$, $p=4.13e^{-08}$). Multivariable Cox regression of the risk group and age showed the risk group to be an independent predictor of survival irrespective of age (Group HR=1.61, 95% CI=1.30-1.99, $p=1.40e^{-5}$; Age HR=1.03, CI= 1.02-1.04, $p=2.50e^{-3}$). As males have poorer outcome in glioblastoma (Krex et al., 2007), the risk score was included in a Cox regression model with gender, and was found to be similar in the male and female groups (median 1.48 in each group).

2.3.3. Assessment of the risk score in glioblastoma subtypes and in relation to other prognostic factors.

I then determined the risk groups for each of the TCGA-defined glioblastoma molecular subtypes (Brennan et al., 2013): proneural G-CIMP (glioma CpG island methylator phenotype) positive ($n=36$), proneural G-CIMP negative ($n=88$), neural ($n=77$), classical ($n=128$) and mesenchymal ($n=143$). MicroRNA-defined risk group was associated with survival in all subtypes except proneural G-CIMP negative glioblastoma (Fig. 2.6 A-E).

The groups were then fitted to a Cox regression model incorporating age in each patient subtype. The score remained significant in the classical (HR=1.73, 95% CI=1.13-2.64, $p=0.011$) and neural (HR=2.03, 95% CI=1.23-3.38, $p=0.007$) subgroups and age was a confounding factor in the mesenchymal group (HR=1.46, 95% CI=0.95-2.23, $p=0.084$). This suggests that the signature is more likely associated with age rather than survival in mesenchymal glioblastoma. The performance of the risk score in the proneural G-CIMP positive group could not be calculated because all samples but one stratified to the low risk group. This was expected because G-CIMP glioblastoma is the subtype of glioblastoma with the best prognosis. The risk score in the proneural G-CIMP negative group was not significant (HR=1.15, 95% CI=0.70-1.86, $p=0.059$). This may suggest that this subtype of glioblastoma, which is particularly refractory to treatment, has different microRNA biology to the other subtypes (Brennan et al., 2013). The survival groups also had significantly different PFS by log-rank ($p=9.91e^{-08}$) (Fig. 2.6 F), which indicates the signature is predictor of PFS as well as OS. There were 26 samples in the cohort with *IDH1* mutations, only one of which stratified to the high-risk group, which suggests the signature is selecting for a subtype with already known survival differences.

The risk score was evaluated by fitting a Cox model incorporating the risk group and other factors involved in glioblastoma prognosis (gender, *MGMT* promoter methylation, *IDH* mutation, patient subtype, extent of resection and KPS score). In each case, the score was significant and was not related to these factors (Table 2.5).

Table 2.5. Results of multivariable Cox regression incorporating prognostic factors in glioblastoma.

Factor	Hazard ratio of variable	CI variable	p-value variable	Hazard ratio for risk group	CI risk group	p-value risk group
Age	1.03	1.02-1.04	1.60E-05	1.6	1.29-1.98	2.30E-13
Gender	1.01	0.82-1.25	9.20E-01	1.88	1.52-2.33	5.30E-09
MGMT meth	meth=0.85 unmeth=1.17	0.66-1.09 1.92-1.50	2.00E-01 2.10E-01	1.87 1.87	1.51-2.31 1.51-2.31	9.00E-09 9.00E-09
IDH	R132G= 0.94 R132H = 0.41	0.13-6.76 0.23-0.71	9.50E-01 1.70E-03	1.76 1.76	1.42-2.18 1.42-2.18	2.80E-07 2.80E-07
Subtype	WT= 0.81 GCIMP=0.44	0.64-1.02 0.27-0.71	6.70E-02 7.50E-04	1.76 1.75	1.42-2.18 1.4-2.19	2.80E-07 9.00E-07
	Mesenchymal= 0.93 Neural= 1.03	0.70-1.23 0.75-1.41	6.20E-01 8.60E-01	1.75 1.75	1.4-2.19 1.4-2.19	9.00E-07 9.00E-07
	Proneural=1.43	1.05-1.94	2.30E-02	1.75	1.4-2.19	9.00E-07
Extent of resection	Excisional Biopsy=0.68	0.40-1.16	1.60E-01	1.86	1.51-2.31	1.10E-08
	Fine needle aspiration biopsy =0.33	0.04-2.50	2.90E-01	1.86	1.51-2.31	1.10E-08
	Incisional Biopsy=1.74	0.23-13.04	5.90E-01	1.86	1.51-2.31	1.10E-08
	Other method=0.30	0.04-2.25	2.40E-01	1.86	1.51-2.31	1.10E-08
	Tumor resection=0.77	0.48-1.22	2.60E-01	1.86	1.51-2.31	1.10E-08
KPS (n=361)	0.98	0.97-0.99	1.00E-06	1.79	1.39-2.29	4.50E-06

In all cases the score significantly predicted survival therefore is a survival predictor independent of these variables. The hazard ratio of each variable is given in column 2, along with 95% confidence intervals in column 3. The p-value for this relationship is provided in column 4. The hazard ratio and 95% confidence intervals for the risk group from the multivariable Cox regression are represented in columns 5 and 6 along with associated p-values in column 7.

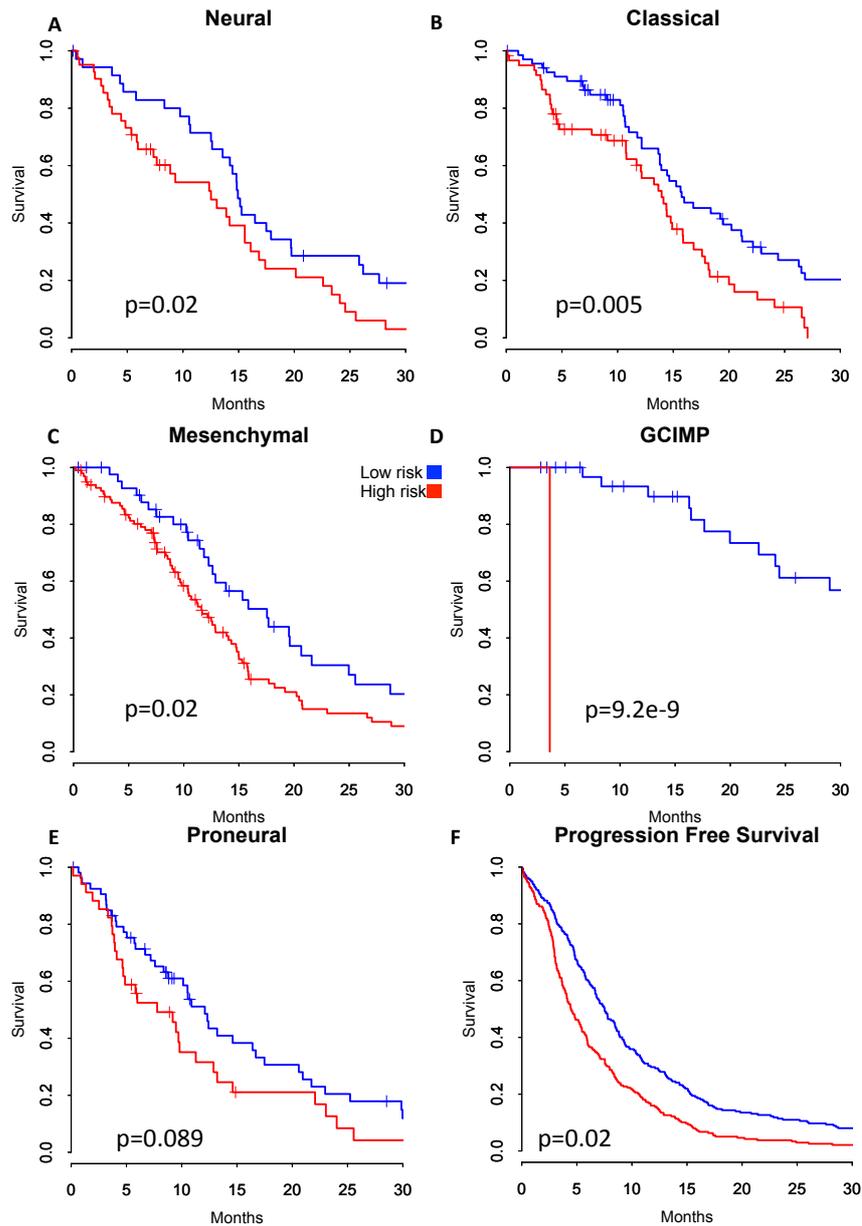


Figure 2.6. Log-rank of the low-risk and high-risk groups in subgroups of glioblastoma.

Risk scores were calculated with the same threshold as the whole cohort for each subtype of glioblastoma. These were then split into groups based on the median and log-rank performed to assess survival association of the groups. The Y-axis represents the fraction of patients alive at each time-point, and the time in months from diagnosis is represented on the X-axis. Blue lines are the low-risk group and red lines are the high-risk group. The risk groups were significant by log-rank test (non-age adjusted) in all molecular subtypes of glioblastoma but proneural G-CIMP negative (A-E). Risk score is also a significant predictor of progression free survival (F).

I then calculated the risk score solely in the group of patients treated with the most common chemotherapy agent, temozolomide ($n=219$). This group showed a significant association between risk score and survival using log-rank ($p=8.6e^{-04}$) (Fig. 2.7 A). These results indicate that, in patients treated with the standard treatment, the microRNA signature predicts survival.

The predictive power of the signature was compared to that of *MGMT* promoter methylation status by the log-rank test. In the 304 patients for whom *MGMT* promoter methylation status was available (Brennan et al., 2013), multivariable Cox regression indicated that the microRNA signature (HR=1.88, CI=1.42-2.48, $p=9.4e^{-06}$) predicted survival. In the same group of patients, *MGMT* promoter methylation also predicted survival but with less significance than the microRNA signature (HR=1.47, CI=1.12-1.93, $p=0.006$). Comparing the power of the two predictions, the microRNA signature results in a 1.88-fold increased risk when stratified to the high-risk group and the *MGMT* promoter methylation signature results in a 1.47-fold increase in risk when the *MGMT* promoter is unmethylated. In the group treated with temozolomide only ($n=219$) there was a 1.76-fold increase in risk by stratification to the low-risk group; this stratifies patients better than the *MGMT* signature, which shows a 1.65-fold increase in risk when stratified to the unmethylated group in the TCGA dataset.

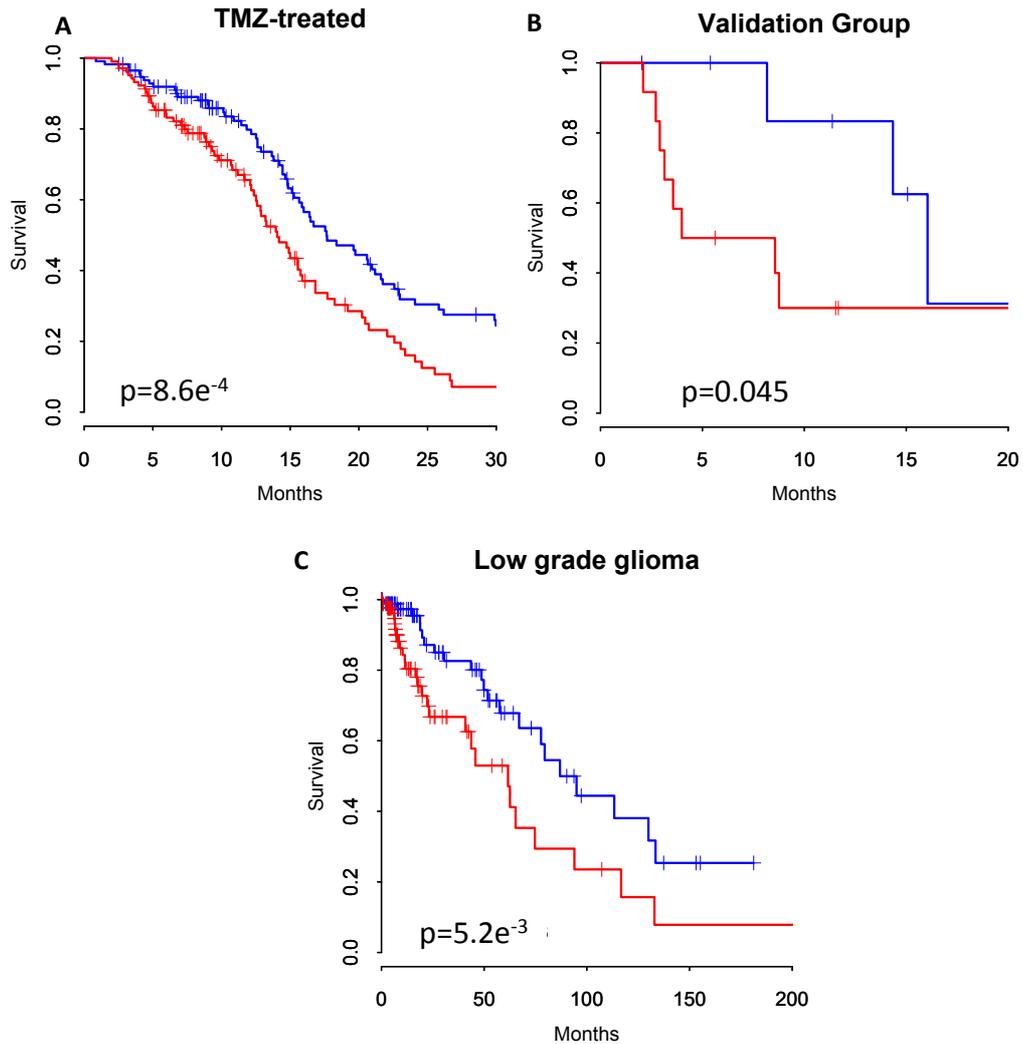


Figure 2.7. Assessment of risk groups in TMZ treated patients, the validation cohort and lower grade glioma.

Kaplan-Meier curves of the signature in the different groups are shown. The Y-axes represent the fraction of patients alive at each time point, and the times in months are on the X-axes. A) The subgroup of patients treated with the chemotherapy agent temozolomide was significantly delineated using the signature ($p=8.4e^{-4}$). B) MicroRNA expression determined by qRT-PCR (quantitative real-time PCR) in an independent cohort of 20 glioblastomas stratified patients by survival based on the signature ($p=0.045$, one-tailed test). C) MicroRNA sequencing data of 178 lower grade glioma samples (55 astrocytoma, 47 oligoastrocytoma, and 75 oligodendroglioma, 1 not stated) significantly separated these samples into high and low risk groups by log-rank test.

2.3.4. Risk score validation in an independent dataset.

The nine-microRNA signature has been generated and tested on the same set dataset. This is possible because LASSO employs leave-one-out cross validation, where the model is rerun iteratively, each time leaving a sample out. This reduces over fitting of the model to the dataset. Ideally, validation should be performed on an independent dataset from a different population, or with a different technique. Risk scores were calculated for an independent dataset from the University of Central Lancashire of 20 glioblastoma samples (Table 2.6), with microRNA expression generated using qRT-PCR (quantitative real-time PCR) and was significantly associated with survival (HR=10.7, p=0.036). This patient group had an overall worse prognosis (80% died earlier than the conventional median of 450 days (Mountz et al., 2014)) than those in the TCGA (70% died earlier than the conventional median of 450 days), and therefore, expecting more patients to fall into the high-risk group, the patients were dichotomized based on the 60th percentile (0.76 DCt). This resulted in 12 patients in the high-risk group with a median survival of 6.27 months and 8 patients in the low-risk group with a median survival of 16 months. These groups predict survival using a one-sided log-rank test (HR=3.01, p=0.045) (Fig. 2.7B).

Characteristic	Number of patients (n=20)
Age (median=68)	
<60 years	5
≥60 years	15
Gender	
Male	11
Female	9
Days to death/ last follow-up (median 268 days)	
<450 days	16
≥450 days	4
<30 days	0

Table 2.6. Patient characteristics in the independent dataset used for validation.

This group had an overall worse prognosis than the TCGA dataset (the ratio to the conventional median was 10% higher than the ratio in the TCGA dataset) and therefore the cut-off for the high and low risk groups was shifted to the 60th percentile.

2.3.5. Risk score assessment in lower grade glioma.

To serve as a further validation, and also to see whether the score is applicable to lower grade glioma, I calculated risk score on the TCGA dataset of lower grade gliomas. Risk scores were also calculated for grade II and III gliomas ($n=178$) consisting of 81 grade II and 96 grade III samples; 55 were astrocytomas, 47 oligoastrocytomas and 75 oligodendrogliomas, using TCGA sequencing data. This was done using the 9 microRNAs derived in the glioblastoma signature. The cohort was dichotomised into high- and low-risk groups using the median (-19541.96 reads per million) as a cut-off. As observed in the glioblastoma dataset, the score proved to be a significant predictor of survival using log-rank (Fig. 2.7C, $p=5.2e^{-03}$) and in a Cox model with age (Group HR=0.62, CI=1.05-3.31, $p=3.5e^{-02}$; Age HR=1.06, CI=1.04-1.10, $p=2.2e^{-07}$). The low-risk group comprised of 44 grade II and 45 grade III samples; 22

were astrocytomas, 22 oligoastrocytomas and 45 oligodendrogliomas. The high-risk group comprised of 37 grade II samples and 51 grade III samples (1 not stated); 33 were astrocytomas, 25 oligoastrocytomas and 30 oligodendrogliomas.

2.3.6. Predicted targets of these microRNAs.

Bioinformatic analysis was used to investigate targets of signature microRNAs to identify the associated pathways involved. Firstly, genes associated with long and short survival groups in glioblastoma were identified. The mRNA expression between two groups of patients with extremely poor prognosis (survival time < 115 days, $n=14$, minimum KPS at diagnosis=80) and extremely good prognosis (survival time >1825 days, $n=14$) were compared in the TCGA dataset. The genes with a p-value of less than 0.05 and greater than 1.5-fold change in expression were used as genes associated with survival. A total of 1154 genes were associated with short and 400 genes with long survival. Predicted interactions of the 9 microRNAs with the survival-associated genes were assessed in the Miranda, Pictar and Targetscan databases (Krek et al., 2005; Lewis et al., 2005; Miranda et al., 2006). This led to the identification of 162 significant predicted microRNA/mRNA interactions, 10 of which had an inverse correlation of at least 0.25 across all glioblastoma samples (Table 2.7). This was chosen because the maximum correlation detected was 0.5 between an mRNA and microRNA and therefore half this was used as a threshold. Using DIANA miRPath (Vlachos et al., 2012) I identified the top pathways that the signature microRNAs are predicted to target. The most significant pathways identified included adherens junction, MAPK signalling, focal adhesion, axon guidance and WNT signalling (Appendix 2.2).

MicroRNA	LASSO penalized Coefficient (log2)	Gene symbol	Gene change with increasing survival	Spearman's Correlation	Fold difference in GBM to non-tumour tissue	P-value of GBM/normal (FDR adjusted)
hsa-miR-9	-0.032	TGFBI	4.499	-0.649	11.487	0.000
hsa-miR-9	-0.032	P4HA2	2.527	-0.615	1.108	0.999
hsa-miR-9	-0.032	FBN1	2.054	-0.53	1.808	0.001
hsa-miR-222	0.162	KHDRBS2	0.189	-0.496	0.024	0.000
hsa-miR-9	-0.032	SLC25A24	3.574	-0.473	2.17	0.000
hsa-miR-9	-0.032	SLC31A2	2.384	-0.463	0.593	0.039
hsa-miR-9	-0.032	FNDC3B	2.171	-0.406	3.828	0.000
hsa-miR-182	-0.021	F13A1	10.982	-0.309	1.785	0.106
hsa-miR-9	-0.032	LMNA	2.034	-0.292	2.25	0.000
hsa-miR-9	-0.032	WNT4	2.038	-0.265	0.691	0.003

Table 2.7. Predicted target interactions of the signature microRNAs with significant correlation in expression.

The ten interactions predicted between the 9-microRNA signature and the mRNAs identified to be involved in survival in glioblastoma (GBM), which also showed a significant inverse correlation in expression of at least 0.25 across the patient set. Two of these mRNAs, *FBN1* (fibrillin 1) and *TGFBI* (TGFB induced), exhibited particularly high correlations in expression with miR-9 (-0.53 and -0.65 respectively) as well as significant differential expression between glioblastoma compared to non-tumour tissue.

Targets implicated most strongly in patient survival were identified for miR-9, which showed a significant correlation with eight mRNAs. In order to assess whether these may be functional targets, a glioblastoma cell line was transfected with a miR-9 mimic and the expression levels of the predicted targets were assessed using qPCR. *LMNA* (lamin A), *WNT4* (Wingless-Type MMTV Integration Site Family, Member 4), *FBN1* (fibrillin 1), *P4HA2* (prolyl 4-hydroxylase, alpha polypeptide II) and *SLC25A24* (Solute Carrier Family 25 (Mitochondrial Carrier; Phosphate Carrier), Member 24) had significantly lower levels of expression when transfected with the mimic in comparison to a scrambled control (Fig. 2.8) suggesting miR-9 may directly target these mRNAs

in glioblastoma cells. These mRNAs have predicted targets for miR-9 in their 3' UTR and also decrease when miR-9 is overexpressed in glioblastoma cells. Thus, bioinformatic analysis of signature microRNAs has identified potential targets and biological processes known to be involved in glioblastoma biology, further supporting the relevance of the 9-microRNA signature.

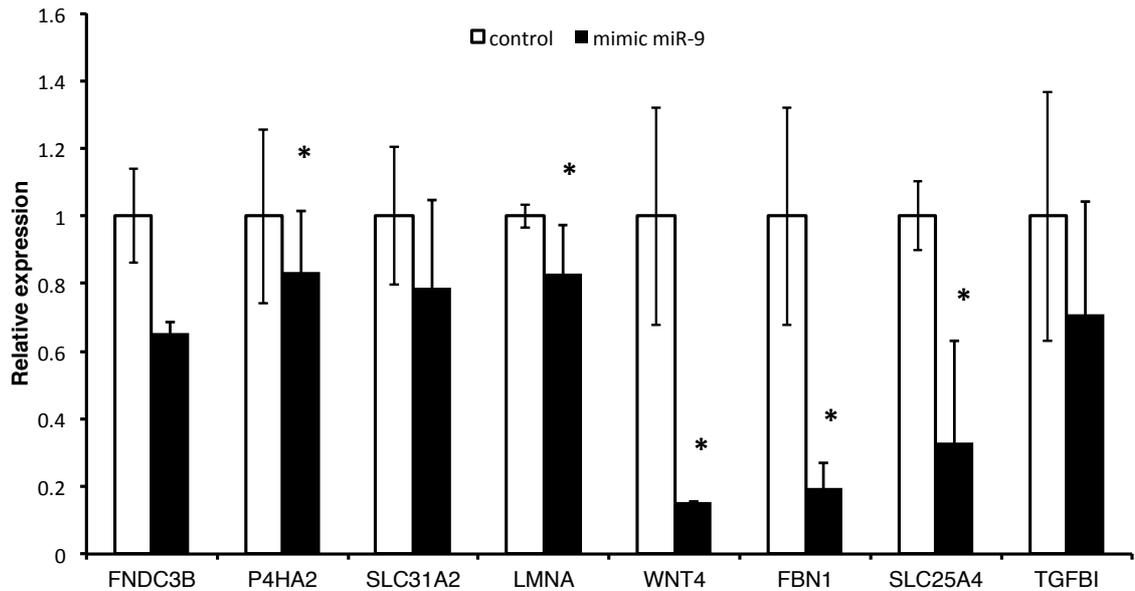


Figure 2.8. Expression of the predicted targets following transfection of a miR-9 mimic into LN229 cells relative to a scrambled control.

miR-9 was predicted to target the most mRNAs identified to be prognostic compared to all other microRNAs in the signature. This bar chart shows the levels of expression of the 8 predicted targets after transfection of a miR-9 mimic or scramble into LN229 cells. Significant decrease in expression (t-test, $p < 0.05$) was observed for P4HA2, LMNA, WNT4, FBN1 and SLC25A4 48 hours after transfection of the mimic. Results are representative of duplicate experiments. The Y-axis represents the mRNA expression relative to GAPDH control mRNA. This is shown in white boxes for the control, scrambled mimic sequence and black boxes for the miR-9 mimic. Values have been normalised to the scrambled control.

2.4. Discussion.

2.4.1. The nine microRNA signature is a molecular indicator of prognosis.

Using LASSO regression, this study has identified and independently validated a biologically relevant 9-microRNA signature that predicts survival in glioblastoma. The signature separates patients into high- and low-risk groups with respect to OS and PFS and may have clinical utility for decisions on patient management. For example, patients stratified to the high-risk group could be monitored more closely, or targeted towards novel treatments. The signature is valid in all glioblastoma subtypes except proneural G-CIMP negative tumours (which represents only 19% of this dataset), and also predicts survival in patients treated with the standard chemotherapy drug temozolomide.

The independent dataset used here is relatively small ($n=20$) and therefore confounding factors for patient age, treatment received and extent of resection could not be accounted for. The independent dataset results were generated using qRT-PCR and indicate that the signature can be implemented using techniques that would be more conducive to a clinical diagnostic laboratory and these are the methods that should be explored further. A limitation of this approach is that a different technique has been used for validation and therefore a single, defined cut-off could not be ascertained.

Further validation, ideally prospective, and calculation of sensitivity and specificity, is required before this signature could be implemented clinically. This would require a prospective clinical trial. This would involve stratifying patients treated with standard treatment into high- and low- risk groups and data collection of their survival times as well as other clinical factors such as extent of resection and KPS.

Prognostic signatures using microRNAs have been formulated previously in glioblastoma but these have not been validated or evaluated within different subgroups of the disease, or in relation to molecular characteristics of the disease (Kim et al., 2011b; Niyazi et al., 2011; Srinivasan et al., 2011; Sana et al., 2014). A more recent study identified prognostic microRNAs for each subtype of glioblastoma using TCGA data (Li et al., 2014b) and five microRNAs in our signature overlap; miR-222, which they report predicts prognosis in classical and neural glioblastoma, miR-370 which predicts prognosis in neural

glioblastoma and miR-34a, miR-145 and miR-182 which predict prognosis in the proneural non-G-CIMP glioblastoma group. Interestingly, 3/9 microRNAs in my signature are present in their model for proneural G-CIMP negative glioblastomas yet my signature did not significantly stratify patients in this subtype. This may be because the other microRNAs in my signature are acting in different ways in this subtype, therefore creating noise in the score for these patients.

The LASSO regression model was chosen to improve on other approaches by utilising all 475 patients, and all microRNAs available to build the signature. This allows a small number of microRNAs for use in a diagnostic signature with maximal information but does not identify all predictors in the dataset involved in survival. This provides a signature with the prediction power similar, or better than, that of *MGMT* promoter methylation. It must be noted however that *MGMT* promoter methylation was assessed in an unselected population, with the Infinium methylation bead chip (Bady et al., 2012), which is not the gold standard employed in a diagnostic laboratory and therefore may lack sensitivity compared to clinical results. *MGMT* promoter methylation was also not assessed in the validation dataset due to lack of methylation data so this finding requires further confirmation. This signature has a manageable number of microRNAs for a prognostic indicator, and is well below the number of predictors employed in commercialised kits for other cancer signatures such as the gene tests, Mammaprint and ms-14 in breast cancer (Sorlie et al., 2001; Cheang et al., 2009).

2.4.2. Roles of the microRNAs in the signature in glioma biology.

All microRNAs in this signature, with the exception of miR-370, have been previously associated with glioma biology (Zhang et al., 2010; Fowler et al., 2011; Gabriely et al., 2011; Kim et al., 2011a; Genovese et al., 2012; Song et al., 2012; Rani et al., 2013; Tan et al., 2013; Mucaj et al., 2014). For example, the microRNA with the most predicted targets involved in survival, miR-9, has been shown to be associated with resistance to temozolomide and its actions in glioma biology are explored further in Chapter four of this thesis. Additionally, miR-148a has been shown to target the EGFR regulators MIG6 and BIM in glioblastoma and inhibition of this microRNA decreased growth of glioma stem

cell and xenograft growth *in vivo* (Kim et al., 2014). This has not been shown for previous glioblastoma microRNA signatures (Lakomy et al., 2011; Srinivasan et al., 2011.; Zhang et al., 2012a, Sana et al., 2014). Although miR-370 has not been reported to have a role in glioblastoma, it targets TGFB-RII (TGFB receptor 2) (Lo et al., 2012), which has a role in glioblastoma cell growth and invasion (Kaminska et al., 2013). These studies, and the data presented here, suggest a potential role for miR-370 in glioma biology. Establishing a defined role for these microRNAs in glioma biology requires further work to determine the link between the biology these microRNAs regulate and patient prognosis. In addition to their established roles in glioma biology, 5 of the 9 signature microRNAs have been associated with sensitivity to TMZ; miR-9 (Munoz et al., 2013), miR-145 (Yang et al., 2012b), miR-148a (Hummel et al., 2011), miR-182 (Tang et al., 2013) and miR-222 (Chen et al., 2012a). It is likely however, that the microRNAs are providing a measure of all treatment including not just chemotherapy but radiotherapy and surgery. For example, glioblastoma tumours with a mesenchymal subtype exhibit more infiltration and therefore extent of resection will not be as high as in the proneural subtype (Beier et al., 2012). These observations suggest that the microRNA signature reflects roles in both tumour biology and treatment resistance, which combined lead to robust effects on patient survival.

2.4.3. Translational relevance of the signature.

This prognostic signature has potential applicability to the clinic by stratifying patients, and identifying those less likely to respond to current treatments. The signature ultimately may facilitate confidence in treatment decisions and recognising candidates for new therapies. It may be that the most powerful use of the signature is in combination with *MGMT* promoter methylation status. Technologies such as the NanoString nCounter platform may provide highly accurate quantitative measurements of transcripts for tumour diagnosis as has been shown for medulloblastoma (Northcott et al., 2012), and is readily applicable to microRNA studies.

In conclusion, I have identified and validated a 9-microRNA-expression signature using biologically relevant markers of use in prediction of prognosis in glioblastoma. Analysis of targets of these microRNAs has identified potential

key players in glioblastoma networks that could be targeted to combat the aggressive disease. The LASSO approach may be more broadly applicable in the identification of relevant microRNA and gene expression signatures in large datasets.

3. Prognostic microRNAs in high-grade glioma reveal a link to oligodendrocyte precursor differentiation.

'It is my belief that the basic knowledge that we're providing to the world will have a profound impact on the human condition and the treatments for disease and our view of our place on the biological continuum.' **Craig Venter, June 2000.**

3.1. Introduction.

Molecular subtypes of glioblastoma can be defined by clustering according to cell type-specific mRNA expression patterns (Verhaak et al., 2010; Brennan et al., 2013). Verhaak *et al.* identified classical, proneural, neural, and mesenchymal subtypes of glioblastoma using mRNA expression, somatic mutation, and copy number data obtained from TCGA (The Cancer Genome Atlas) (Verhaak et al., 2010; TCGA, NIH). Interestingly, clustering analysis of signature gene expression patterns of the four subtypes with expression patterns from murine neural cells showed that they are reminiscent of specific neural cell types, for example the proneural class of glioblastoma has an oligodendrocyte rather than astrocyte signature. The proneural glioblastoma subtype is also particularly refractory to the current standard treatment of radiotherapy and temozolomide and a recent study by Ozawa *et al.* indicates that most glioblastoma subtypes can arise from a common proneural-like precursor cell (Verhaak et al., 2010; Ozawa et al., 2014). A consistent body of literature supports the notion that the presence of less differentiated cells in cancer confers a poorer prognosis and it may therefore be possible to identify common signatures of aggressive clinical behaviour in glioma based on progenitor cell types (Dirks, 2010; Garrido et al., 2014; Waghmare et al., 2014; Auffinger et al., 2014).

In this context, microRNAs may be relevant, as changes in microRNA expression are emerging as a common feature of both neural development and glioma biology (Godlewski et al., 2010b). MicroRNAs have roles in the maintenance of brain functions throughout life and are extensively dysregulated in cancer (Carroll & Schaefer, 2012; Stahlhut & Slack, 2013). In brain tumours, they have been shown to promote 'stemness' or inhibit differentiation, consequently maintaining tumourigenesis (Schraivogel et al., 2011). Their expression is also altered in stem-like compartments of both brain tumours and other tumours and has been reviewed by Stappert *et al.* (Stappert et al., 2014).

In addition, microRNAs modulate neural differentiation and their expression patterns have been shown to be distinct at different cellular stages of differentiation, including oligodendrocyte precursor (OP) differentiation (Letzen et al., 2010). The presence of stem-like cells in brain cancer has been shown to be associated with more aggressive, treatment resistant tumours (Dirks, 2010; Garrido et al., 2014; Auffinger et al., 2014). It is established that microRNAs have a role in maintaining a specific differentiation phenotype but it remains unclear whether prognostic microRNA signatures are exclusively tumour grade and/or molecular subtype-specific, or whether common signatures, for example associated with differentiation status, can be identified (Aldaz et al., 2013). Here I have used a computational approach to test the hypothesis that differential microRNA expression profiles in groups of glioma patients with good and poor prognosis reflect changes in progenitor development pathways. I therefore correlated the microRNA expression changes between good and poor prognosis groups with microRNA expression changes in the OP differentiation pathway. OP differentiation can be modelled *in vitro* using embryonic stem cells (ESCs) that adopt an oligodendrocyte cell fate in a step-wise fashion using instructive cell culture conditions (Letzen et al., 2010). The differentiation steps include embryoid body (EBs), a neural progenitor cell state (NP), the oligodendrocyte progenitor stages OP1, OP2, and OP3 and the fully differentiated oligodendrocyte lineage (OL) (Figure 3.1). Analysis of microRNA profiles of these cell types showed that expression changes during OP differentiation correlate with prognostic microRNA expression changes in malignant glioma. This correlation is most apparent for the OP1 cell stage, which consistently predicts survival (in >500 gliomas), hence suggesting a prognostic signature of aggressive clinical behaviour that is independent of grade and malignant brain tumour subtype.

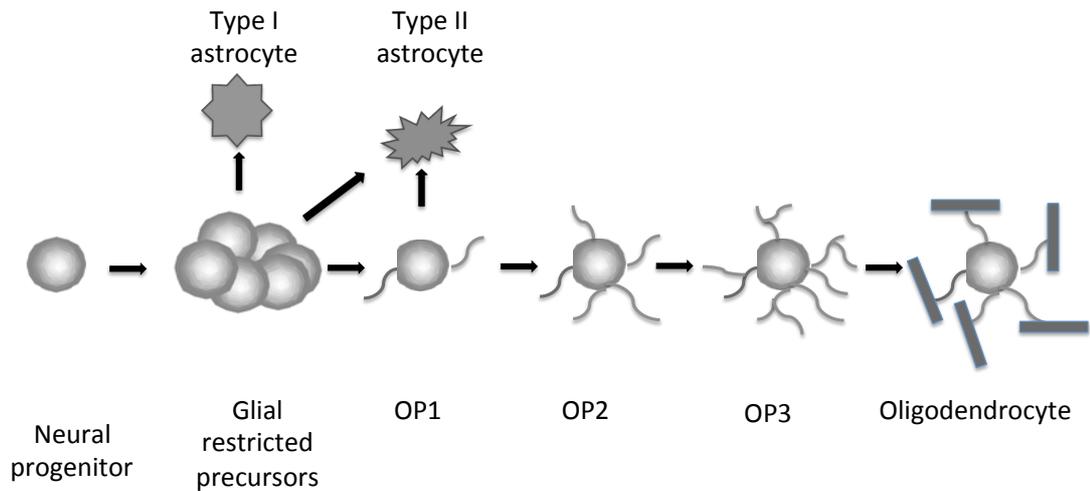


Figure 3.1. The cell stages in the oligodendrocyte precursor (OP) differentiation pathway. Oligodendrocyte precursors are intermediates between neural progenitors and astrocytes and oligodendrocytes.

3.2. Methods.

3.2.1. MicroRNA and mRNA expression analysis.

All computational work was performed in R (v2.15.1). Level 3 Agilent microRNA 8×15k microarray, G4520A microarray gene expression data and clinical information for glioblastoma and non-tumour samples were downloaded from TCGA (TCGA, NIH). Level 3 Illumina HiSeq sequencing data for mature microRNA and mRNA expression plus clinical information for lower grade gliomas were also downloaded from TCGA. There are differences in normalisation methods and quantification artefacts of microarray and sequencing platforms, and therefore merging of the data in normalised format may result in artefacts and spurious results. This is particularly apparent for microRNA expression below the median, where microarray underperforms (C. Wang et al., 2014a). Therefore I chose to perform differential expression analysis prior to merging the two sets of results. The expression changes within the grades were ascertained using appropriate statistical packages developed for microarrays and sequencing (Smyth, 2005; Robinson et al., 2009) and then these data were merged, rather than direct merging of the data prior to differential expression analysis.

The good and poor prognosis groups of these glioma datasets were selected according to the published survival data in the TCGA database (Table 3.1).

		GIIIA		Glioblastoma	
		Good prognosis (>48 months, n=6)	Poor prognosis (<18 months, n=10)	Good prognosis (>48 months, n=13)	Poor prognosis (<4 months, n=14)
Age at Diagnosis		40.5	59.5	41.5	61.9
Mean Overall Survival (months)		87	8.6	83.4	2.7
Gender	Male	67%	40%	62%	50%
	Female	33%	60%	38%	50%
IDH1 mutation status	Mutated	100%	10%	0%	0%
	WT	0%	90%	100%	100%

Table 3.1. Characteristics of the grade III astrocytoma and glioblastoma TCGA tumours in poor and good prognosis groups.

As expected, in both GIIIA and glioblastoma age is increased in the poorer prognosis groups. There are only IDH1 mutations in the GIIIA group. In glioblastoma, IDH1 mutation is usually present in tumours following progression from a lower grade, and as the TCGA has sought samples from mainly primary glioblastoma, this is expected.

EdgeR, a package designed for analysis of differential gene expression from RNA-seq data, first calculates the interlibrary variation for each gene using tagwise dispersion, then determines differential expression using the exact negative binomial test (Robinson et al., 2009). EdgeR was used to compare microRNA and mRNA expression sequencing data between the two GIIIA survival groups and 139 *IDH* mutated and 39 *IDH* wild-type grade II and III tumours (Robinson et al., 2009). The LIMMA (linear models for microarray data) package fits a linear model to the expression data for each gene and then tests the differences between the parameters of the model (Smyth, 2005). LIMMA was used to compare microarray-based microRNA expression data for the poor and good prognosis groups in glioblastoma.

For each microRNA or mRNA, r , the z-scores associated with GIIIA (grade III astrocytoma) and glioblastoma (IV) prognosis were calculated separately from their log(fold change, FC) and corresponding standard error, SE:

$$Z_{r,III} = \log(FC_{r,III})/SE_{r,III} ; Z_{r,IV} = \log(FC_{r,IV})/SE_{r,IV}$$

Under the joint null hypothesis, $\log(FC_{r,III}) = \log(FC_{r,IV}) = 0$, the two z-scores are $N(0,1)$ distributed and independent, so the sum $Z_{r,III} + Z_{r,IV}$ is $N(0,2)$. The p-values corresponding to the joint null hypothesis were adjusted for multiple testing using the Benjamini-Hochberg method (Benjamini & Hochberg, 1995).

3.2.2. Pathway prediction.

Miranda, Pictar and Targetscan were used to predict targets for differentially expressed microRNAs from the differentially expressed mRNAs using the RmiR package (Krek et al., 2005; B.P. Lewis et al., 2005; Miranda et al., 2006; RmiR, Favero). Targets were only considered if they were present in at least two of these databases. The resulting targets were entered into the pathway analysis program Metacore[®] (Thomson Reuters). This program performs pathway enrichment analysis generating a p-value, which represents the probability to randomly obtain the intersection between the microRNAs inputted and the pathway genes following hypergeometric distribution.

3.2.3. Analysis of the differentiation pathway.

I used data published in Letzen *et al*, which describes the microRNA expression fold changes between each cell differentiation stage within the OP differentiation pathway including embryonic stem cells (ESCs), neural embryoid bodies (EB), neural progenitors (NP), glial restricted precursors (GP), oligodendrocyte precursors (OP) I, OP II, OP III and the oligodendrocyte lineage (OL) (Letzen et al., 2010). Spearman's correlation (a correlation indicating a relationship, not necessarily linear) was performed on the fold change between good and poor prognosis groups within GIIIA and glioblastoma, with the expression changes of the significantly differentially expressed microRNAs with at least 2-fold change at each stage in the OP differentiation pathway. The fold changes of all microRNAs of significance between OP cell types were used, regardless of their significance for survival. As a control, expression values from Taqman PCR microRNA expression between ESCs and haematological precursors (HP) as described in Risueño *et*

al. and between neural stem cells (NSCs) and NPs as described in Goff *et al.* were used to calculate the Δ Ct and perform Spearman's correlation with the prognosis-associated fold differences in GIIIA and glioblastoma (Goff *et al.*, 2009; Risueño *et al.*, 2012;). Only microRNAs significantly differentially expressed between the ESCs and HPs were used (139 microRNAs) in the correlation analysis, to prevent a false correlation coefficient due to 'baseline' microRNA expression.

3.2.4. Correlation of microRNA expression of the OP pathway with malignant glioma tumours.

Microarray expression data was processed as described in Letzen *et al.* (Letzen *et al.*, 2010) using Agilent Feature Extraction software and the gTotalGeneSignal was correlated with the level 3 expression data from the TCGA GIIIA astrocytomas ($n=39$), glioblastoma tumours ($n=558$) and non-tumour samples ($n=10$). Only microRNAs detected on all platforms (Agilent microarray G4470C and custom TCGA Agilent microarray, and Illumina HiSeq sequencing) were included resulting in 150 microRNAs. Glioblastomas were classified according to Brennan *et al.* (Brennan *et al.*, 2013). The correlation pattern of each cell type for every tumour was analysed for association with survival using Cox regression and log-rank tests.

3.3 Results

3.3.1 Identification of a high-grade glioma microRNA signature associated with poor patient survival

To investigate candidate prognostic microRNAs that are associated with high-grade brain tumours (GIIIA and glioblastoma) through a differential TCGA microRNA expression analysis, I developed the computational pipeline shown in Figure 3.2.

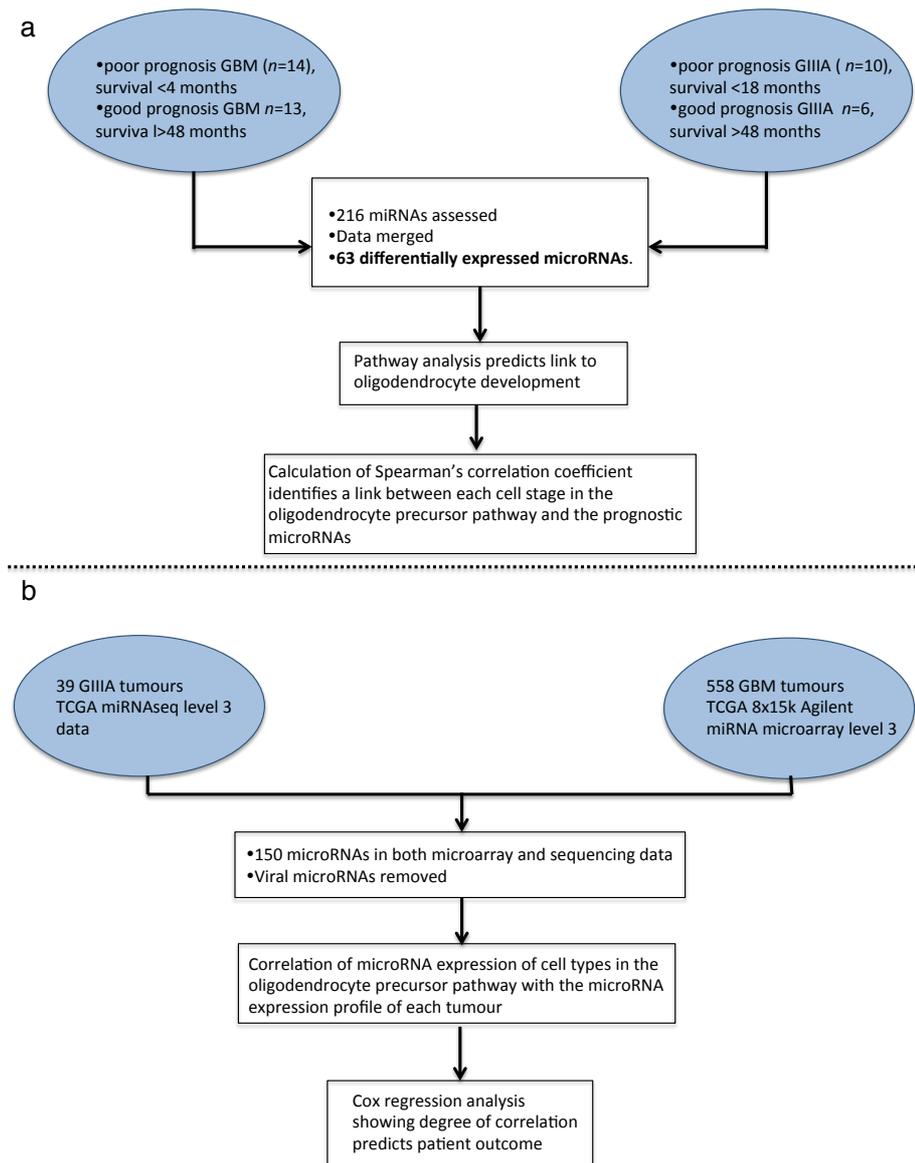


Figure 3.2. The computational analysis pipeline to identify common prognostic molecular signatures in high-grade astrocytoma.

(A) Prognostic microRNAs were identified separately in GIII A and glioblastoma and merged to create a common high-grade microRNA profile associated with prognosis. Predicted pathway analysis suggests that gene expression pathways associated with OP cells may predict patient outcome. Fold change data for differentially expressed microRNAs between cell types in the OP differentiation pathway were correlated with microRNA fold change data calculated between prognosis groups in GIII A and glioblastoma. (B) MicroRNA expression profiles for all 597 TCGA malignant glioma (GIII A and glioblastoma) were correlated with the expression values of each cell type in the OP differentiation pathway (Letzen et al., 2010).

Based on TCGA patient survival data (TCGA, NIH), I defined suitable filter criteria indicative of good prognosis (>48 months for GliIIA and glioblastoma) and poor prognosis (<10 months for GliIIA and <4 months for glioblastoma). These cut-offs were decided by determination of the 10% shortest and the 10% longest survival in the TCGA cohort and including patients with sufficient clinical and microRNA data. This yielded a total of 534 mature microRNAs from 27 glioblastoma and 16 GliIIA tumours, respectively (Fig. 3.2, Table 3.1). Based on this dataset, I first determined the microRNAs that are differentially expressed between the good and poor prognosis groups within glioblastoma and GliIIA specimens, separately. To minimise the false discovery rate, I used EdgeR and LIMMA including multiple testing correction procedures for next generation sequencing and microarray analysis respectively (Hochberg, 1995; Smyth, 2005; Benjamini & Robinson et al., 2009). My approach identified 11 microRNAs that are significantly differentially expressed (with log fold changes between -1.27 and 6.39) in good versus poor prognosis groups in glioblastoma, and 19 in GliIIA (with log fold changes between -1.28 and 2.20). Two of the 19 GliIIA microRNAs were lower in the poor prognosis GliIIA group (Fig. 3.3 A) and 5 of the 11 candidate glioblastoma microRNAs were lower in the poor prognosis glioblastoma group (Fig. 3.3 B). The most strongly (>5 fold) altered microRNAs (miR-10a, miR-196b, miR-211) were all increased within the poor prognosis GliIIA group. This is in line with previous data for miR-10a and miR-211 suggesting their implication in progression and treatment resistance in malignant glioma (Ujifuku et al., 2010).

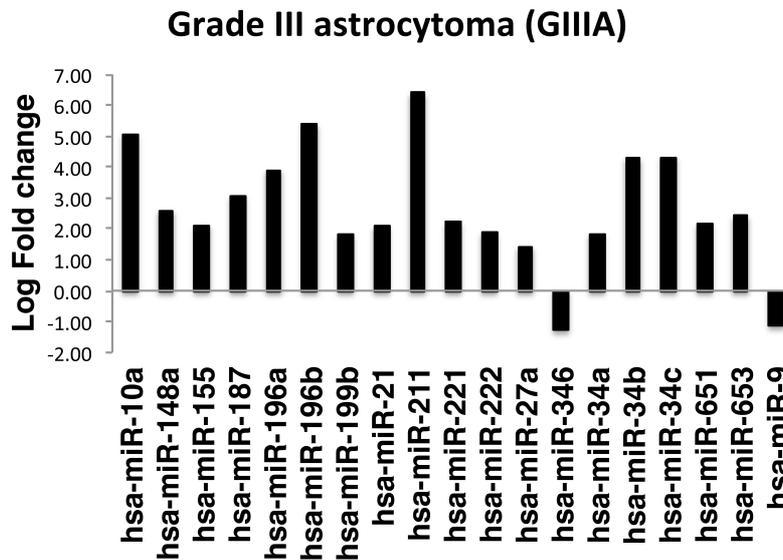
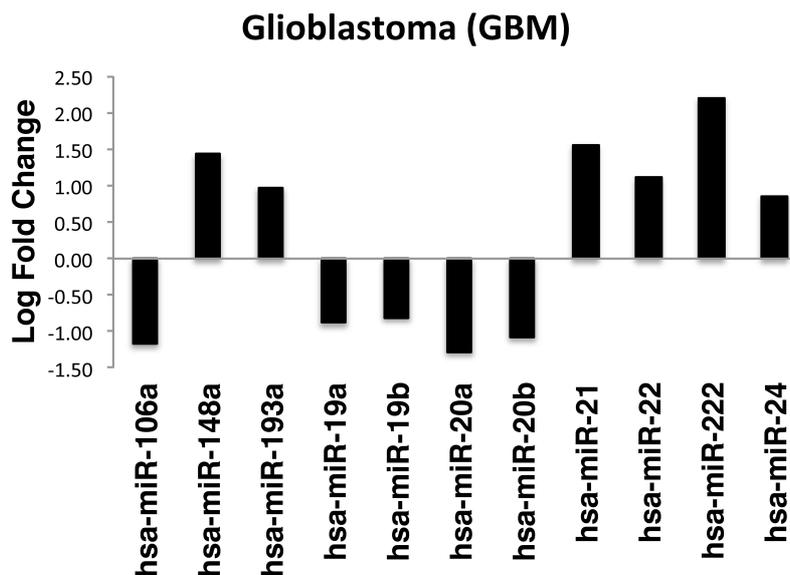
A**B**

Figure 3.3. Fold changes of the differentially expressed microRNA expression between the good and poor prognosis groups in GIIIA (A) and glioblastoma (B).

The majority of the differentially expressed microRNAs in GIIIA are increased with poorer prognosis, whereas glioblastoma shows a more even spread of increased and decreased microRNAs in poorer prognosis patients. The Y-axis refers to the log fold change in the poor prognosis subgroup, compared to the good prognosis subgroup.

Overall, my intra-grade glioma microRNA comparison of good and poor prognosis only yielded three microRNAs; the oncomiR miR-21, the apoptosis regulator miR-148a, and the tumour suppressor regulator miR-222 that could serve as candidate predictors of poor prognosis in both glioblastoma and GIIIA (Zhu et al., 2008; Quintavalle et al., 2012b; Kim et al., 2014). This low overlap between glioblastoma and GIIIA candidate prognostic microRNAs raises the question as to whether it is possible to identify a common microRNA signature for high-grade glioma, or whether the statistical power of the intra-grade comparison approach is insufficient to reveal a glioblastoma/GIIIA poor prognosis signature. To address this question and to increase statistical power in my differential microRNA expression analysis, I combined the z-values ($Z_{r,combined}$) from the good and poor prognosis groups of GIIIA ($Z_{r,III}$) and glioblastoma ($Z_{r,IV}$) accounting for differences in microRNA expression profiling platforms using a computational algorithm based on the z-score merging performed by Stouffer (Stouffer, 1949). This was suitable because Stouffer showed that division by standard error generates a value most similar to a merged z-score. The probability that the score could be less than the z-score was then determined using the pnorm function and 1 minus this value generates a p-value. The resulting p-value was then multiplied by 2 for a two-sided test.

Z-score merging used the formula for each microRNA, r , including fold change, FC, and standard error (SE):

$$Z_{r,III} = \log(FC_{r,III})/SE_{r,III} \quad Z_{r,IV} = \log(FC_{r,IV})/SE_{r,IV}$$

$$Z_{r,combined} = (Z_{r,III} + Z_{r,IV})/\sqrt{2}$$

Under the null hypothesis that $Z_{r,III}$ and $Z_{r,IV}$ are both $N(0,1)$ and independent, $Z_{r,combined}$ will also be $N(0,1)$ and can therefore be interpreted as a z-value. This approach yielded a pool of 216 microRNAs whose differential expression was analysed across all relevant poor/good prognosis glioblastoma and GIIIA TCGA specimens, thereby creating z-scores and p-values for the individual microRNAs. Using the multiple testing corrected p-values for each microRNA yielded 63 microRNAs that significantly change expression between good and

poor prognosis high-grade gliomas as indicated by a >2 fold change of standard deviations from the mean microRNA fold change (FDR<0.05) (Fig. 3.4 A).

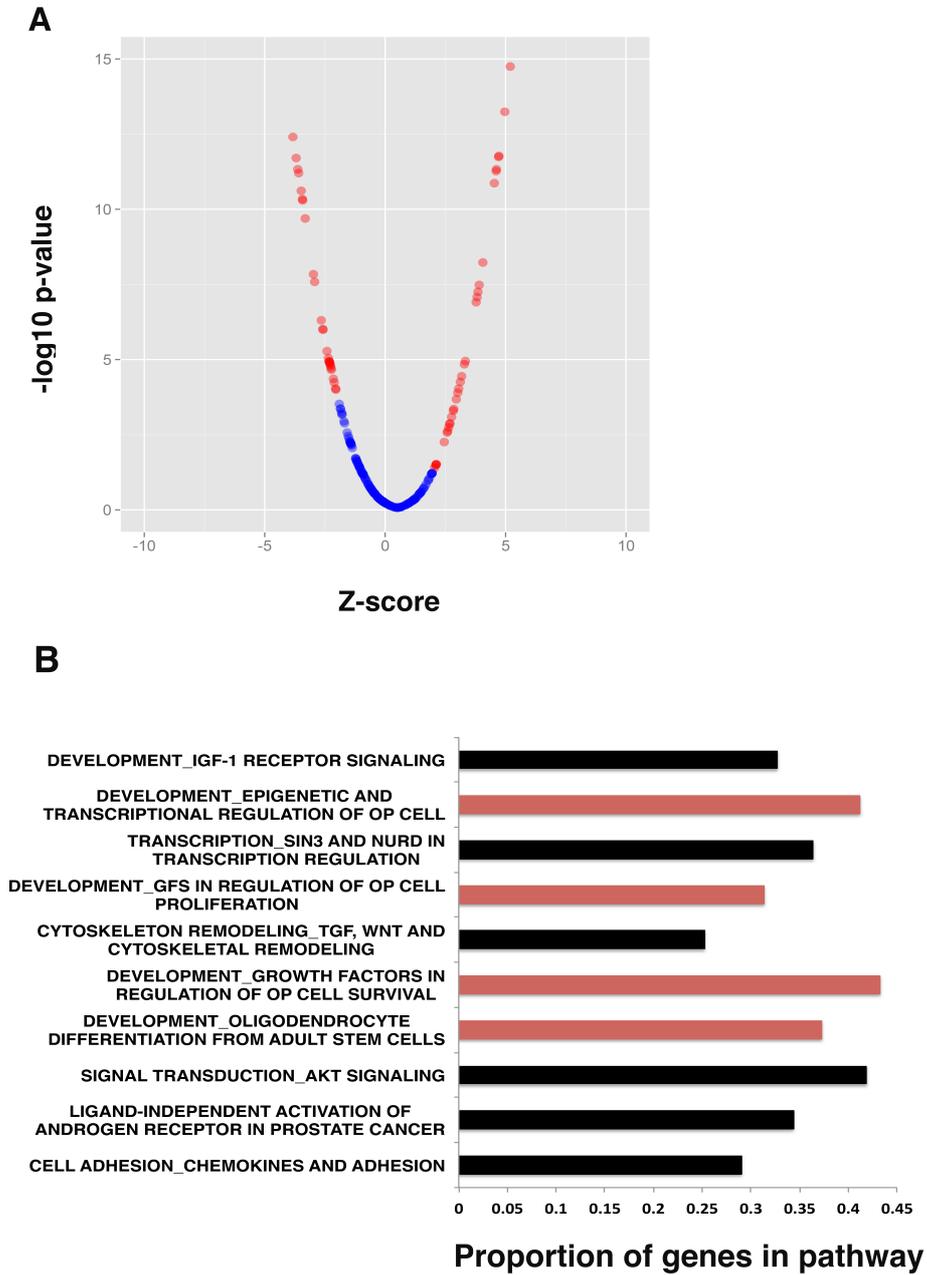


Figure 3.4. Differentially expressed microRNAs in good and poor prognosis groups of glioma point to OP-related pathways.

(A) Plot of the microRNAs differentially expressed between good and poor prognosis groups when data from glioblastoma and GliIIA are combined. 63 microRNAs are significantly altered between good and poor prognosis groups ($p < 0.05$, in red) and have a z-value of at least 2/-2. (B) The targets of the 63

microRNAs associated with patient outcome were predicted and pathway analysis revealed a significant enrichment of genes involved in several OP-related pathways.

This result suggests that a pool of 63 microRNAs form part of a molecular network that is associated with and/or drives aggressive clinical behaviour in high-grade gliomas. To identify the molecular pathways that are likely regulated by the 63 candidate prognostic microRNAs, I predicted their mRNA targets using standard bioinformatic approaches. In order to focus on the mRNA targets that are involved in prognosis, I first enriched for those that are associated with either good or poor prognosis. I compared good prognosis and poor prognosis mRNAs in Glioma and glioblastoma (Table 3.1) using the same criteria as those described above for microRNA analysis. The mRNA data (z-scores and p-values) for Glioma and glioblastoma were merged resulting in 4259 mRNAs with significant ($p < 0.05$) > 2-fold changes. The targets of the 63 microRNAs associated with patient outcome were predicted from the 4259 mRNAs using the target prediction databases Miranda, Pictar and Targetscan (Krek et al., 2005; B.P. Lewis et al., 2005; Miranda et al., 2006). In order to improve target prediction and reduce false positives, I only used targets that were present in at least two of these databases, resulting in 1618 predicted targets for the microRNAs. Subsequently, I entered these mRNAs into the Metacore software and carried out a pathway analysis revealing significant enrichment of genes involved in several cancer-related pathways (Fig. 3.3 B). These pathways included IGF and AKT (V-Akt murine thymoma viral oncogene) signalling, epigenetic and transcriptional regulation, growth factor, androgen and chemokine- effectors and cytoskeletal remodelling. Interestingly, four of these pathways are linked with OP cell fate decisions such as survival, proliferation, differentiation, and myelination. This provides correlative evidence to suggest that the microRNAs associated with survival in high-grade glioma have roles in OP differentiation pathways.

3.3.2. Determination of the role of OP gene expression in prognosis of glioma.

To further determine whether the activity of microRNAs in different cell stages of the OP differentiation pathway are associated with malignant glioma patient

outcome, I accessed published data describing microRNA profiles associated with stages in the differentiation of ESCs into oligodendrocytes (Letzen et al., 2010). My initial hypothesis was that presence of less differentiated oligodendrocyte cells in glioma confers a poorer prognosis. To this end, I questioned whether microRNA expression changes throughout OP differentiation resemble the prognostic microRNA expression pattern of malignant glioma. First, I calculated fold changes between each progenitor cell type in the OP differentiation pathway and correlated these with the fold differences between poor prognosis and good prognosis samples of GIIIA or glioblastoma (Fig. 3.2 A-B). Only microRNAs that are significantly differentially expressed between each stage of the OP differentiation pathway and with at least a 2-fold change in expression were used. The OP2 to OP3 stage was omitted, as there were too few differentially expressed microRNAs between these cell types, suggesting that these two cell stages don't have a significantly altered microRNA expression pattern. In GIIIA, the microRNA expression differences between good and poor prognostic cases correlated directly with the changes associated with differentiation from NP to GP (correlation coefficient = 0.50, $p < 0.05$), which was not evident in glioblastomas. In both grades, the expression differences between good and poor prognosis showed a negative correlation with the changes associated with differentiation from OP1 to OP2 (correlation coefficient -0.54 for GIIIA and -0.47 for glioblastoma, $p < 0.05$) (Fig. 3.5).

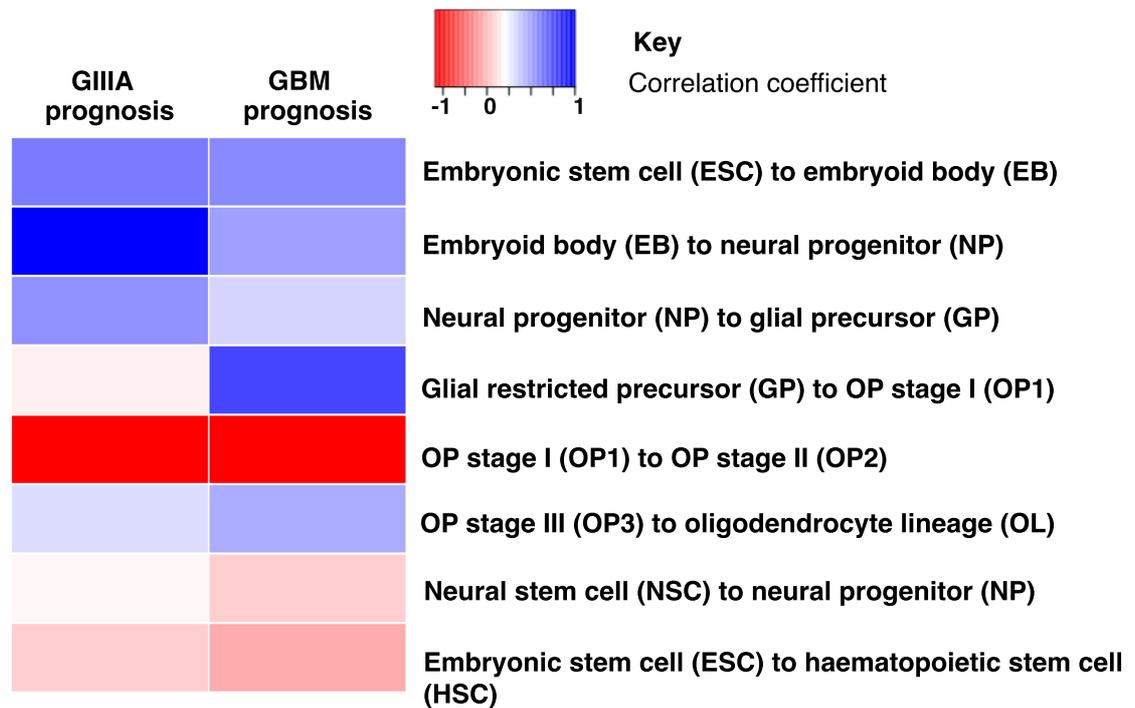


Figure 3.5. Correlation coefficients comparing the fold change of microRNA expression between each stage in the OP pathway and the GIII A and glioblastoma good and poor prognosis groups.

This heatmap represents coefficients from correlation tests between *in vitro* microRNA expression data and the prognostic microRNA expression data. The top 6 rows are data from correlations with Letzen et al. The bottom rows refer to correlations with data from Goff et al. and Risueño et al (Goff et al., 2009; Risueño et al., 2012). Significant correlations (with p -value generated using $p=r/\text{Sqrt}(r^2)/(N-2)$) were between neural progenitor and glial restricted precursors in grade III data ($p=0.009$) and OP1 to OP2 in both grade III and GBM ($p=2.7e^{-3}$ and $p=4.2e^{-4}$ respectively). The highest negative correlation is the transition from OP1 to OP2 and the highest significant positive correlation is the transition from GP to OP1 in glioblastoma.

Next, I tested whether these correlations are a result of non-specific correlations with any ESC differentiation pathway (including non-neural lineages), or whether these high correlations are specific for neural differentiation. I used expression data from a study comparing microRNA expression between ESC cells and hematopoietic progenitors (HPs) and between neural stem cells (NSCs) and neural progenitors (NPs) (Goff et al., 2009, Risueño et al., 2012;). I

correlated the differences in the differentially expressed microRNAs between ESCs and HPs and NSC and NP cells with the differential microRNA expression patterns between good and poor prognosis in GIIIA and glioblastoma. This approach revealed no significant correlations ($p > 0.05$) indicating that the microRNA expression differences between good and poor prognosis of malignant glioma are specifically correlated with microRNA expression changes in OP differentiation, and not with other differentiation pathways (Fig. 3.5).

A notable difference in good and poor prognosis GIIIA patients studied here was their *IDH* mutation status (Table 3.1), which is used to classify patients clinically; those with the mutation are usually proneural tumours and have a favourable prognosis (Riemenschneider et al., 2010; Killela et al., 2013). In my cohort, all the good prognosis patients had an *IDH* mutation, while only one poor prognosis patient's tumour was *IDH* mutated. It could be argued that the difference in microRNA expression between these two groups is simply due to different biology associated with the presence or absence of an *IDH* mutation. To test this possibility, I obtained sequencing data for *IDH* mutated (*IDHmut*, $n=139$) and *IDH* wild-type (*IDHwt*, $n=39$) gliomas (combining data for both grade II and III glioma for added statistical power as it is simply the mutation under analysis and not the tumour context) from the TCGA and determined microRNA expression differences between the two groups using the criteria previously stated. The microRNA expression fold differences between *IDHwt* and *IDHmut* were correlated with the fold changes between each stage in the OP differentiation pathway. The only significant correlation observed was an inverse correlation between *IDHmut* and *IDHwt* and OP1 to OP2. Good versus poor prognosis GIIIA and glioblastoma also correlated with this OP differentiation stage. Critically, the fold differences between *IDHmut* and *IDHwt* did not correlate with the changes during differentiation from NP to GP ($p < 0.05$, correlation coefficient < -0.34). Therefore I conclude that the correlation I have shown between prognosis and OP stage differentiation is independent of *IDH* mutation status.

3.3.3. Correlation of glioma tumour microRNA expression with the OP cell stage.

Correlations of the microRNA expression differences between good and poor prognosis cases and between neural differentiation stages imply that correlation with the OP1 cell type is most closely related to prognosis.

In order to examine this hypothesis, I correlated the microRNA expression profiles of each differentiation stage/cell type in the oligodendrocyte differentiation pathway with microRNA expression profiles of 39 GIIIA, 558 glioblastoma and 10 non-tumour samples from the TCGA. Only microRNAs present in all platforms (sequencing and microarray) were used (150 microRNAs) (Appendix 3.1). The majority of the 597 tumours positively correlated with each cell type in the OP differentiation pathway. Seven GIIIA tumours did not correlate with OP2 or OP3 microRNA expression, and two GIIIA tumours did not correlate with OP2 expression. Across all tumours assessed, the highest correlations were with OP1 and OL cell types microRNA expression patterns (Fig. 3.6 A-B). The cell type whose microRNA expression correlated most positively with tumour microRNA expression was OP1. MicroRNA expression of this cell type was most correlated with expression of each tumour type and also ten non-tumour samples.

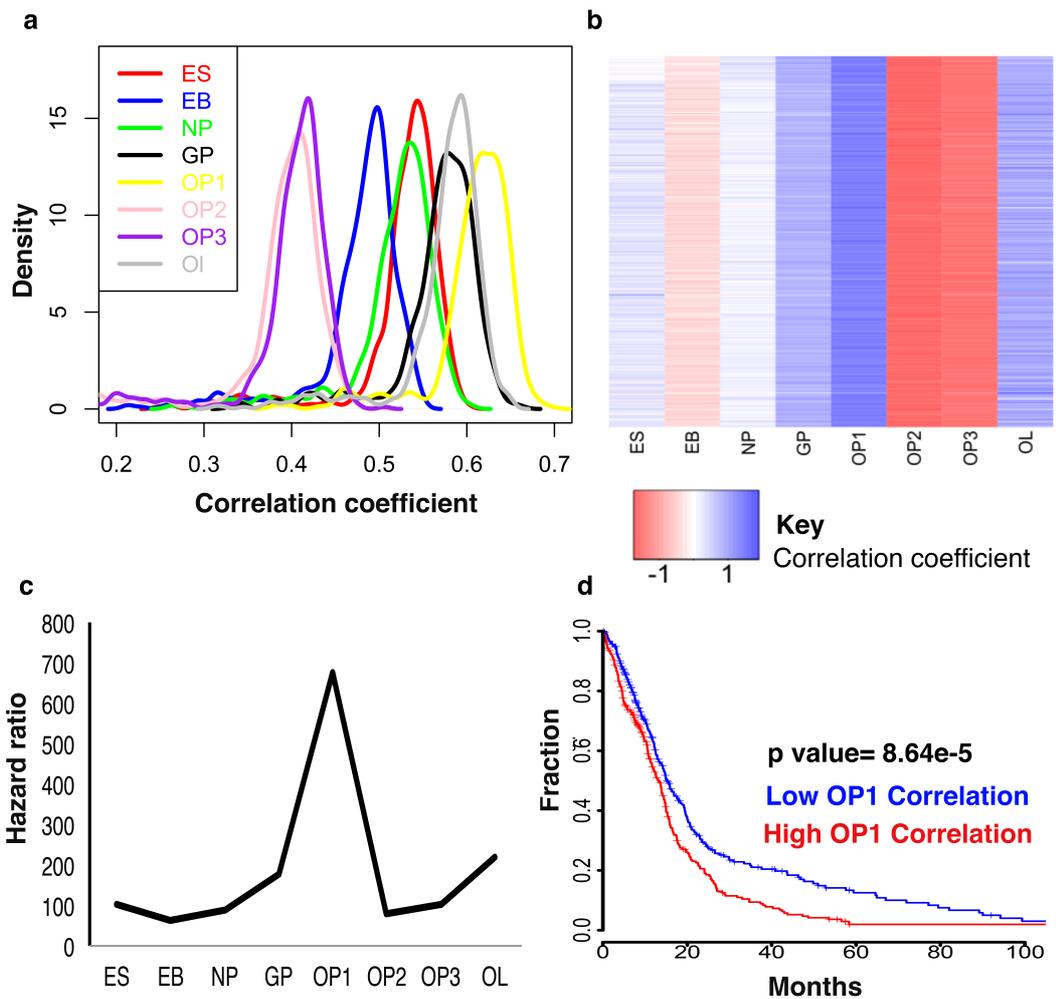


Figure 3.6. The correlation of microRNA expression between each cell type with glioma tumours in the TCGA.

(A) Plot of density (Y-axis) of the Spearman's correlation coefficients (X-axis) for each cell type with all GliIIA and glioblastoma tumours in the TCGA. All cell types in the OP differentiation pathway (from Letzen et al, data obtained through personal communication) show significant positive correlation of microRNA expression with each tumour, with the oligodendrocyte lineage and OP1 cells showing the highest positive correlations. (B) Heatmap of correlation of each GliIIA/glioblastoma tumour with each OP cell type. (C) Hazard ratios from Cox regression analysis of the correlation patterns of each cell type shows that OP1 microRNA expression correlation is the most predictive in terms of prognosis. MicroRNA profiles of all cell types were significantly associated with survival ($p < 0.05$). (D) Kaplan-Meier plot of the OP1 correlation coefficients for

grade III and IV gliomas. Groups are separated above and below the median correlation of microRNA expression between OP1 and tumour.

Twenty-three glioblastomas had the highest correlation with OL and one glioblastoma had the highest correlation with glial restricted precursor (GP). Average correlation with OP1 was 0.60 for GIIIA, 0.93 for glioblastoma, 0.69 for the mesenchymal subtype of glioblastoma ($n=155$), 0.67 for classical glioblastoma ($n=143$), 0.36 for G-CIMP glioblastoma ($n=38$), 0.54 for neural glioblastoma ($n=82$), 0.58 for proneural glioblastoma ($n=97$) (Brennan et al., 2013)¹, and 0.88 for non-tumour samples. These results indicate that glioblastoma is most positively correlated with OP expression patterns. GIIIA alone has lower correlations with OP cells than non-tumour samples.

To determine the association of the OP differentiation cell stages with high-grade glioma patient survival, the correlation values for each of the eight cell types in the OP differentiation pathway with all 597 tumours in the TCGA were assessed for association with survival using Cox regression analysis. Rho (ρ) values (Spearman's coefficient) for all cell types were significant negative predictors ($p < 0.05$) of survival. The highest hazard ratio was for correlations with the OP1 cell type (Fig. 3.5 C), which indicates that gliomas with microRNA expression patterns similar to OP1 cells have a poorer patient outcome (Cox regression HR = 13.02, 95% CI = 3.77-45.04, $p = 5.02e^{-05}$) (Fig. 3.5 D). Taken together, my results suggest that the most aggressive malignant gliomas (both GIIIA and glioblastoma) have a microRNA expression pattern that aligns with expression patterns characteristic of the OP1 cell stage.

3.4. Discussion.

3.4.1. Prognostic glioma microRNAs align with OP pathways.

There has been considerable discussion over subtyping of glioblastoma based on expression and copy number data. However, so far this approach has not delivered robust clinical biomarkers and the field is further complicated by data confirming that subtypes can co-exist within the same tumour thereby creating a diversity of oncogenic transcriptional programs that contribute to treatment resistance (Sottoriva et al., 2013; Patel et al., 2014).

Models of glioma suggest these tumours may be defined by the initiating cell type or the type of initiating mutation (Lei et al., 2011). Despite these

observations, the glial cell of origin in different histological types of glioma remains unclear (Sukhdeo et al., 2011). It has been proposed that OPs may fill this role in some subtypes and this is supported by data suggesting that mesenchymal glioblastoma can arise from a proneural-like precursor (Liu et al., 2011a; Ozawa et al., 2014).

Using integrated mRNA and microRNA expression data I have identified that prognostic microRNA expression patterns in malignant glioma correlate with microRNA expression changes during oligodendrocyte differentiation. This study is novel in identifying grade-independent and subtype-independent prognosis prediction using microRNAs as biomarkers, which are stable in clinical samples and may be appropriate for implementation into clinical practice (Hall et al., 2012).

MicroRNA expression changes associated with cellular transitions between OP1 and OP2 suggest that more aggressive tumours have more cells with OP1-like expression patterns. Whether these are non-malignant OP1s present within the tumour mass, malignant cells with similarities to these cells, or are simply less differentiated, cannot be ascertained from my data. The tumour samples under study here were defined by the TCGA as having at least 70% tumour nuclei which suggests this is unlikely to be a non-malignant population of cells.

In line with my computational results, OPs have been shown to stimulate a more aggressive phenotype by promoting neo-vascularisation of glioma and are present at the invasive front of high grade tumours (Huang et al., 2014). Initial neoplasia-generating aberrations in NSCs can only become transforming upon differentiation into an OP, suggesting that these cell types are important in tumour initiation, as well as defining its behaviour (Sukhdeo et al., 2011). Supporting this notion, both proneural and mesenchymal glioblastoma tumours have been shown to arise from a common precursor (Ozawa et al., 2014). OP cells are also implicated in maintaining self-renewal by means of asymmetric cell division, supporting both self-renewal and proliferation in the tumour (Sugiarto et al., 2011). OP cells are also defined by their PDGFRA expression, and recent studies show that amplification of this is an initiating event in gliomagenesis (Ozawa et al., 2014; Zhang et al., 2014; Havrda et al., 2014).

Oligodendrocyte precursor cells are highly migratory, and this is mediated by different cues at different regions in the CNS (Jarjour and Kennedy, 2004). Substrates such as laminin promote OPC migration, whereas collagen inhibits mobility (Milner et al., 1996). Enhanced migratory capacity may be an explanation for the poorer prognosis in patients exhibiting microRNA patterns similar to GPs and OPCs. The biological significance of these results may suggest that it is a particular stem-like cell that is important in patient outcome, and detection of these cell types could be of importance in assessment of the level of aggressiveness of certain cancers.

3.4.2. Translational relevance of the OP1 prognostic signature.

My results suggest a more OP1-like phenotype is associated with a more aggressive tumour, and presence of these OP1 microRNA patterns predicts poorer prognosis. MicroRNA signatures predicting more aggressive tumours have been developed in the past, yet their relevance to tumour biology is not well understood (Lakomy et al., 2011; Srinivasan et al., 2011; Zhang et al., 2012a, Sana et al., 2014). Also, subtype-specific signatures are not easy to implement into the clinical routine of standard healthcare laboratories due to logistic challenges (i.e. multiple testing procedures) and the need for diverse state-of-the-art profiling platforms (i.e. next generation sequencing) as well as high-level bioinformatics/computational support. Hence, it would be desirable to replace complex prognostic signatures with a few key biomarkers wherever possible. For example, Letzen *et al* describe peaks of miR-10a and miR-21 expression in OP cells (Letzen et al., 2010) compared to other neural cell types (these microRNAs were both significantly increased in poor prognosis malignant glioma in my study) and it may therefore be possible that these microRNAs alone have the potential to be exploited as biomarkers for the presence of OP-like cells. Prospective observational clinical trials will be needed to address this hypothesis.

Taken together, I provide preliminary evidence that classification of malignant glioma based on microRNA expression patterns seen in OPs may predict the outcome of the disease, which could not only inform patient management but also guide development of novel treatments. The statistical power of future work

is likely to be increased due to the availability of more samples in TCGA and other repositories. This is also a principle that could be extended to other tumour types, to elucidate the characteristic microRNA profiles exhibited in particular by poor prognosis tumours.

4. Investigation of microRNA-9 in malignant glioma.

'Not just microbes, not just the worm, not just animals but plants...and related to RNAi. This really sets it up as a tiny RNA universe.' Gary Ruvkin, 2008.

4.1. Introduction.

Results from my prognostic signature in glioblastoma shown in Chapter two revealed miR-9 as a prognostic microRNA, with reduced tumour expression levels resulting in poorer prognosis. This suggests that miR-9 may be a tumour suppressor. When targets were predicted for all microRNAs in the prognostic signature in Chapter two, more targets were identified for miR-9 than any other microRNA. Kim *et al.* also reported that miR-9 had the largest correlation network of all microRNAs in glioblastoma and Sun *et al.* named miR-9 as one of four microRNAs in a 'hub' contributing to gliomagenesis (Kim *et al.*, 2011b; Sun *et al.*, 2013). This microRNA is highly expressed in the brain, extremely abundant in glioma, and has been shown to promote neural cell differentiation in conjunction with miR-124a by inhibiting STAT3 (signal transducer and activator of transcription 3) phosphorylation (Sempere *et al.*, 2004; Krichevsky *et al.*, 2006; X. Tan *et al.*, 2012). This is likely to occur through the JAK (janus kinase) family of proteins, which are all targeted by miR-9 and phosphorylate STAT3 (Kim *et al.*, 2011b). These results support the notion that miR-9 may be an important microRNA in the biology of glioma.

Despite the finding that miR-9 has a lower expression level in tumours from patients who have a worse outcome, several studies report that miR-9 may be oncogenic in nature. Inhibition of miR-9 was indicated to reduce migration and induce sensitivity to temozolomide, as well as hinder neurosphere formation and promote differentiation (Schraivogel *et al.*, 2011; X. Tan *et al.*, 2012; Munoz *et al.*, 2013; Munoz *et al.*, 2014).

In support of the survival analysis results, it was reported that EGFRvIII can suppress miR-9 which causes up-regulation of its target FOXP1 (forkhead box P1) resulting in increased tumour growth (Gomez *et al.*, 2014). The developmental taxonomy of glioblastoma defined by microRNAs showed miR-9

to be expressed at a higher level in the oligoneural subclass, which is a relatively good prognosis group of tumours (Kim et al., 2011b).

An individual microRNA has the ability to target at least 200 genes (Yue et al., 2009), and the effects of a particular microRNA may be different depending on the cellular context. For example, miR-9 only has the ability to affect neural cell lineage differentiation when in the presence of miR-124a (Krichevsky et al., 2006). The fact that a microRNA has multiple targets can aid its research, as pathway enrichment may be determined using these targets, which I performed in Chapter three. However, study of this number of targets in a heterogeneous disease such as glioblastoma, with many cellular contexts in one tumour, can prove problematic. This may be the cause of the conflicting results for miR-9. The aim of Chapter four was to clarify why tumours from patients with poorer prognosis have lower expression levels of miR-9 yet miR-9 confers an advantage to tumour cells when overexpressed.

In this study, I have combined results from patient samples, which reflect the overall expression levels of miR-9 in a tumour, with the expression levels of miR-9 in cultured glioma stem cells. I have interpreted the results in light of the current molecular classifications for both glioblastoma tumours (Verhaak et al., 2010; Brennan et al., 2013) and glioma stem cell lines (Mao et al., 2013). These results suggest that miR-9 is expressed at lower levels in the mesenchymal group of tumours and the phenotypic characteristics of miR-9 correlate with hallmarks of this glioblastoma subtype, such as infiltration and necrosis. With further validation miR-9 could be used as a biomarker associating with the mesenchymal subtype, which would eliminate the need for multiple markers to be tested.

4.2. Methods.

4.2.1. Cell culture and transfection.

Glioblastoma stem-like cell lines G33, G35, X6, 157, 528 and G44 were obtained from freshly resected malignant glioma samples (from Ichiro Nakano, Ohio State University with patient consent and permissions, Appendix 4.1 (Mao et al., 2013)). These were cultured at 37°C in 5% CO₂ in stem cell media consisting of neurobasal/Glutamax medium (Life Technologies, Carlsbad, CA) supplemented with 2% B27 (Life Technologies), which is designed to support

neural cell culture, and 20ng/ml epidermal growth factor and 20ng/ml fibroblast growth factor (PeproTech, Rocky Hill, NJ). Cells were grown as adherent monolayers in laminin-coated dishes (Invitrogen) or as neurospheres in 75cm² ultra-low attachment flasks (Corning, NY).

U251, U87, U373 and LN229 adult glioblastoma cells (American Type Culture Collection, ATCC), and KNS42, Res186, UW479 and SF188 paediatric cell lines (obtained from Chris Jones at The Institute of Cancer Research, Surrey under a material transfer agreement (Gaspar et al., 2010)) were cultured in RPMI 160 medium (Life Technologies) containing 10% foetal bovine serum (Sigma-Aldrich, St Louis, MO) at 37°C in 5% CO₂ in 75cm² flasks (Corning). GBM-1 and GBM-4 (obtained from freshly resected malignant glioma samples as recorded in Wurdak *et al* (Wurdak et al., 2010) and transferred under a material transfer agreement) were cultured at 37°C in 5% CO₂ in stem cell media consisting of neurobasal/Glutamax medium (Life Technologies) supplemented with 2% B27 (Life Technologies) and 1% N-2 (Life Technologies), both designed to support neural cell culture. 20ng/ml epidermal growth factor and 20ng/ml fibroblast growth factor (PeproTech) were also added to the media. Cells were grown as adherent monolayers in laminin-coated 25cm² flasks (Life Technologies).

Cells were transfected with 100nM miR-9 mimic or scrambled control oligonucleotides (GE Healthcare Dharmacon Inc, Lafayette, CO), using 5µl of lipofectamine 2000 (Life Technologies) per 2.5ml of transfection mix in six-well plates containing 200,000 cells/well and were assayed after 48 hours. Plasmid transfections were performed in the same way with 10µl lipofectamine 2000 and 4µg of plasmid. Human foetal neural stem cells (SCP2743) obtained from Columbus Nationwide Children's Hospital were grown as monolayer cultures by Dr. Choi-Fong Cho (Harvard Medical School).

Hypoxic cultures were performed at 1% oxygen rather than 21% (atmospheric oxygen level) by adjustment of the nitrogen levels in the incubator. The incubator was not opened during the 72-hour incubation period and RNA extraction was performed immediately after harvesting the cells. Hypoxia response was confirmed by quantifying mRNA levels of a HIF1A target, GLUT1 by qRT-PCR.

4.2.2. RNA extraction and quantitative real-time PCR.

For mRNA quantification RNA was extracted using TRIzol[®] Reagent (Life Technologies) according to the manufacturer's guidelines 48 hours post-transfection and first-strand synthesis performed using SuperScript[®] II Reverse Transcriptase (Life Technologies). Quantitative PCR (qPCR) analyses were performed in duplicate with SYBR[®] Select Master mix for CFX (Applied Biosystems) on a BIO-RAD CFX Connect[™] real-time PCR detection system. Primers were manually designed for mRNAs and ordered from Integrated DNA technologies (IDT, San Jose, CA) or Sigma-Aldrich. The control used for these assays was a primer for 18S ribosomal RNA, which has been shown to have stable and abundant expression across treatments and samples by Life Technologies and Selvey *et al* and is often used in RT-PCR studies for glioblastoma (Selvey et al., 2001; Fassl et al., 2012; Ozawa et al., 2010). Primer sequences are included in Appendix 4.2.

For microRNA quantification, RNA was extracted using TRIzol[®] Reagent (Life Technologies) according to the manufacturer's guidelines 48-72 hours post-transfection and reverse transcription was performed using Taqman MicroRNA Reverse Transcription Kit (Applied Biosystems, ABI, Waltham, MA). Quantitative PCR (qPCR) analyses were performed with Taqman 2x Universal PCR master mix with no UNG (ABI) and measured on a BIO-RAD CFX Connect[™] real-time PCR detection system. Taqman MicroRNA assays (ABI) were used as microRNA primers. The control used for these assays was U6 snRNA as this is considered to be stable across human tissues, both cancerous and non-cancerous (Peltier & Latham, 2008).

4.2.3. RNA extraction and Illumina microRNA Sequencing (sequencing performed by Dr. Sally Harrison, Leeds Teaching Hospitals).

RNA was extracted using miRNeasy (Qiagen, Gaithersburg, MD). Library preparation was performed using the NEBNext[®] Multiple X Small RNA Library Prep Set for Illumina (New England Biolabs Inc, Ipswich, MA) according to manufacturer's guidelines. Illumina multiplex HiSeq sequencing (Illumina, Inc)

was performed according to manufacturer's guidelines on a HiSeq 2000 with 50 base reads and 16 samples per lane.

4.2.4. Cell adhesion assays.

For cell-cell adhesion assays a single glioma stem-like cell suspension of 1 million cells in 3ml stem cell medium was added to a 15ml Falcon tube (Thermo Fisher Scientific Inc, Waltham, MA) and incubated on a 360° rotator at 37°C in 5% CO₂. Every 1-hour 100µl was removed and fixed in 16% paraformaldehyde (PFA, Electron Microscopy Sciences, Hatfield, PA). To visualise the spheres, cells were stained with 1:1000 Hoechst 33342, Trihydrochloride, Trihydrate 10mg/ml solution in water (Life Technologies) and the size and number of the spheres were imaged using a bright-field microscope (DM6000B, Leica, Wetzlar, Germany) connected to a charged-coupled device camera. These images were then quantified by area using the 'Analyze Particles' function of ImageJ 1.48v which applies a mask to the image so all small particles (single cells) are removed from the image.

For cell substrate adhesion assays, established cell lines U251 and U373 were seeded at 5000 cells per well of a 96-well plate in 100µl of serum containing media (see above). The plate was incubated at 37°C in 5% CO₂ and at 4 hours incubation time the non-adherent cells and media were carefully removed from appropriate wells. 16% PFA was added to each well to fix the cells. Cells were stained with 1:1000 Hoechst 33342 for 30 minutes. One field per well at 4x magnification was imaged using a bright-field microscope (DM6000B, Leica) connected to a charged-coupled device camera. The number of cells was then quantified by area using the 'Analyze Particles' function of ImageJ 1.48v.

4.2.5. Cell viability assays.

Cells were assessed for viability using PrestoBlue[®] Cell Viability Reagent (Life Technologies). Glioma stem-like cells were seeded at 5000 cells/well in 90µl stem cell media. 10µl PrestoBlue[®] was added and incubated at room temperature for 10 minutes. Absorbance at 570nm was measured on a POLARStar[®] Omega platereader (Bmg Labtech, Ortenberg, Germany).

4.2.6. Transwell migration assays.

Migration of glioma stem cells (GSCs) was assayed using FluoroBlock™ permeable transwell inserts (Corning®) with 8µm pore sized membranes. Membranes were coated with 600µl 1:100 Corning® Matrix Growth Factor Reduced Matrigel® in DMEM for 3 hours at 37°C in 5% CO₂ and then washed with 200µl phosphate buffered saline (PBS, Life Technologies). 50000 GSC cells in 100µl stem cell medium were added to the insert. Then 600µl stem cell medium was added to the well. Cells were incubated for 6 hours at 37°C in 5% CO₂, and then non-migrating cells in the insert were removed with a PBS wash and cotton swabs. Cells that migrated to the bottom of the membrane were then stained with 600µl crystal violet for 15 minutes at 25°C and washed with water. The inserts were allowed to dry for 3 days. For quantification, the crystal violet was removed using 600µl acetic acid and absorbance at 570nm was measured on a POLARStar® Omega platereader (Bmg Labtech). The level of absorbance provides an indication of the number of cells present on the membrane as they stain with crystal violet.

4.2.7. Western blotting.

For western blotting of targets Invitrogen apparatus was used. Cells were washed in cold PBS, lysed in radioimmunoprecipitation assay (RIPA) buffer which contains 25mM Tris-HCl pH 7.6, 150mM NaCl, 1% NP-40, 1% sodium deoxycholate and 0.1% SDS (Thermo Scientific) with 1x PhosStop phosphatase inhibitors (Roche diagnostics, Basel, Switzerland) and 1x protease inhibitor cocktail set (Calbiochem, Billerica, MA) and sonicated for 5 seconds. 20µg protein were separated on a 10% precast polyacrylamide SDS (sodium dodecyl sulphate) gel (BIORAD, mini protean TGX gel) and blotted onto nitrocellulose membrane (BIORAD, Hercules, CA). Membranes were blocked for one hour under shaking in 5% non-fat milk solution (Lab Scientific, Highlands, NJ) in 1 x Tris-buffered saline with Tween 20 (TBST, 20 mM Tris pH 7.5. 150 mM NaCl. 0.1% Tween 20, BIORAD). After washing in TBST, the membrane was incubated overnight at 4°C under shaking with 10ml primary polyclonal rabbit anti-SHC antibody (Cell Signalling Technology, Danvers, MA) at a dilution of 1:1000 in 5% non-fat milk solution (Lab Scientific). The membrane was again washed and then incubated for 1 hour in 1:10000

peroxidase-conjugated goat anti-rabbit IgG antibody (Cell Signalling Technology). The Thermo Scientific Supersignal West Femto Maximum Sensitivity substrate and the BIORAD Chemidoc were used to develop the signal. The membrane was then washed in TBST and then incubated for 1 hour with anti-GAPDH fluorescent antibody (Abcam, Cambridge UK) and visualised on the BIORAD Chemidoc.

4.2.8. Luciferase reporter assays.

U251 cells were co-transfected in six well plates (Corning Costar) with either scrambled control oligonucleotide or miR-9 mimic (Dharmacon) and pEZX negative control (empty vector) or pEZX-SHC1 3' UTR luciferase construct harbouring the 3' UTR including binding sites for miR-9 (GeneCopoeia, Rockville, MD, HmiT017080, Fig. 4.1). 24 hours later 100µl was transferred to wells of a 96-well white bottom plate (Cellstar, VWR, Radnor, PA) to allow cells to settle in the appropriate volumes for the assay. 24 hours later 100µl Steady-Glo[®] Luciferase reagent (Promega, Madison, WI) was added and incubated at room temperature for 5 minutes to allow lysis. Luminescence was measured using a POLARStar[®] Omega platereader (Bmg Labtech).

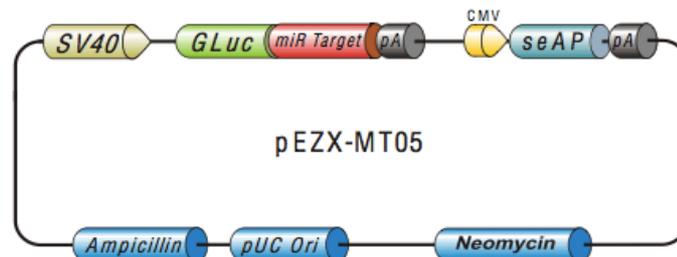


Figure 4.1. The pEZX reporter plasmid with the 3' UTR of SHC1.

In order to assess SHC1 as a target of miR-9, cells were co-transfected with miR-9 mimic and this plasmid containing the SHC1 3' UTR sequence (miR target), or an empty (without the SHC1 3'UTR) pEZX plasmid as a control. Image from Genecopoeia manufacturer's guidelines.

4.2.9. TCGA data, statistical analysis and target prediction.

Level 3 Agilent microRNA 8x15k microarray and G4520A microarray gene expression data plus clinical information for 558 glioblastoma and 10

unmatched non-tumour samples were downloaded from TCGA (TCGA, NIH (accessed October 2012). Illumina HiSeq sequencing data (level 3, reads per million of total reads mapping to a mature microRNA and precursor sequences) for microRNAs were downloaded for all samples with grade II or III glioma from TCGA ($n=178$; 55 astrocytoma, 47 oligoastrocytoma, 75 oligodendroglioma, 1 not stated; 95 grade II, 112 grade III, 1 not stated). Survival analysis was performed in R v2.15.1 using Cox regression for mature sequences and precursor sequences separately. Subgroups of the patient samples were defined using data from Brennan *et al.* and tumour cell percentage and extent of necrosis were obtained from TCGA clinical data files (Brennan *et al.*, 2013). Targets of miR-9 were identified using Miranda (Miranda *et al.*, 2006) and Targetscan (Lewis *et al.*, 2005) irrespective of conservation. Gene ontology analysis was performed using Metacore v6.16 (Thomson Reuters) modified exact Fisher's test. Correlation of microRNA and mRNA expression was performed using Spearman's correlation on all 558 glioblastoma samples.

4.2.10. Sequencing bioinformatics pipeline (performed by Dr. Lucy Stead).

Fastq files were generated using Illumina software and instructions. Adapter sequences were removed and trimmed according to quality to remove low quality reads and adapter contamination. Reads of less than 13 bases were removed. Data was collapsed into unique sequences and their associated counts. Reads were then aligned to miRBase v19 (Griffiths-Jones *et al.*, 2008) precursors (downloaded May 2013). Reads aligning more than once were assigned to the most probable location of origin using SeqEM (Martin *et al.*, 2010). This produced a read count for every precursor in the database. A further file was generated from this, which only included sequences aligning to mature microRNAs defined by MiRBase.

4.3. Results.

4.3.1 Expression of miR-9 in samples with different prognosis, and with different molecular subtypes.

I performed survival analysis for miR-9 in glioblastoma using TCGA microarray microRNA expression data. This showed that in 558 patients miR-9 is expressed at lower levels in tumours from patients with a poorer prognosis

(HR= 0.91, CI=0.85-0.99, p=0.019, Fig. 4.2 A). A similar analysis was performed using the level 3 sequencing data for mature microRNAs for lower grade glioma (encompassing grade II and III) and this also showed that miR-9 was expressed at lower levels in patients with poorer prognosis (log miR-9 HR = 0.28, CI=0.13-0.59, p=9.8e⁻⁴, Fig. 4.2 B).

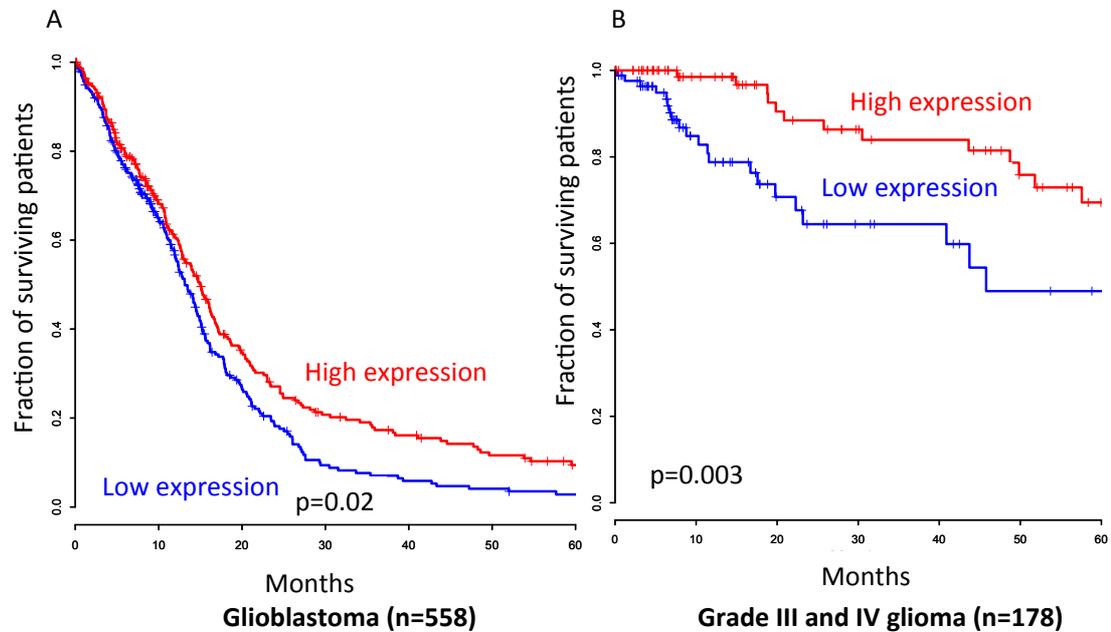


Figure 4.2. In glioma miR-9 is a prognostic microRNA that decreases with poorer patient outcome.

Survival analysis was performed for miR-9 using TCGA microRNA expression data. These are Kaplan-Meier curves showing the association of miR-9 expression levels with patient survival. The Y-axis represents the fraction of patients alive at each time, shown in months on the X-axis. A) In 558 glioblastoma patients, miR-9 was shown to be at lower expression levels in patients with a poorer outcome (separated on the median expression level). B) In 178 grade II and III glioma patients, mature miR-9 expression levels were also lower in tumours of patients with poorer prognosis.

In order to assess whether the expression of miR-9 was altered in different subtypes of glioblastoma, levels were assessed in each subgroup (as defined by Brennan *et al*). This showed that the mesenchymal glioblastoma subgroup had much lower levels of miR-9 compared to the other subgroups (Fig. 4.3 A) and a LIMMA (linear models for microarray data) test comparing miR-9 in proneural TCGA tumours with mesenchymal tumours showed a significant decrease in expression in the mesenchymal subgroup ($p=3.05e^{-11}$, Fig. 4.3 B).

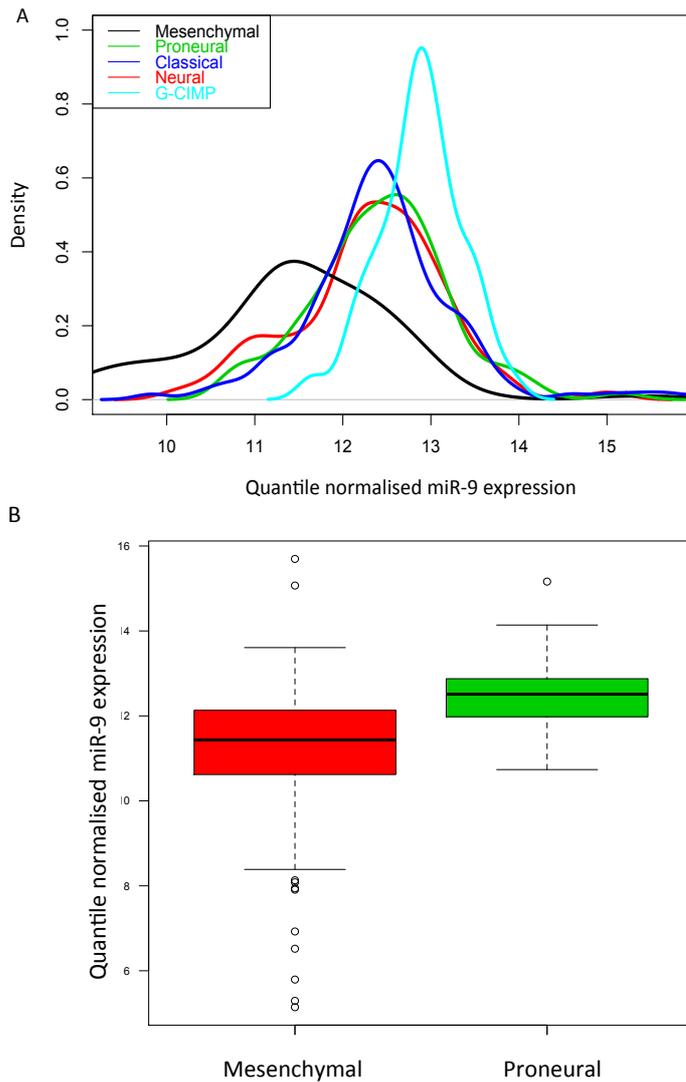


Figure 4.3. miR-9 expression is lower in mesenchymal glioblastoma.

Levels of miR-9 expression were assessed using quantile normalised microarray expression data for molecular subtypes of glioblastoma. A) A density plot of the microRNA expression in all subgroups of glioblastoma. miR-9 is highest in the G-CIMP tumours and lowest in the mesenchymal subtype. B) Comparison of the miR-9 levels in proneural and mesenchymal tumours by LIMMA test showed a significant decrease in the mesenchymal subgroup compared to the proneural group. The Y-axis represents quantile normalised microRNA expression for the two probes of miR-9 on the microarray. The boxes represent the interquartile range and the heavy line within the box is the median expression. The whiskers are the minimum and maximum expression within the nominal range (the upper/lower quartile plus 1.5 times the interquartile distance) and the circles are outliers from this range.

4.3.2. Association of miR-9 precursor expression with prognosis and abundance of mature miR-9 in glioma.

So far I have shown that mature miR-9 expression is prognostic in patients with glioblastoma and lower grade glioma. In order to determine whether the precursor sequences are prognostic I used TCGA sequencing data where reads aligning to precursor sequences have been quantified, to perform survival analysis. I performed survival analysis on the three precursors of miR-9, which showed that all precursors are prognostic (Fig. 4.4 A-C). In-house sequencing of two GSC lines and a non-tumour sample also shows that the levels of the precursors differ between samples, but no one particular precursor is altered in the same way across samples (Fig. 4.4 D).

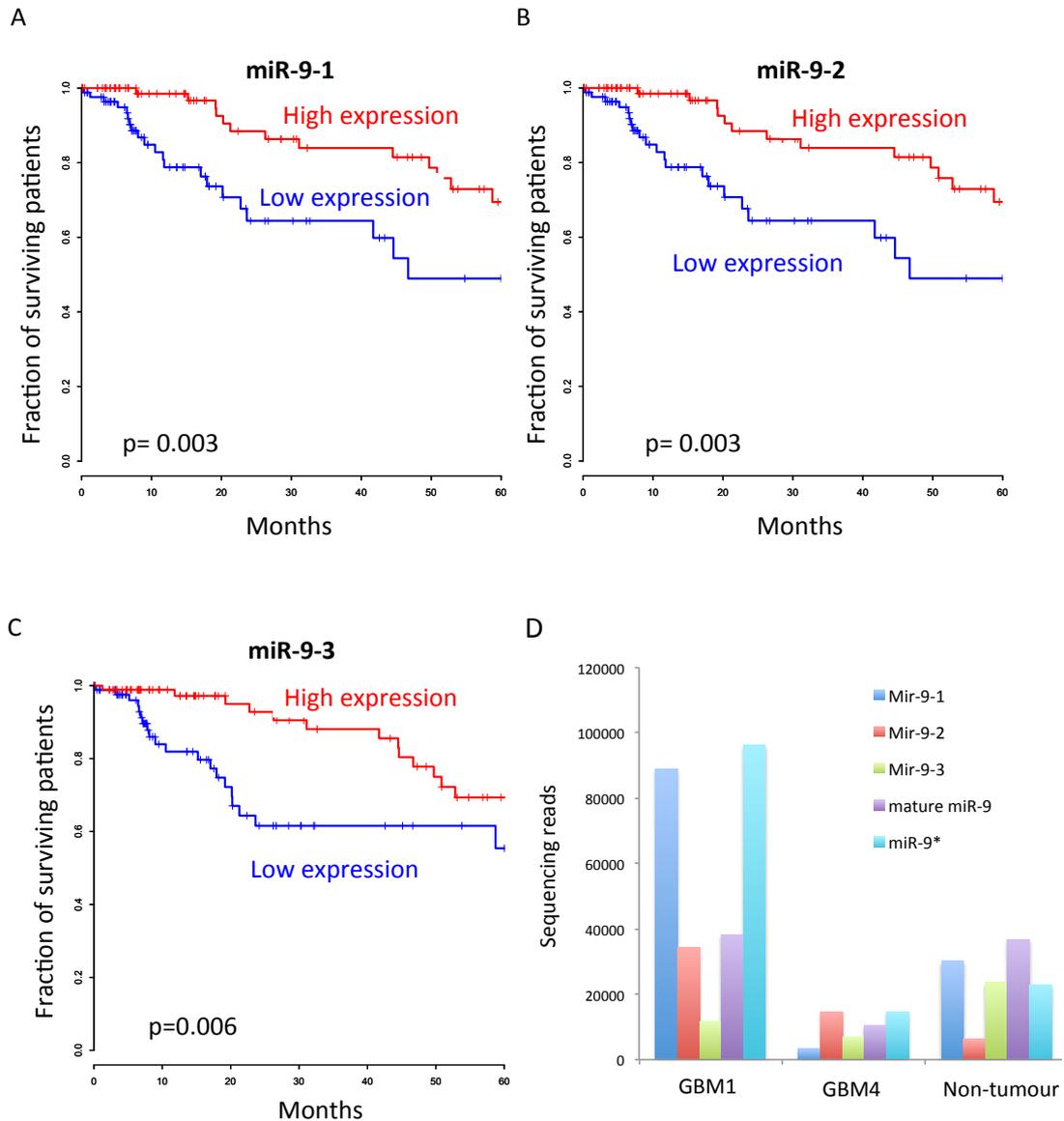


Figure 4.4. All precursors of miR-9 are associated with survival but levels of the precursors differ across samples.

A-C) Sequencing data from TCGA grade II and III glioma samples showed all precursor sequences are prognostic. D) Read counts per million (Y-axis) of miR-9 precursor in two glioma stem cell lines grown on laminin, and a non-tumour sample. Characterisation according to the signature of Mao et al. (Mao et al., 2013) loosely indicates GBM-1 may be more proneural-like and GBM-4 more mesenchymal-like (Appendix 4.3). Lower miR-9 levels in GBM-4 to GBM-1 reflect the similarity of GBM-4 to a mesenchymal subtype with lower miR-9 levels.

MiR-9 has been reported to be highly expressed in brain (Sempere et al., 2004) and so the levels of miR-9 sequences were assessed compared to all other microRNA sequences in a glioblastoma sample and a non-tumour sample. This showed that miR-9 represents nearly 30% of all microRNAs in one glioblastoma sample, and 6% of all microRNA in a non-tumour sample (Fig. 4.5).

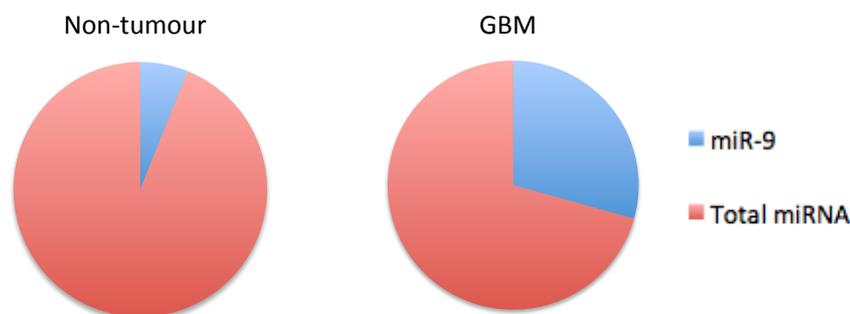


Figure 4.5. miR-9 sequences are highly represented in the total population of microRNA sequences with brain and glioblastoma samples.

The proportion of miR-9 sequences compared to all other microRNA was assessed to determine its abundance in glioblastoma (GBM) and non-tumour tissue. These pie charts show the numbers of sequencing reads per million mapping to a miR-9 mature sequence or precursor compared to reads mapping to all other microRNAs. miR-9 is more highly represented in the total microRNA from a glioblastoma patient sample. Note these samples are not matched.

4.3.3. Levels of expression of miR-9 in tumours with different percentages of tumour cell content.

Following the finding that miR-9 is expressed at lower levels in mesenchymal tumours I sought to determine whether this is linked to a recent result showing that mesenchymal tumours have a lower percentage of tumour cells than other glioblastoma subtypes (Meyer et al., 2015). The 558 samples in the TCGA were split based on the median expression of miR-9 and the percentage tumour cells compared in the two groups using a Student's T-test. This showed that the group with a below median expression of miR-9 also have a significantly lower tumour cell percentage ($p=2.2e^{-16}$, Fig. 4.6). Spearman's correlation of

percentage tumour cells with miR-9 levels was also significant for a positive correlation ($Rho= 0.23$, $p=5.63e^{-08}$).

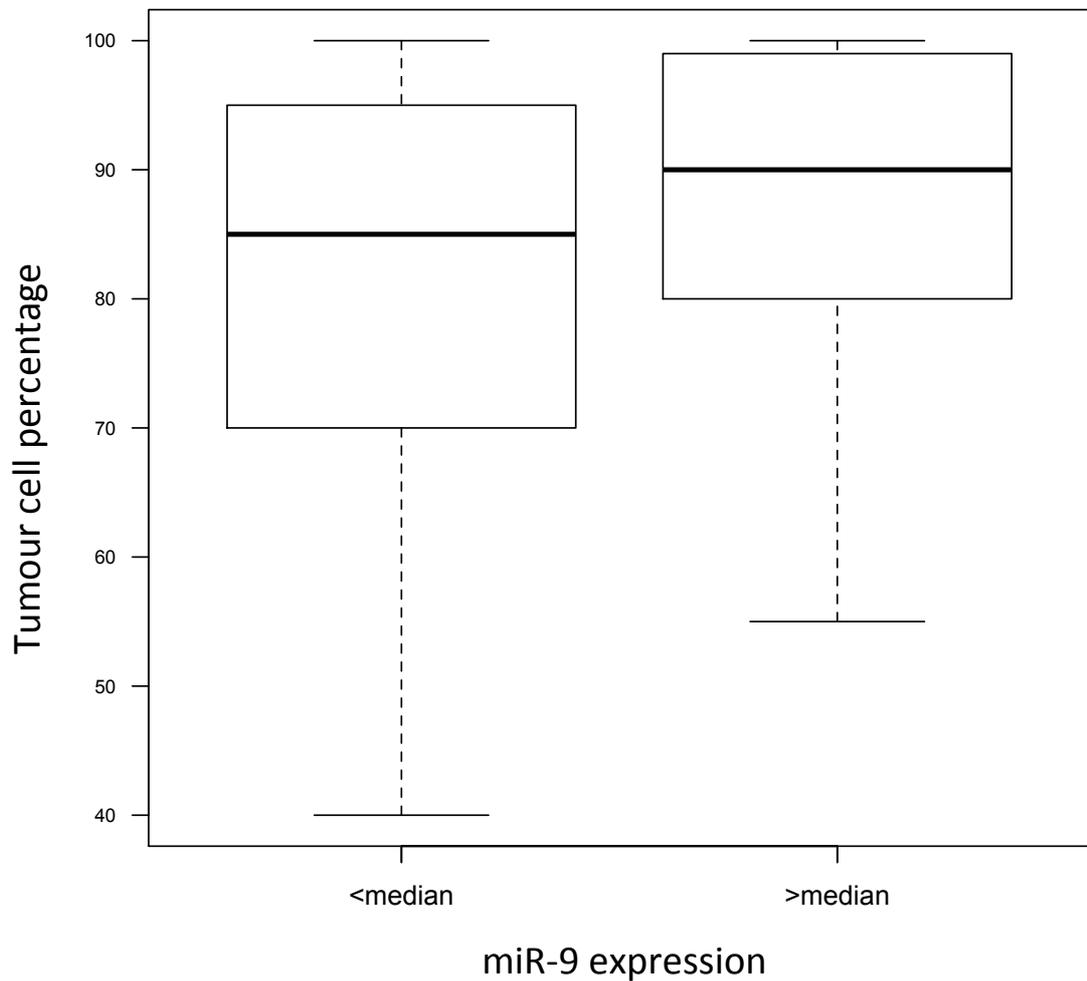


Figure 4.6. The patient group with below median expression of miR-9 have fewer tumour cells in the sample.

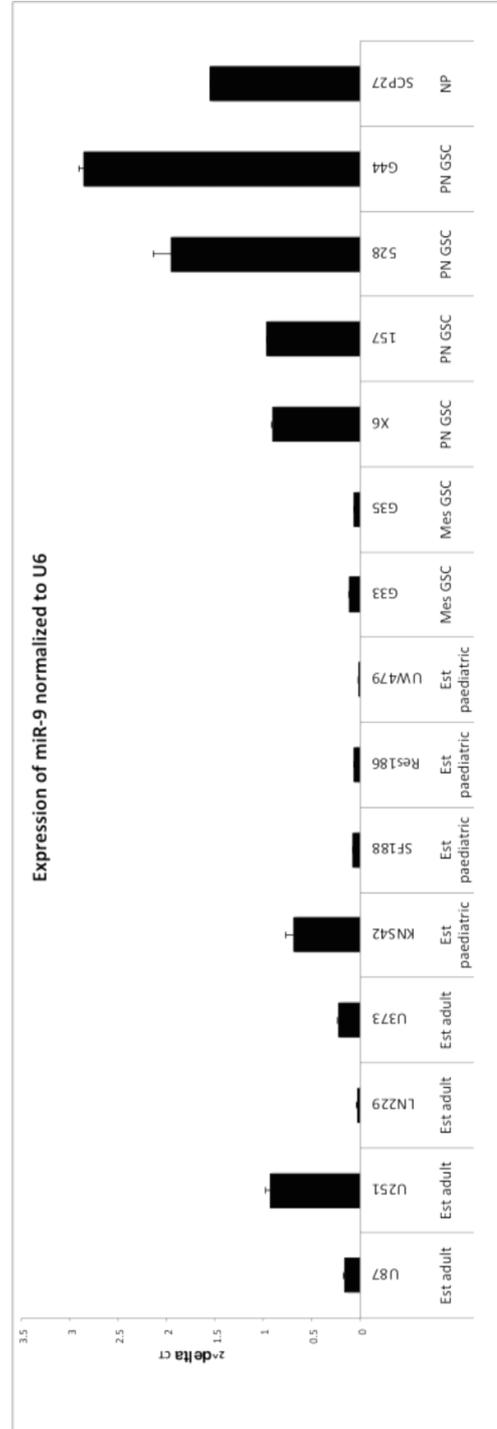
The Y-axis represents the percentage tumour cells in each sample defined by the TCGA. The X-axis represents the below median miR-9 expression level and above median miR-9 expression level based on the merged data from two probes for miR-9 on the microarray. The boxes represent the interquartile range and whiskers represent the maximum and minimum expression level. Patients with below median expression levels of miR-9 in their tumour show a lower percentage of tumour cells. This may suggest that samples with more infiltration have less miR-9.

4.3.4. Expression of miR-9 in mesenchymal glioma stem cells.

Tumours with more extensive infiltration may have lower miR-9 simply because the infiltrating cells have lower miR-9 expression, thus diluting the levels of miR-9 in the total sample. In order to test whether this contributes to why miR-9 is lower in the mesenchymal glioblastoma subtype and also prognostic, I measured the levels of expression in glioma stem cell lines. These lines were characterised as proneural or mesenchymal according to the Mao *et al* signature (Mao et al., 2013) (characterisation previously performed by Dr. Marco Mineo, Appendix 4.3). A neural progenitor cell line, SCP27, was also assessed for miR-9 levels and Dr Marjorie Boissinot performed qRT-PCR for established adult and paediatric cell lines (Fig. 4.7). A student's T-test showed that mesenchymal stem cells have lower expression of miR-9 ($p=7.9e^{-4}$), suggesting that the infiltrates are not the sole explanation for lower miR-9 expression in mesenchymal glioblastoma. The other cell lines showed varying expression of miR-9. Highest levels were observed in proneural glioma stem cell lines.

Figure 4.7. Levels of miR-9 are low in mesenchymal glioma stem cell lines; other cell lines have varying expression levels.

Levels of miR-9 were tested in four established adult cell lines; U87, U251, LN229 and U373, four established paediatric cell lines KNS42, SF188, RES186 and UW479, two mesenchymal glioma stem cell lines; G33 and G35, four proneural glioma stem cell lines; X6, 157, 528 and G44 and one neural progenitor cell line SCP27. The Y-axis represents $2^{\text{delta Ct}}$ where delta Ct is U6 snRNA expression Ct minus miR-9 expression Ct.



4.3.5 Association of miR-9 with necrosis.

One of the hallmarks of mesenchymal glioblastoma is the more extensive necrosis (Naeini et al., 2013). Therefore it is plausible that miR-9 is expressed at lower levels in necrotic regions. In order to assess this I performed Spearman's correlation of percentage of necrosis with the levels of miR-9 across 558 glioblastoma samples in the TCGA. In accordance with my hypothesis above, this showed that samples with a lower miR-9 expression have more necrosis by T-test ($p=2.2e^{-16}$, Fig. 4.8). Spearman's correlation also shows negative correlation of miR-9 with necrosis ($Rho=-0.17$, $p=3.37e^{-05}$).

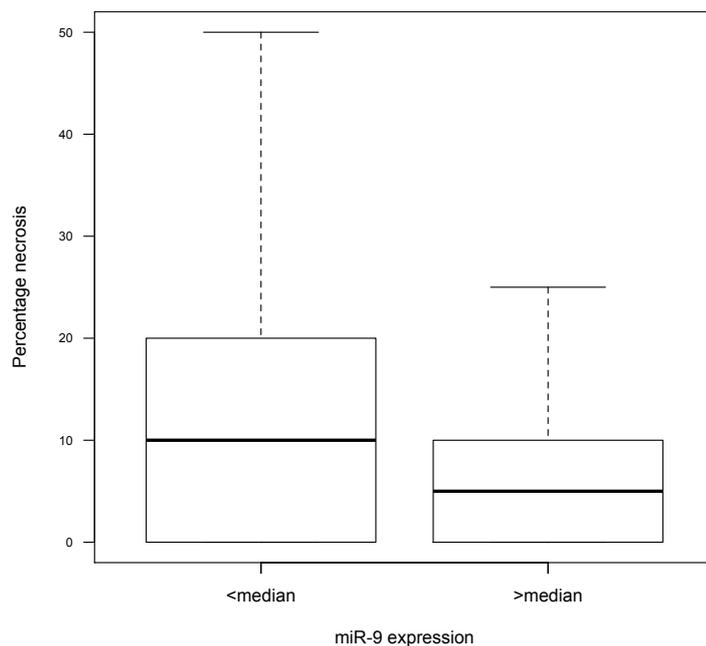


Figure 4.8. Glioblastoma tumours with lower miR-9 expression have more extensive necrosis.

The Y-axis represents percentage necrosis as defined by the TCGA and the X-axis represents quantile normalised miR-9 expression based on two probes for miR-9 in the microarray, separated on median miR-9 expression in all glioblastoma samples in the TCGA. The dark lines represent the median, the boxes represent the interquartile range and the whiskers show the maximum percentage necrosis recorded in these groups of patients. As expected, patients with below median miR-9 expression had a higher percentage of necrosis suggesting miR-9 levels may be lower in necrotic regions, and therefore contributing to the lower levels of miR-9 in mesenchymal glioblastoma samples.

4.3.6. Effect of hypoxia on miR-9 expression levels.

Another hallmark of mesenchymal glioblastoma is extensive hypoxia and therefore it is plausible that miR-9 may decrease when cells are exposed to hypoxia. As mesenchymal cells have more extensive hypoxia it could be that these pockets in the tumour have even less miR-9 expression. I tested this by exposing a mesenchymal glioma stem cell line; G35, as a monolayer and as spheres, to 1% oxygen for 72 hours and performing qRT-PCR for miR-9. This showed that miR-9 does indeed decrease with hypoxia (Fig. 4.9). In order to monitor hypoxia response of the cells, a HIF-1a target was also measured, which showed an increase at lower oxygen levels (Appendix 4.4).

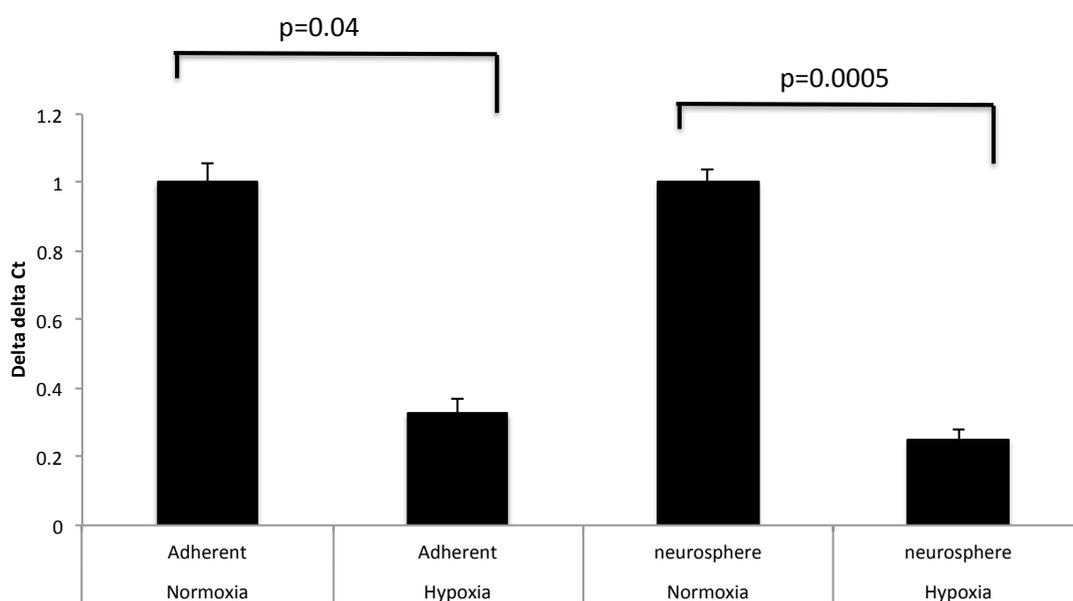


Figure 4.9. When cells are exposed to hypoxia, levels of miR-9 are lower.

In order to determine what the effect of hypoxia is on miR-9 levels, a mesenchymal stem cell line was exposed to hypoxia and normoxia and levels of miR-9 were quantified after 72 hours. This was tested with cells growing as a monolayer and as spheres. The Y-axis represents the delta delta Ct G35 cells (U6 Ct minus miR-9 Ct in normoxia/ U6 Ct minus miR-9 Ct in hypoxia/normoxia), and the X-axis represents the treatment condition. Both had lower expression of miR-9 after exposure to 1% oxygen in comparison to those exposed to 20% oxygen. These are merged results of duplicate experiments.

4.3.7. Pathway enrichment of predicted miR-9 targets.

In order to predict the functions of miR-9, targets were predicted using the intersection between the targets predicted from microRNA target prediction algorithms Miranda (Miranda et al., 2006) and Targetscan (Lewis et al., 2005). This identified 1148 targets, which were used for pathway analysis using Metacore, which performs pathway enrichment using hypergeometric analysis. This method uses the hypergeometric distribution, which describes the probability of the number of successes in a number of draws, without replacement, from a finite population size. The intersection between the target genes and those in a pathway is calculated and under the null hypothesis of no enrichment the probability of occurrence of an intersection of a certain size by chance follows the hypergeometric distribution. The top ten pathways are shown in Fig. 4.10.

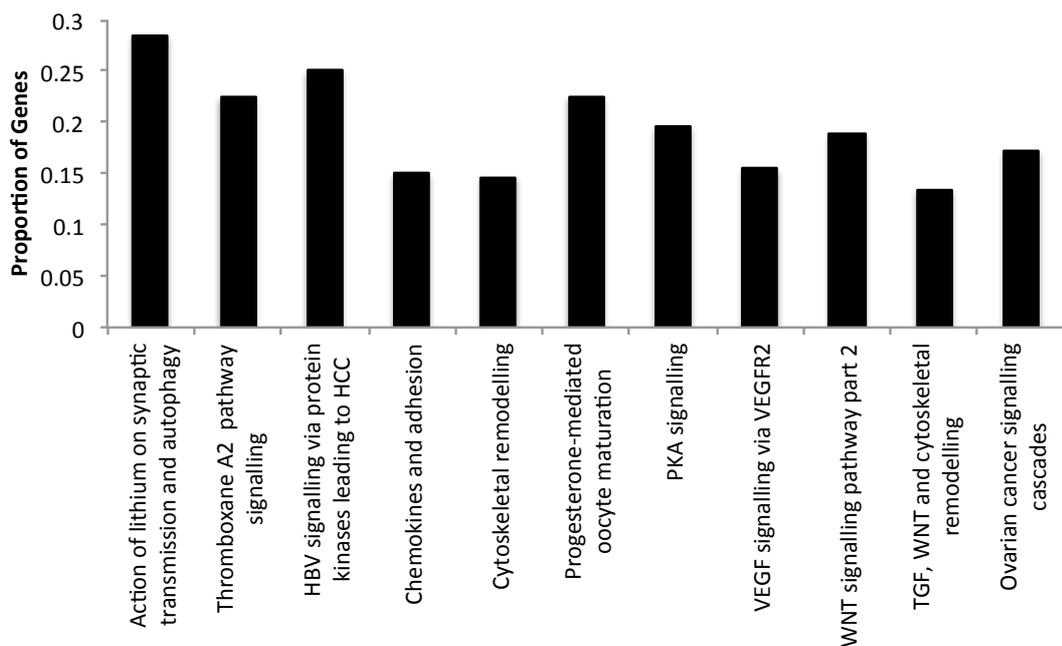


Figure 4.10. The top ten enriched pathways of the predicted targets for miR-9, with representation of proportion of genes in the pathway.

This shows the proportion of genes in the pathway predicted to target miR-9 along the Y-axis and the pathways defined by Metacore on the X-axis, in order of significance from left to right. In the top ten pathways enriched for predicted miR-9 targets there is high representation of cytoskeletal remodelling, and chemokines and adhesion. Cell adhesion and cytoskeletal modelling were

chosen for further study as these pathways are of interest in the transition to a mesenchymal subtype.

4.3.8. Effect of miR-9 on adhesion.

In order to determine whether miR-9 is involved in cell adhesion, as indicated by the pathway analysis of predicted target genes, I assessed cell adhesion following overexpression of miR-9 in GSCs as these are able to form neurospheres *in vitro*. miR-9 was overexpressed using Dharmacon mimics and compared to a scramble mimic sequence (and quantified by RT-PCR with every experiment to confirm overexpression). Cell-cell adhesion assays were performed using glioma stem cell lines and measured the number and size of spheres after 4 hours of seeding single cells and culturing in rotating suspension. This showed that overexpression of miR-9 increased the number/size of spheres compared to the scrambled control (Fig. 4.11).

Cell substrate adhesion assays were performed using established cell lines (as these do not require substrate such as laminin or fibronectin to adhere to) on tissue culture plates. After 4 hours non-attached cells were washed away and the number of cell remaining were assessed. This showed that overexpression of miR-9 increased the ability of cells to attach to the plastic (Fig. 4.12).

The design of the cell adhesion experiments was different for cell cell adhesion and cell substrate adhesion. Different cell lines were used because they have different abilities to attach to other cells, and also to substrates. For the cell cell adhesion the fold change from T0 was used because it could not be confirmed that all cells were single cells at this stage. For cell substrate adhesion however, fold change was not required and therefore assessment was at the end of the 4 hours.

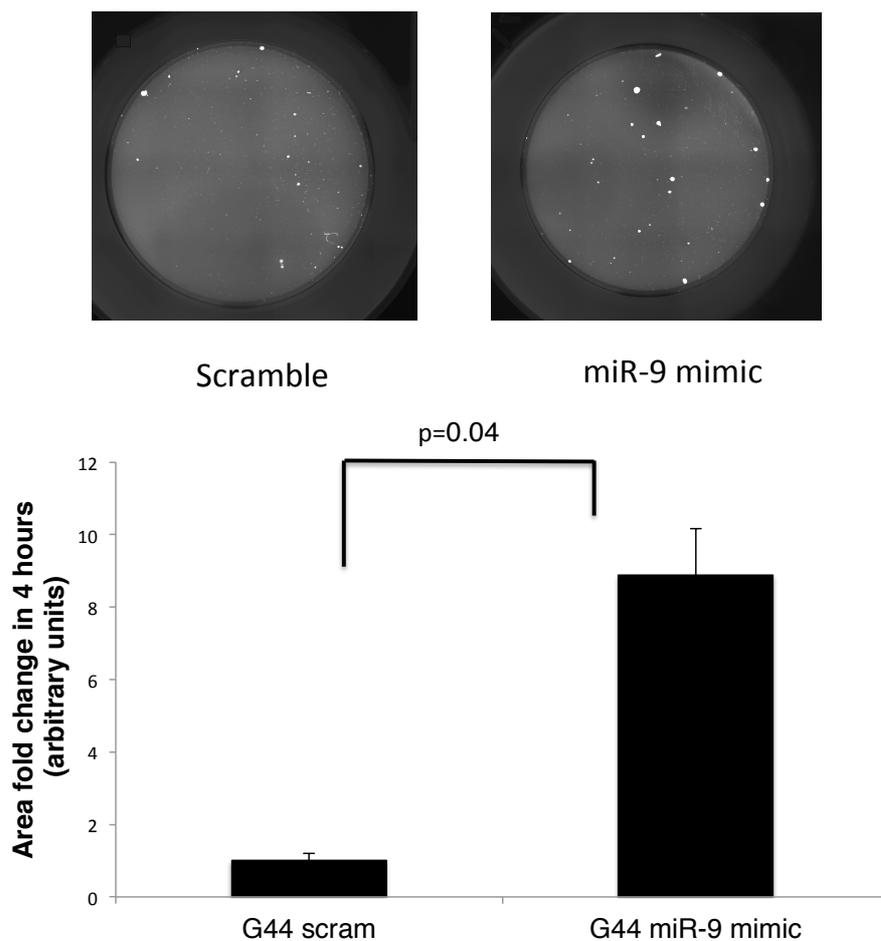


Figure 4.11. Cell adhesion assays show that miR-9 increased the ability of cells to adhere to each other.

After four hours of incubation, there were more cells in a sphere when cells had been transfected with the miR-9 mimic compared to cells transfected with the scramble. The Y-axis represents the area increase after four hours, which was quantified by applying a mask to the well image, which removed particles less than a threshold size to remove single cells from the image. G44 scramble has been normalised to 1. The stained area remaining was then compared to the area present with mask applied at the start of the experiment (this was done in case some cells were not completely dissociated at the start of the experiment). Representative of two experiments.

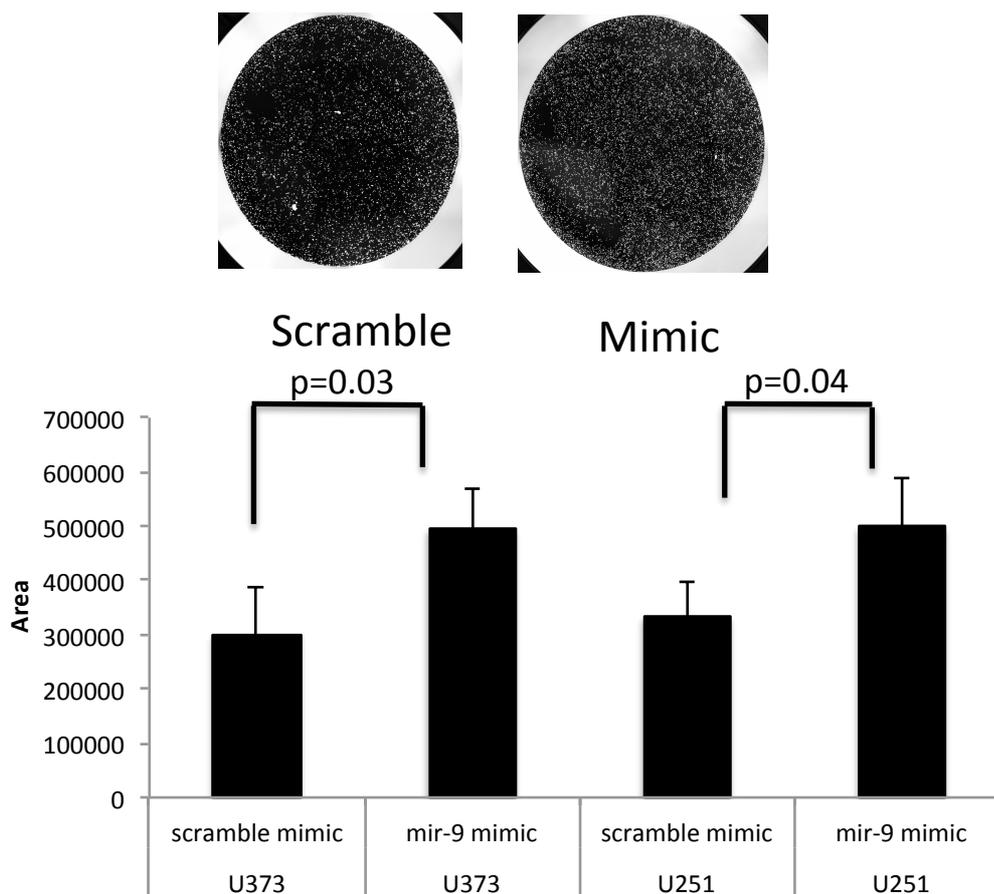


Figure 4.12. Cell adhesion assays show that miR-9 increased the ability of cells to adhere to a substrate.

After four hours of incubation, there were more cells attached to the bottom of a well in cells overexpressing miR-9 compared to a scrambled control. The Y-axis represents the area after four hours quantified in the same way as for the cell cell adhesion assay (Fig. 4.11) without normalisation to time 0. Shown as the merge of two experiments.

4.3.9 Effect of miR-9 on cell viability.

It is plausible that the effects seen in hypoxia, and observations of cell adhesion, are due to miR-9 altering cell viability. I therefore tested whether miR-9 has an effect on the viability of a cell, which may have implications in other assays. Cell viability assay Presto Blue, which uses cell permeable redox indicator resazurin to detect cell reducing power, showed that miR-9 overexpression does not alter cell viability (Fig. 4.13).

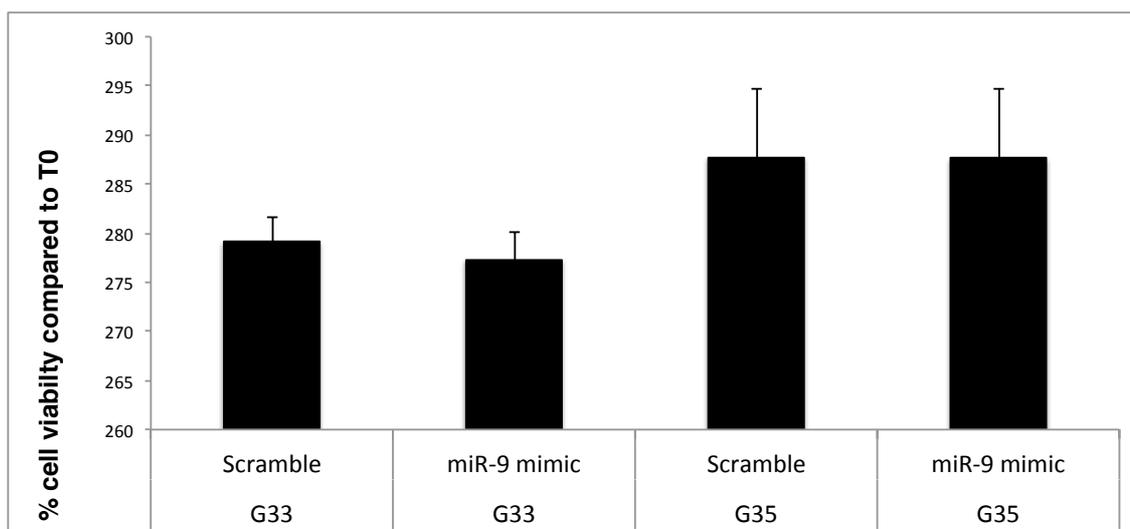


Figure 4.13. miR-9 overexpression does not affect cell viability.

The effect of miR-9 on cell viability was assayed because it may affect the results of other experiments. Cells transfected with a miR-9 mimic did not have altered cell viability compared to those transfected with a scramble (48-hour transfection). The Y-axis represents the percentage cell viability determined by Presto Blue assay compared to the cell viability prior to transfection.

4.3.9 A target of miR-9 is SHC1, a cytoskeletal remodelling protein.

I have ascertained that miR-9 is lower in mesenchymal glioblastoma and also has a role in cell adhesion, potentially through cytoskeletal remodelling. To identify a target of miR-9 that may be eliciting this effect I sought to determine which predicted target of miR-9 is also overexpressed in mesenchymal glioblastoma, and is associated with the Gene Ontology terms “Cell adhesion” and “Cytoskeletal Organisation”. I determined which mRNAs are significantly altered between 156 mesenchymal tumours from the TCGA and 96 proneural tumours using LIMMA and determined which of these intersected with miR-9 targets. Then I intersected this list with the Gene Ontology lists for cell adhesion and cytoskeletal organisation. The final list contained nine genes; *CLDN2* (claudin 2), *DSG2* (desmoglein 2), *ATXN3* (ataxin 3), *TESK2* (testis-specific protein kinase 2), *ARFGEF1* (ADP-ribosylation factor guanine nucleotide-exchange factor 1), *DIAPH2* (diaphanous-related formin 2), *MYPN* (myopalladin), *SHC1* (Src homology 2 domain containing transforming protein

1). These were individually assessed and *SHC1* was chosen as a candidate to study further, due to high scoring target prediction sites as determined by Targetscan (Fig. 4.14).

Position 209-215 of SHC1 3' UTR	5' ... CCCCAAACAUAUAUCACCAAAGU ...
hsa-miR-9	3' AGUAUGUCGAUCUAUUGGUUUCU
Position 1465-1471 of SHC1 3' UTR	5' ... ACGGGUCUCACUUAUACCAAAGG ...
hsa-miR-9	3' AGUAUGUCGAUCUAUUGGUUUCU

Figure 4.14. The predicted target sites in SHC1 3' UTR for miR-9.

Of all microRNAs predicted to target *SHC1*, miR-9 has the top score using both scoring methods in Targetscan. The total context score, which is calculated by site-type contribution, 3' pairing contribution, local AU contribution, position contribution, target site abundance contribution and seed-pairing stability contribution, is -0.34. The Aggregate P_{CT} which scores based on the probability of conservation based on microRNA selection rather than by chance, is 0.97. There are two predicted target sites for miR-9 in this 3' UTR.

Levels of SHC1 protein were assessed after miR-9 overexpression using western blotting. This showed that miR-9 overexpression decreased SHC1 protein levels (Fig. 4.15 A). To further confirm this result a plasmid with the *SHC1* 3' UTR downstream of the *Gaussia princeps* luciferase (GLuc) reporter gene under the control of a SV40 promoter was co-transfected into U251 cells with either the miR-9 mimic or a scramble sequence (Fig. 4.15 B).

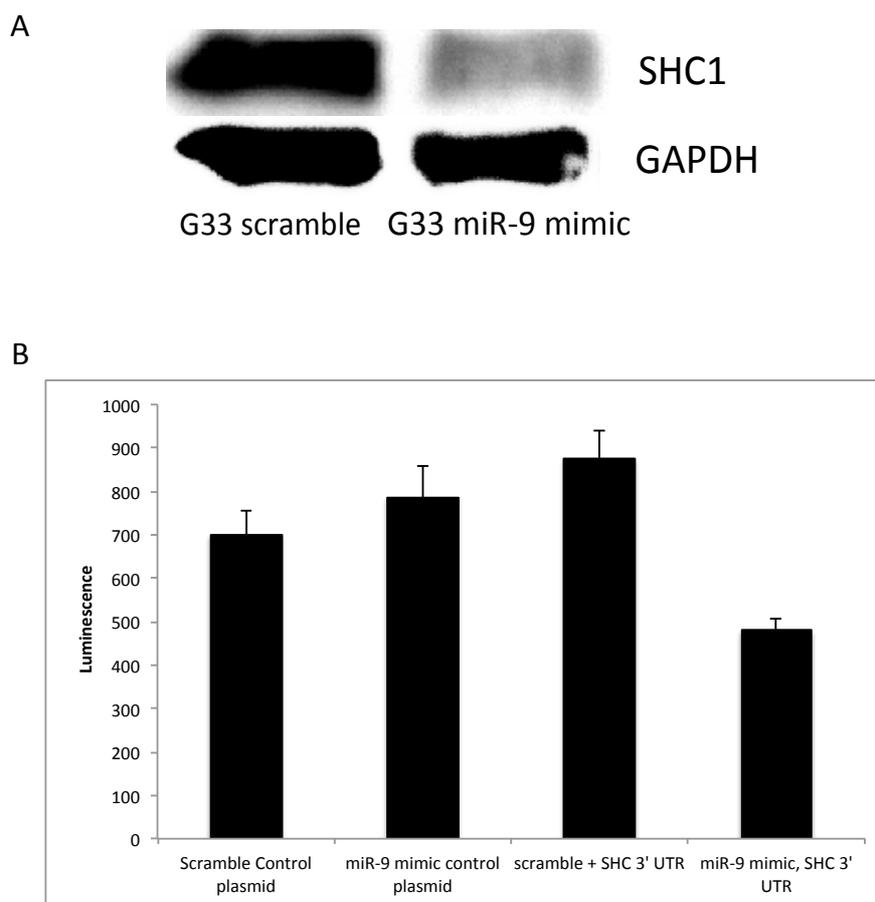


Figure 4.15. Overexpression of miR-9 decreases SHC1 protein levels.

SHC1 had a high score on Targetscan as a predicted target for miR-9, therefore levels of *SHC1* protein and expression of a luciferase reporter under the control of *SHC1* 3' UTR were measured with overexpression of miR-9. A) Western blotting to measure protein levels of *SHC1* after 48 hours of transfection with a miR-9 mimic or scramble. The p66 isoform of *SHC1* is shown here and GAPDH was used as a control. Results are representative of duplicate experiments. B) Results of a 48-hour co-transfection of the miR-9 mimic with the plasmid and scramble with the plasmid. A paired two-tailed *T*-test between these showed that overexpression of miR-9 significantly decreased luminescence ($p=4.43e^{-5}$). Both the miR-9 mimic and scramble sequence were also co-transfected with a plasmid without the *SHC1* 3' UTR present as controls.

4.4 Discussion.

Glioblastoma can be separated into different molecular subtypes with different clinical and phenotypic characteristics (Verhaak et al., 2010; Brennan et al., 2013). Current methods to subtype glioblastoma in patient samples rely on the quantification of a large number of mRNAs using the Verhaak *et al.* signature (Verhaak et al., 2010). This is not conducive to clinical testing due to the cost and labour intensity. My research has shown that miR-9 is lower in the mesenchymal subtype of tumours and also in glioma stem cell lines expanded from patient tumours. This has been tested using microarray analysis on patient samples, sequencing analysis and qRT-PCR analysis in glioma stem cell lines. With further validation, miR-9 would be an ideal candidate as a biomarker delineating this subtype, which confers an aggressive tumour type with a good response to standard treatment (Mao et al., 2013; Verhaak et al., 2010). The cause for the lower miR-9 expression levels in this subtype is not entirely clear from my data. Patient samples are heterogeneous, and different clones will have varying levels of miR-9, which culminate to produce a lower level from a mesenchymal glioblastoma tumour sample. Using miR-9 expression from the TCGA patient samples I have shown that despite the heterogeneity within the tumour, lower miR-9 expression is detectable within mesenchymal tumours without the need for multiple sampling of a particular tumour. My data suggests that the levels of miR-9 could be lower in these tumours because of more extensive infiltration. As miR-9 is a neural microRNA (Sempere et al., 2004), levels are likely lower in non-neural cells (Landgraf et al., 2007). A tumour highly infiltrated with other cells, such as lymphocytes, is therefore expected to have a lower overall miR-9 expression level. However, miR-9 levels remain low when glioma stem cells are expanded from these samples in optimal conditions, representing a more homogenous population of cells. This indicates that the mesenchymal glioblastoma tumour cells themselves have a lower level of the microRNA.

Analysis of the precursors of miR-9 showed that all are prognostic in TCGA patient samples. The sequencing data showed that it could be a combination of miR-9-1 and miR-9-2 that cause higher levels of miR-9 in proneural glioblastoma. Compared to normal brain, miR-9* is, like miR-9, also expressed

at higher levels in the GSC line GBM-1 to GBM-4, which suggests regulation at the transcriptional level rather than stability of the precursor transcript. In both of the cell lines the levels of miR-9* are higher than that of miR-9, which may indicate an alteration in processing generating more of the 3p transcript.

As previously mentioned, mesenchymal glioblastoma has extensive regions of hypoxia and necrosis and levels of miR-9 may fluctuate within these regions. I have shown by correlation that tumours with more necrosis have lower levels of miR-9, and by qRT-PCR that the cellular response of a glioma stem cell to hypoxia causes a decrease in miR-9 levels. This may suggest that certain regions of a mesenchymal tumour may have lower expression levels of miR-9 compared to others. The levels of oxygen termed as 'hypoxia' and 'normoxia' here refer to an adjustment on atmospheric oxygen (1% and 21% oxygen respectively), and not a comparison between physiological levels of normoxia (<1%-10% (Evans et al., 2004)). If the experiment had been performed at these levels, results may have been different, or possibly more significant.

Pathway analysis of predicted targets of miR-9 suggested a role for miR-9 in adhesion and cytoskeletal remodelling. Adhesion assays confirmed that miR-9 promoted cellular adhesion. Mining for a target that may mediate this function in mesenchymal glioblastoma pointed to SHC1, an adapter protein involved in signal transduction. TargetScan (Lewis et al., 2005) also shows that miR-9 is predicted to target other SHC family members; SHC2 and SHC3. Three different isoforms of SHC1 protein exist, all obviously with the same 3' UTR sequence present and therefore under the control of miR-9. An important role for SHC1 is as a sensor for external signals, prompting a form of programmed cell death called anoikis. This occurs when anchorage dependent cells, such as epithelial cells, detach from the extracellular matrix (Chiarugi & Giannoni, 2008). Epithelial cells exhibit strong cell-cell adhesion properties, and this can be likened to differences seen *in vitro* between proneural and mesenchymal glioblastoma stem cells (Mao et al., 2013). The epithelial mesenchymal transition (EMT) is characterised by a loss of these interactions and release from the epithelial layer (Lamouille et al., 2014).

The mesenchymal subtype of glioblastoma has low levels of miR-9, yet breast cancer samples with high expression of vimentin and CD44 (mesenchymal markers) have higher levels of miR-9 (Gwak et al., 2014), which may reflect

differences in the extracellular matrix of these cancers. Several experiments have validated e-cadherin as a target of miR-9, and increase of miR-9 in carcinomas has been shown to contribute to the 'cadherin switch' and loss of e-cadherin-mediated adhesions in the epithelial-mesenchymal transition (Ma et al., 2010; Sethupathy et al., 2006). However, this switch has been shown to be absent in glioblastoma cells (Mikheeva et al., 2010). Tan *et al.* showed that miR-9 acts in a negative feedback loop with CREB (cAMP response element-binding protein) (Tan et al., 2012), and this protein has the ability to regulate WNT, TGF- β (transforming growth factor beta) and Notch signalling (Zoni et al., 2014), which were also identified in my pathway analysis of miR-9 targets. Another well-characterised target of miR-9 is NF- κ B (nuclear factor kappa-light-chain-enhancer of activated B cells) (Bazzoni et al., 2009), which mediates mesenchymal differentiation in glioma (Bhat et al., 2013). These findings point to a strong role for miR-9 targets as oncogenic drivers of the mesenchymal subtype of glioblastoma.

5. MicroRNAs predicting response to the anti-angiogenic drug bevacizumab.

'For the sick it is important to have the best.' Florence Nightingale.

5.1. Introduction.

Bevacizumab is an anti-angiogenic monoclonal antibody that acts by slowing the growth of new blood vessels through inhibition of VEGFA (vascular endothelial growth factor A). In glioblastoma, there have been two recent, prospective, randomised, placebo controlled clinical trials to assess whether bevacizumab improves survival in patients with newly diagnosed glioblastoma; AVAglio and RTOG 0825 (Chinot et al., 2014; Gilbert et al., 2014). In AVAglio patients received 10mg/kg of bevacizumab or placebo every two weeks with concurrent TMZ/radiotherapy ending with a dose on the last day of TMZ. Then, after a break of 4 weeks bevacizumab was administered with TMZ maintenance. Following this, 15mg/kg of bevacizumab or placebo alone were given every three weeks until disease progression (Chinot et al., 2014). In RTOG 0825 bevacizumab or placebo was administered 10mg/kg every two weeks from the fourth week of radiotherapy, until disease progression, toxicity or completion of adjuvant therapy up to a maximum of 24 cycles (Gilbert et al., 2014). Both studies reported an improved progression-free survival but no overall survival benefit. Despite these results there is anecdotal evidence, and some evidence from the tails of Kaplan Meier curves, that some patients may benefit from bevacizumab treatment, and appropriate methods for identification of these patients is an unmet need (Field et al., 2014, Prados, 2014; Reardon & Wen, 2014).

Prediction of response to bevacizumab in glioblastoma patients has been attempted previously, using a nine-gene expression signature representative of the mesenchymal glioma subtype (Colman et al., 2010), which was assessed using tumour samples from the RTOG 0825 patients. The results showed a significant association between worse OS (overall survival) and PFS (progression free survival) with the mesenchymal subtype. A smaller, more recent trial (the BELOB trial) in the Netherlands assessed bevacizumab alone and in combination with the alkylating nitrosourea compound lomustine for recurrent glioblastoma, with a primary endpoint of overall survival (Taal et al.,

2014). This involved administration of lomustine 110 mg/m² once every 6 weeks, bevacizumab 10 mg/kg once every 2 weeks, or a combination treatment including lomustine 110 mg/m² every 6 weeks and bevacizumab 10 mg/kg every 2 weeks. Results showed that the EGFR amplified, classical glioblastoma subtype responded well to combination therapy and the mesenchymal subtype showed poor response to single agent bevacizumab (although there were only 28 patients in the mesenchymal single agent bevacizumab group) (Eraslan et al., 2014). This result is supported by the fact that anti-angiogenic therapy resistance is associated with the mesenchymal transition, and tumours with more infiltration are more resistant to these drugs (Piao et al., 2012; Piao et al., 2013). As previously mentioned in Chapter four, these aspects of a tumour are often observed together. In another study, Omuro *et al.* found patients with proneural tumours had a reduced overall survival compared to other subtypes when treated with bevacizumab, TMZ and hyperfractionated stereotactic radiotherapy (Omuro et al., 2014).

In an immune context, infiltration by monocytes expressing the tyrosine kinase Tie2 (tunica internal endothelial cell kinase 2) has been shown to confer anti-angiogenic therapy resistance to tumours (Gabrusiewicz et al., 2014). Tie2, and Ang1 (angiopoietin 1) have been shown to predict progression free survival in patients with ovarian cancer treated with bevacizumab (Backen et al., 2014). The AVAglio trial ruled out VEGFA and VEGFR2 (vascular endothelial growth factor receptor 2) as predictive or prognostic biomarkers, although a VEGFA SNP rs2010963 is associated with vascular toxicity (Field et al., 2014; Di Stefano et al., 2014). MMP2 (matrix metalloproteinase 2) has also been shown to be associated with response and survival in bevacizumab treated patients (Tabouret et al., 2014). MicroRNAs have not been studied in the prediction of response to bevacizumab in glioblastoma to date.

So far, there has been no validation of a bevacizumab response biomarker for any disease type to date, although biomarkers have been suggested (Bruhn et al., 2014). The AVAglio and RTOG 0825 trials reported an improvement in PFS alone and this is solely meaningful if it translates to an overall survival improvement (Field et al., 2014). These data therefore do not provide enough evidence to continue with bevacizumab use in newly diagnosed glioblastoma.

Different results may have been found if patients were selected according to a biomarker.

Identification of a biomarker for bevacizumab treated patients is particularly complicated as patients are treated at different stages of disease, with varying lengths and numbers of cycles due to toxicity. Patients are also treated with different combination therapies, and may have been treated with a different drug prior to administration of bevacizumab. The cost of the drug limits its access to just a small number of institutions, and therefore obtaining samples from these patients is difficult. In this study, I have attempted to identify a microRNA signature prognostic in bevacizumab treated patients (newly diagnosed and recurrent glioblastoma) with overall survival as an endpoint. I have adapted my LASSO (least absolute shrinkage and selection operator) method used in Chapter two to predict prognosis in glioblastoma to this subgroup of patients. This showed an 8-microRNA signature defined a subgroup with better prognosis, and when applied to patients treated with all treatments was less powerful suggesting that this signature is related to bevacizumab.

5.2. Methods.

5.2.1. TCGA clinical information and expression data.

Level 3 Agilent microRNA 8x15k microarray expression data plus clinical and treatment information for 562 glioblastoma samples, 90 of which were from patients treated with bevacizumab (either as adjuvant, progression or recurrence treatment), were downloaded from TCGA (TCGA, NIH). Level 3 data provide an expression value for each microRNA by merging of probes values. Patients had been treated using varying numbers of 2-3 week cycles of bevacizumab therefore treatment time was determined as the date of start of treatment to the date of end of treatment. Samples were taken at diagnosis. Three patients were removed due to lack of start date information, resulting in 87 patients (Table 5.1). These were randomly split into test and training set groups of 50 and 37 patients respectively. This was done because it was not possible to obtain an independent dataset, which would be the ideal validation. These numbers were chosen to generate maximum power in generation of the model, whilst allowing a validation cohort for testing of the model.

	TRAINING SET No. of patients (n=50)	TEST set No. of patients (n=37)
Age	median 54.5 years	median 56 years
<60 years	33	27
≥60 years	17	10
Gender		
Male	27	23
Female	23	14
Karnofsky Performance score		
≤70	21	18
>70	29	19
Days to death/ last follow-up		
<450 days	25	23
≥450 days	25	14
<30 days	0	0
Treatment regimen		
Adjuvant	8	8
Progression	29	16
Recurrence	5	2
Not available	8	11
Mean treatment length (days)	205.9	177.4

Table 5.1. Characteristics of the patients included in the study.

MicroRNA expression levels for 87 patients from the TCGA were included in the study. These were separated into a training set of 50 patients for generation of the model and a test set of 37 patients for testing of the model. The test and training set patients were allocated randomly, with more patients in the training set to ensure a robust model. All patients had been treated with bevacizumab; whether this treatment was adjuvant (16% of training set patients and 22% of test set patients), at progression (58% of training set patients and 43% of test set patients), or at recurrence (10% of training set patients and 5% of test set patients) was noted. Mean length of treatment denotes the days from the start of first cycle to the end of last cycle. These groups were randomly assigned. The test set is a smaller group with a slightly poorer survival and more patients with a lower than expected KPS score (47% with a KPS lower than 70 compared to 42% with a KPS below 70 in the training set).

5.2.2. Generation of a risk algorithm for OS in bevacizumab-treated glioblastoma patients using microRNAs.

The training set samples were assessed using LASSO penalized regression (Tibshirani, 1996) with leave-one-out cross-validation using the R software (v2.15.1) and the Penalized package (Goeman, 2010). This produced 8 microRNAs with non-zero coefficients. A risk score was generated using the

sum of microRNA expression values weighted by the coefficients from the LASSO regression.

This was: $E_{\text{miR-n}} = \text{expression of microRNA n.}$

Risk score = $0.055E_{\text{miR-124a}} + 0.309E_{\text{miR-202}} + -0.184E_{\text{miR-204}} + 0.170E_{\text{miR-222}} + -0.194E_{\text{miR-363}} + -0.025E_{\text{miR-630}} + -0.322E_{\text{miR-663}} + 0.161E_{\text{miR-7}}$

The risk score was applied to all samples in the training set. The scores were then plotted on a density plot to determine whether there is a natural cut-off for responders and non-responders. The most prognostic cut-off was chosen based on log-rank tests at each minimum value on the density plot. The training set samples were then separated into responders and non-responders using this cut-off (0). A Cox regression model incorporating age and the log-rank test were used to assess OS of the two groups in the training set. The score was also assessed for PFS. A statistical significance threshold of $p=0.05$ was used throughout. The length of treatment time was correlated with the time of survival in both responder and non-responder groups.

5.2.3. Validation of the risk score in the test set.

The risk score was calculated with the above algorithm using the microRNA expression values for the 37 test set samples. The defined cut-off of 0 was used to separate the test set into two groups of responders and non-responders. A Cox regression model incorporating age and the log-rank test were used to assess OS of the two groups. The length of treatment time was correlated with the time of survival in both responder and non-responder groups.

5.2.4. Testing of the algorithm across all treatment types.

The risk score was applied to all 562 patients in the TCGA (treated with varying treatment regimens) (Table 5.2). This cohort was split into two groups based on the score cut-off of 0 and the two groups assessed by Cox regression and log-rank test.

	No. of patients (n=562)
Age	median 57.9 years
<60 years	286
≥60 years	276
Gender	
Male	342
Female	220
Karnovsky Performance score	
≤70	171
>70	257
Not available	134
Days to death/ last follow-up	
<450 days	372
≥450 days	190
<30 days	24

Table 5.2. Patient information for the whole TCGA dataset of glioblastoma patients.

These patients were treated with a range of treatment regimens, 472 of which did not include bevacizumab.

5.2.5. Characterisation of the two groups defined by the signature.

Each group (test group responders, test group non-responders, training group responders, training group non-responders, all responders from test and training groups and all non-responders from test and training groups) was analysed for the number of patients with mesenchymal glioblastoma according to Brennan *et al.* (Brennan *et al.*, 2013), the percentage of tumour cells and stromal cells in the sample and the type of treatment regimens (adjuvant or recurrent) the patients were treated with. This was done because previous data suggest that non-mesenchymal tumours fare better when treated with bevacizumab (Sulman *et al.*, 2013). These were analysed for significance using Fisher's exact test.

5.2.6. Testing of the signature in a group of National Cancer Institute (NCI) cell lines.

In order to assess the performance of the signature in another set of samples, I downloaded Agilent microarray (V2) microRNA expression data on eight NCI cell lines (different disease types, not including glioblastoma) that had been

tested for levels of apoptosis by western blot 48 hours after treatment with bevacizumab (Hein & Graver, 2013). Mean-centred microRNA expression data was downloaded from NCI (<http://discover.nci.nih.gov/cellminer/>, accessed January 2015). Expression levels were not recorded for miR-124a so this was omitted from the signature. The other seven microRNAs were used to calculate a signature:

This was: $E_{\text{miR-n}}$ = expression of microRNA n.

Risk score = $0.309E_{\text{miR-202}} + -0.184E_{\text{miR-204}} + 0.170E_{\text{miR-222}} + -0.194E_{\text{miR-363}} + -0.025E_{\text{miR-630}} + -0.322E_{\text{miR-663}} + 0.161E_{\text{miR-7}}$

The score was plotted on a density graph and a minimum density at a risk score of 0 was observed. This was used as a cut-off for responder and non-responder groups.

5.3. Results.

5.3.1. An 8-microRNA signature generated from the training set predicts prognosis in bevacizumab treated patients.

In order to identify microRNAs associated with OS in bevacizumab treated glioblastoma, LASSO regression (Tibshirani, 1996) was performed on 50 randomly chosen training set samples of 87 bevacizumab treated patients in the TCGA. This used Agilent microarray microRNA expression data (merged expression data for 534 microRNAs).

Using the LASSO method, 8 microRNAs were identified with non-zero regression coefficients. MicroRNAs given a negative LASSO coefficient are positive predictors of survival and *vice versa* (Table 5.3).

MicroRNA	Lasso regression coefficient	Chromosome location	Band
hsa-miR-124a	0.055	chr8:9760898-9760982 chr8:65291706-65291814 chr20:61809852-61809938	8p23.1, 8q12.3, 20q13.33
hsa-miR-202	0.309	chr10:133247511-133247620	10q26.3
hsa-miR-204	-0.184	chr9:70809975-70810084	9q21.12
hsa-miR-222	0.170	chrX:45747015-45747124	Xp11.3
hsa-miR-363	-0.194	chrX:134169378-134169452	Xq26.2
hsa-miR-630	-0.025	chr15:72587217-72587313	15q24.1
hsa-miR-663	-0.322	chr20:26208186-26208278	20q10
hsa-miR-7	0.161	chr9:83969748-83969857 chr15:88611825-88611934 chr19:4770670-4770779	9q21.32 15q26.1 19p13.3

Table 5.3. The eight microRNAs with non-zero coefficients from LASSO regression.

LASSO regression identified 8 microRNAs with non-zero coefficients, indicating they are suitable for inclusion in a signature algorithm. There are equal numbers of negative predictors and positive predictors. Data was obtained from miRbase (accessed February 2015). These microRNAs are expressed from seven different chromosomes with a high representation of chromosomes 9, 15, 20 and X. miR-7 and miR-204 are both expressed from within chromosome band 9q21; a region defined as having glioma proliferative capacity (Weber et al., 2001), although miR-204 is associated with a good response and miR-7 with a poor response. miR-663 resides within a region of heterochromatin on chromosome 20, which may reduce its expression levels.

5.3.2. A risk score combining expression values of the 8 microRNAs predicts survival in the training set of bevacizumab treated patients.

A risk score was created using the regression coefficients from the LASSO analysis (see methods page 68) to weight the expression value of each of the 8 microRNAs, in a similar way to the risk score in Chapter two.

This was: $E_{\text{miR-n}} = \text{expression of microRNA n.}$

Risk score = $0.055E_{\text{miR-124a}} + 0.309E_{\text{miR-202}} + -0.184E_{\text{miR-204}} + 0.170E_{\text{miR-222}} + -0.194E_{\text{miR-363}} + -0.025E_{\text{miR-630}} + -0.322E_{\text{miR-663}} + 0.161E_{\text{miR-7}}$

The risk score was then plotted as a density plot (Fig. 5.1) and each minimum value on the density plot was assessed for its prognostic power by cutting the data at this risk score value and performing a log-rank test (Table 5.4). This was performed so as to assess points in the dataset as a cut-off for dichotomisation. The peaks and troughs in the curve represent risk score values where there are more or few patients with this risk score. A trough (a minimum) may suggest that a responder and non-responder group exist within the data.

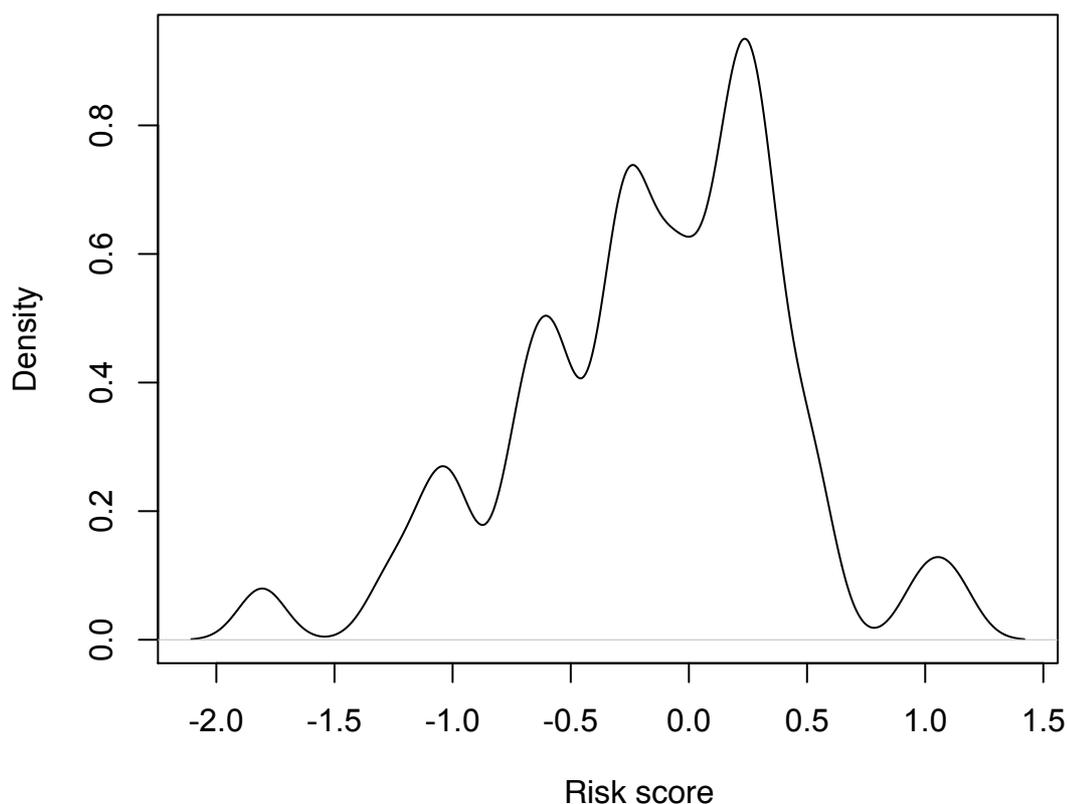


Figure 5.1. Density plot of the risk score in the 50 training set patients.

The risk score was calculated using the signature algorithm. The Y-axis represents the density of each score in the 50 patients and the X-axis represents the risk score, as defined by the expression of the 8 microRNAs. The minima on this graph were used as potential cut-offs for the responder and non-responder group.

Minimum value on graph	Risk score at minimum	Hazard ratio of log-rank	95% confidence intervals	p value
1	-1.54	NA		
2	-0.87	0.1	CI=0.01-0.77	0.027
3	-0.46	0.15	CI=0.05-0.45	7.40E-04
4	0.00	0.16	CI=0.07-0.36	1.20E-05
5	0.79	0.09	CI=0.01-0.95	0.046

Table 5.4. Results of log-rank test on each of the risk score minimum densities.

The risk score was plotted as a density plot, which highlighted 5 minima of density for the risk scores. These risk scores were assessed for their ability to separate the patients into responder and non-responder groups. The first column represents which minimum this refers to on the density plot in Fig. 5.1, numbered from left to right. The second column is the risk score at this minimum. The remaining three columns represent results of the log-rank test on the two groups (responders and non-responders) created by using this minimum as a cut-off; hazard ratios, 95% confidence intervals and p-values are recorded. As minimum 4 showed the most significance, with the most powerful hazard ratio, the risk score at this point (0) was used as a cut-off to separate the training set into responders and non-responders. This produced 29 'responder' patients with a risk score below 0, and 21 'non-responders' with a risk score above 0.

The cut-off was assigned in this way because if the signature is predicting the response of patients to the drug it is assumed that patients either respond or don't respond and therefore the score will fall into two groups. This has its limitations, as it could be complicated by other factors, such as toxicity of the drug to certain patients. The minimum density that occurred around a risk score of 0 showed the highest significance with the best hazard ratio, and so a risk score of 0 was chosen for the cut-off (Fig. 5.2). The median survival time of the responder group was 22 months and the median of the non-responder group was 12 months.

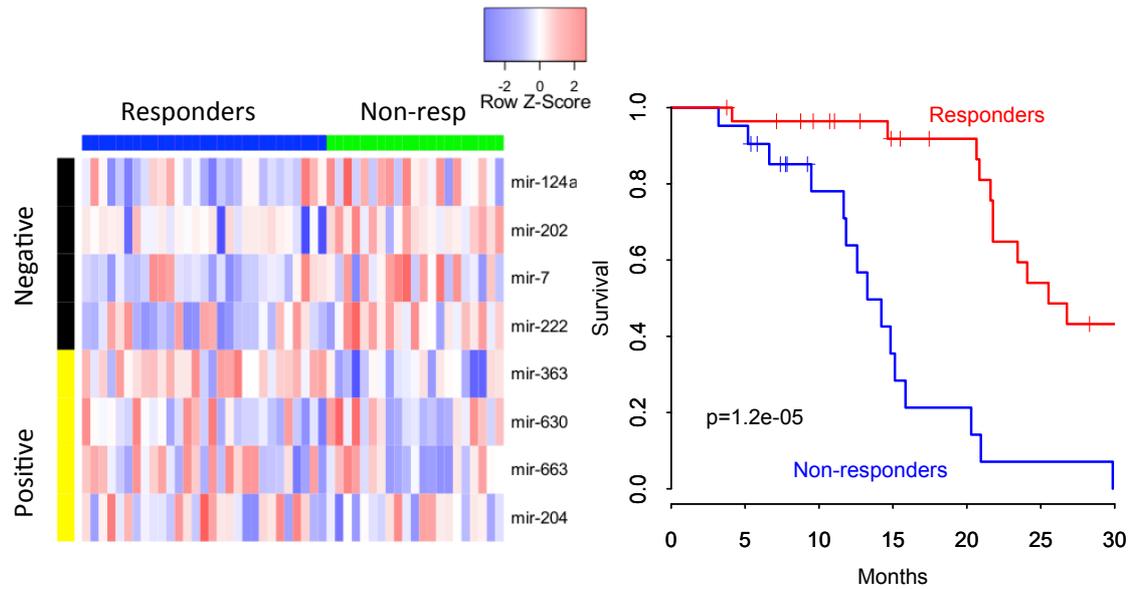


Figure 5.2. The two groups of the training set ($n=50$) defined by using a cut-off value risk score of 0.

The risk score is based on the expression of the eight microRNAs and the cut-off to generate responder and non-responder groups is 0, as defined by assessment of each minimum density in Table 5.4. The log rank of these two groups showed a significant survival benefit to the responders. The median survival of the responder group is 22 months, compared to a non-responder survival median of 12 months. The heatmap shows high microRNA expression in red and low microRNA expression in blue. Positive microRNAs are those that are higher in tumours of patients with poorer prognosis, and negative microRNAs are the vice versa. The responders show lower expression of negative microRNAs and higher expression of positive microRNAs.

Spearman's correlation of treatment length with survival time showed that the responders showed a correlation (correlation coefficient=0.48, $p=0.01$) whereas the non-responders did not (correlation coefficient=0.36, $p=0.12$). Multivariable Cox regression of the risk group and age showed the risk group to be an independent predictor of survival irrespective of age (group HR=0.11, CI=0.04-0.29, $p=5.4e^{-6}$, age HR=1.03, CI=1.00-1.06 $p=3.3e^{-2}$).

5.3.3. Assessment of the signature in the test group of 37 patients.

Risk scores were calculated for the 37 test set patients using the risk score algorithm and the cut-off of 0 was used to separate into responder and non-responder groups. A one-sided log rank test was performed (one-sided because this tests the hypothesis that the responders respond *better* than the non-responders, and not the other way round). This produced a group of 18 responders, with median survival 21 months and a group of 19 non-responders with a median survival of 15 months. The log rank test showed that the responders survive significantly longer than the non-responders (HR=0.34, CI=0.11-1.01, $p=0.026$, Fig. 5.3). Multivariable Cox regression showed that age was not a prognostic factor in the test set group (HR=0.99, CI=0.94-1.04, $p=0.69$), yet the risk group was a prognostic factor (HR=0.33, CI=0.11-0.99, $p=0.049$).

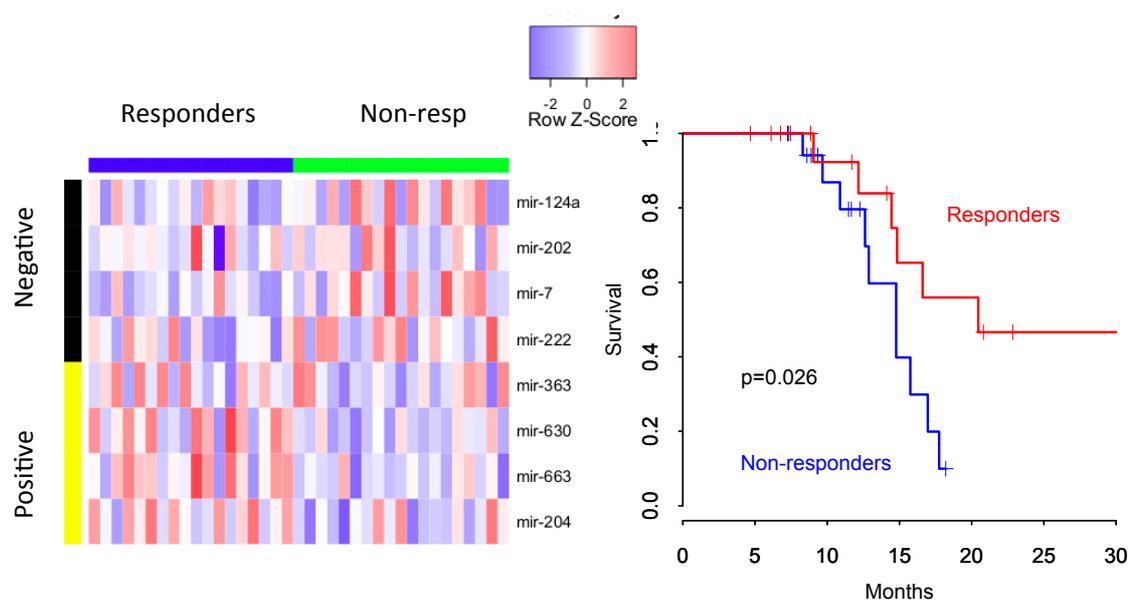


Figure 5.3. The two groups of the test set ($n=37$) defined by using a cut-off value risk score of 0.

The risk score is based on the expression of the eight microRNAs to generate responder and non-responder groups with a cut-off of 0. The one-tailed log rank of these two groups show a significant survival benefit to the responders ($p=0.026$). The median survival of the responder group is 21 months, compared to a non-responder survival median of 15 months. The heatmap shows high microRNA expression in red and low microRNA expression in blue. Positive microRNAs are those that increase in tumours of patients with better prognosis, and negative microRNAs are the vice versa. The responders show lower expression of negative microRNAs and higher expression of positive microRNAs.

Spearman's correlation of treatment length with survival time showed that neither the test group responders nor non-responders showed a correlation (responder correlation coefficient=0.07, $p=0.80$, non-responder correlation coefficient=0.44, $p=0.12$). When the test and training sets were combined, the responder group showed a significant correlation between survival and treatment length (correlation coefficient=0.33, $p=0.04$) whereas the non-responder group did not (correlation coefficient=0.31, $p=0.08$). Although the

survival of the responder group significantly correlates with treatment length and the non-responder groups do not, it must be noted that there is not a great difference in correlation coefficient (0.33 versus 0.31 respectively) and p-value (0.04 versus 0.08 respectively) for these tests. In addition to correlation studies, Cox regression was performed on the responders and non-responders from the 87 patients combined from the test and training sets. This confirmed that the responders show an improved survival with longer treatment length (HR=0.995, CI=0.99-1.00, p=0.01) yet the non-responders do not (HR=0.997, CI=0.99-1.0, p=0.19).

5.3.4. Testing of the signature across all the patients in the TCGA, treated with different treatment types.

The patients tested so far have all been treated with bevacizumab and the signature appears to stratify patients into responder and non-responder groups. In order to test whether this is predictive of patient outcome in general or in response to bevacizumab, I calculated the signature for all 562 patients in the TCGA treated with varying drugs and regimes (Table 5.2). These patients included the 87 patients treated with bevacizumab. This resulted in 305 in the responder group (median survival 10 months) and 257 in the non-responder group (median survival 8 months). A log rank test between the groups showed this signature is prognostic in this group of patients yet is less powerful than in the group of patients treated with bevacizumab (HR=0.73, CI=0.61-0.89, p=0.0016, Fig. 5.4). If a patient is stratified to the responder group and treated with bevacizumab, they have a 75% (average of test and training set hazard ratio) better chance of responding than if they were stratified to the non-responder group. If a patient is stratified to the responder group and treated with a non-specific treatment, they have a 27% better chance of responding than if they were stratified to the non-responder group. This indicates the signature is predicting response to bevacizumab specifically.

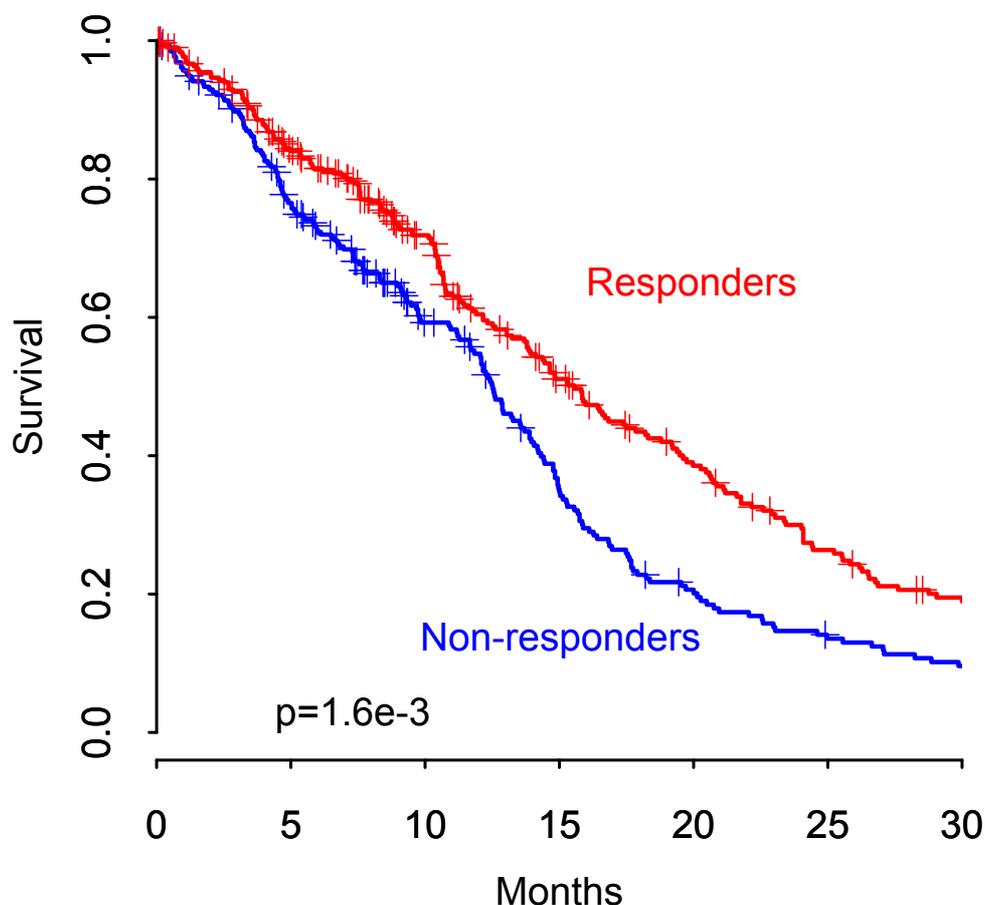


Figure 5.4. The responder and non-responder groups of the whole of the glioblastoma patients in the TCGA as defined by the signature, regardless of treatment.

The risk score is based on the expression of the eight microRNAs and the cut-off to generate responder and non-responder groups is 0. Log rank of these two groups show a significant survival benefit to the responders ($p=0.0016$). The median survival of the responder group is 15.6 months, compared to a non-responder survival median of 12.5 months.

5.3.5. Characterisation of the two groups defined by the signature.

Previous signatures have shown that the mesenchymal subtype does not respond well to bevacizumab (Colman et al., 2010). Therefore I assessed whether the molecular subgroup is associated with the risk score by determining the proportions of each molecular glioblastoma subtype in the responder and non-responder groups and assessing for difference using Fisher's exact test (Table 5.4 A). This showed that there were significantly less

mesenchymal type tumours in the responder group ($p=0.041$). The other subtypes do not show a significant difference in the responder and non-responder groups.

The cohort of bevacizumab treated patients includes different treatment start points during the course of a patient's disease. These include adjuvant, progression and recurrence treatment. It is plausible that the particular timing for bevacizumab treatment in the patient's treatment course may affect their survival time and thus be reflected in the risk group. To test this, I assessed the treatment regimes in each of the responder and non-responder groups. There was no significant difference in any treatment regime between the two groups when tested with Fisher's exact test (Table 5.4 B).

As mesenchymal tumours are associated with more extensive infiltration and necrosis, percentage of tumour cells, stromal cells and necrosis were correlated with risk score in order to assess whether the microRNA expression is affected by any of these histological features (Table 5.4 C). There was no significant correlation between any of these features and risk score indicating the microRNA expression signature does not solely reflect any of these features.

A) Molecular subtypes

	Classical	G-CIMP	Mesenchymal	Neural	Proneural	Missing
Responders	11	2	11	9	9	5
Non-responders	4	1	18	5	8	4
Fisher test p-value	0.154	1.000	0.041	0.56	1.000	

B) Treatment regimen

	Adjuvant	Progression	Recurrence	Missing
Responders	8	24	5	10
Non-responders	8	21	2	9
Fisher test p-value	0.786	1	0.445	

C) Histological features

	Rho	p-value
Tumour cells	-0.11	0.52
Stromal cells	-0.15	0.36
Necrosis	0.22	0.19

Table 5.5. Assessment of the characteristics in the responder and non-responder groups in the 87 bevacizumab treated patients in the TCGA.

The 87 patients were separated into responders and non-responders according to the algorithm and the two groups were assessed for enrichment of particular characteristics using Fisher's test. A) Each molecular subtype was compared in the responder and non-responder groups to determine whether any particular subtype is more or less prominent in responders. This showed that there were fewer mesenchymal tumours in the responder group. B) The treatment regimens were compared between groups to determine whether the signature is discriminatory based on treatment regimen. This showed no significance, which indicates the signature is not associated with survival due to the timing of bevacizumab administration relative to a patient's disease course. C) Risk score was assessed for correlation (Spearman's correlation) with histological features that are reminiscent of a mesenchymal subtype; percentage tumour cells, percentage stromal cells and percentage necrosis. There were no significant correlations between risk score and these features ($p=0.52$ for tumour cells, $p=0.36$ for stromal cells, $p=0.19$ for necrosis).

5.3.6. Ability of the signature to predict progression free survival.

In the RTOG 0825 and AVAglio trials progression free survival improved yet overall survival did not. As overall survival and progression free survival are usually correlated, a signature score is likely to predict both. I tested whether the 8-microRNA signature predicts progression free survival using Cox regression in both the test and training sets. The training set showed that the signature score predicts PFS (HR=2.67, CI=1.42-5.03, p=0.0024) and the two groups separated by a cut-off of 0 have significantly different PFS (HR=0.48, CI=0.25-0.93, p=0.029). In the test set the risk score predicted PFS (HR=2.27, CI=1.02-5.04, p=0.045) but the risk group did not (HR=0.58, CI=0.29-1.13, p=0.11). This may be because the test set is much smaller and when further separated into subsets it loses significance.

5.3.7. Assessment of the signature in NCI cell lines.

In the absence of an independent bevacizumab-treated patient cohort, I attempted to perform a validation using bevacizumab-treated cell lines. The NCI have a panel of 60 cell lines that have drug response information (Shoemaker, 2006). Unfortunately, bevacizumab was not included in the drug screen. However, Hein and Graver assayed apoptosis levels in eight NCI (non-glioblastoma) cell lines after 48 hours of treatment with/without bevacizumab using western blotting with an antibody against cleaved Poly [ADP-Ribose] Polymerase (PARP) (Hein & Graver, 2013). This showed a non-significant increase in apoptosis in two cell lines when treated with bevacizumab compared to a control; A498 (renal cell carcinoma) and HS-578T (breast cancer). My hypothesis was that the cell lines that showed an increase in apoptosis after bevacizumab treatment would be categorised as responders using my signature. I used calculated risk scores using the mean-centred microRNA expression data on treatment naïve cell lines from the NCI and plotted a density plot as with the TCGA data (Figure 5.5). This showed that there was a single minimum at 0; therefore 0 was again used as a cut-off (although this doesn't relate to 0 in the TCGA data which was quantile normalised and not mean-centred).

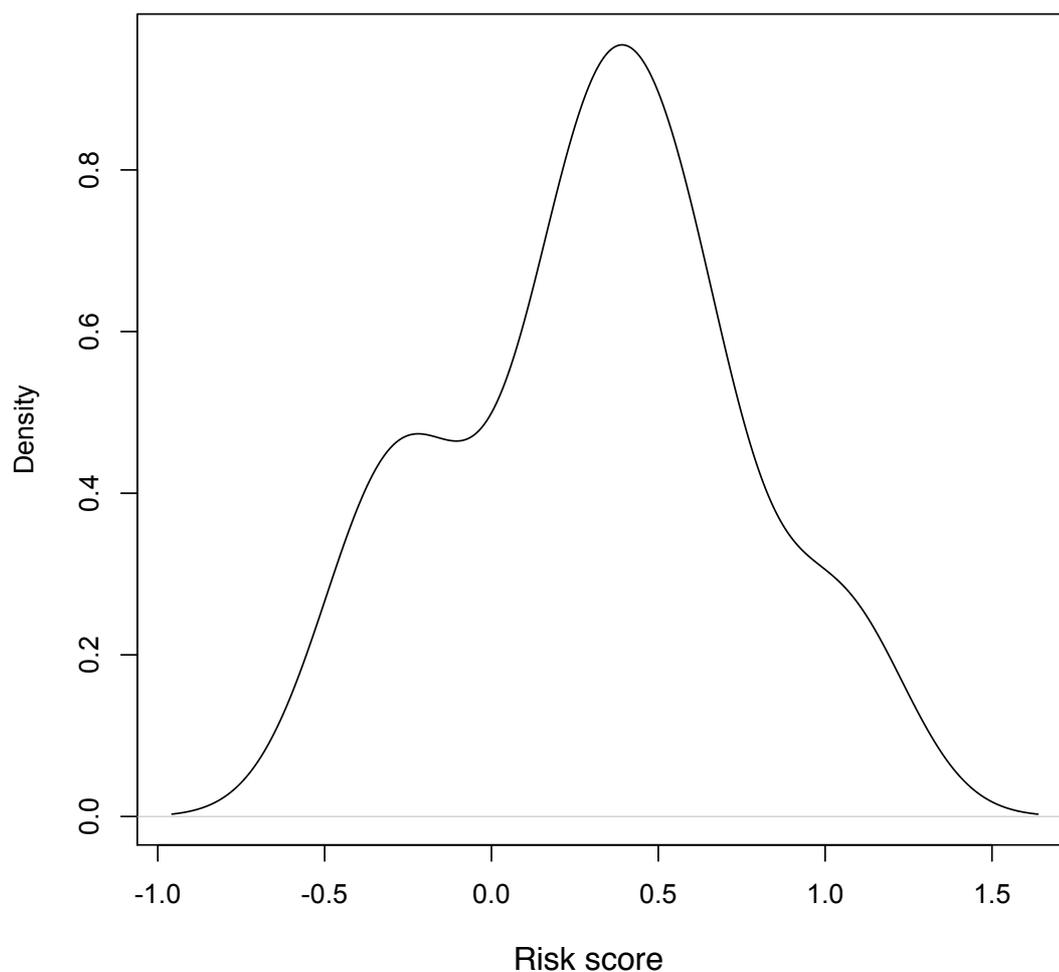


Figure 5.5. Density plot of the risk score in eight NCI cell lines (miR-124a was not included in this risk score).

The risk score was calculated using the signature algorithm without miR-124a. The Y-axis represents the density of each score in the 8 cell lines and the X-axis represents the risk score, as defined by the expression of the 7 microRNAs. The minimum at a risk score of zero was used as a cut-off for the responder and non-responder groups.

This produced a group of two cell lines in the responder category; H522 and A498 and six cell lines in the non-responder category (Fig. 5.6).

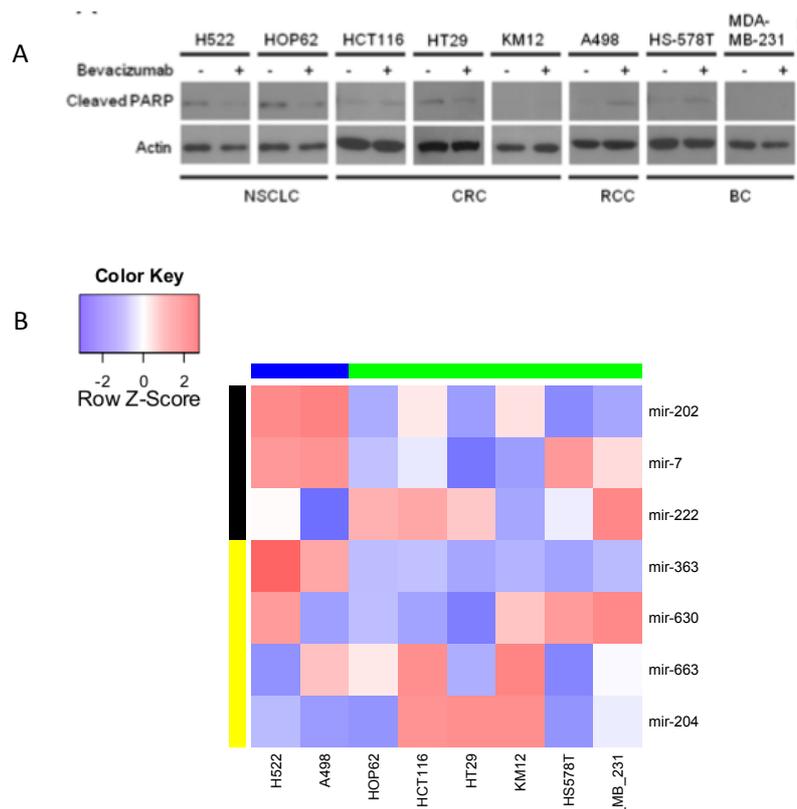


Figure 5.6. The microRNA expression of the 7 microRNAs (original signature minus miR-124a) in eight NCI cell lines.

A) The results of the western blotting performed by Hein and Graver (Hein & Graver, 2013) for cleaved PARP. A498 and HS-578T showed a minor increase in cleaved PARP following bevacizumab exposure. Two lung adenocarcinoma cancer cell lines; H522 and HOP62, both already show presence levels of cleaved PARP in the untreated cell lines, show a slight decrease in cleaved PARP with bevacizumab. B) Heatmap of the signature microRNA expression in the cell lines. The two cell lines stratified by the signature to the responder group were H522 and A498. A489, defined as a responder, is one of the cell lines that showed increased apoptosis following bevacizumab treatment. Interestingly, two of the negative microRNAs, miR-7 and miR-202, are higher in the responders to other samples, yet the score overall still defined them as a responder.

5.4. Discussion.

This study has identified 8 microRNAs that, when combined in a signature, are associated with prognosis in bevacizumab treated glioma patients. The expression of these microRNAs can be incorporated into a risk score that when split at a defined cut-off, separates patients into groups with significantly different survival times. This is more significant in patients treated with bevacizumab than in patients treated with other treatments. This risk score, with further validation, could be used as an indicator at diagnosis to determine whether a patient will respond, at any time during their disease, to bevacizumab. Validation is required ideally in a prospective trial, and at least retrospectively from a previous trial to determine whether this signature predicts response in an independent cohort. The trial would require collection of extensive clinical data and samples in a control arm treated with standard treatment and a treatment arm treated with standard treatment plus bevacizumab. The majority of the patients in this study were treated at progression, and therefore a trial that assesses the signature at diagnosis and starts bevacizumab treatment at progression would be ideal.

As a form of validation, I calculated the signature in the eight cell lines from different cancers from the NCI that had been tested for bevacizumab associated apoptosis (Hein & Graver 2013). This is not ideal because they are not glioblastoma cell lines, and also miR-124a could not be included in the signature, however it serves as a form of testing of the signature in the absence of other options. The cell line microRNA expression performed by the NCI and the cell lines used in the apoptosis assessment would have been of different passages. It must also be noted that tumour cell apoptosis, although an indicator of response, is not the major effect of bevacizumab. Bevacizumab can alter the tumour vasculature, which is likely to improve patient outcome (Field et al. 2014). Despite these problems, one of the two cell lines defined as a responder (A498) showed increased apoptosis after bevacizumab treatment, which is encouraging. This may suggest that my signature is a reflection of tumour response. Vascular response may also be a factor that is involved in the signature, especially as some of the microRNAs are in angiogenic pathways.

Considerable further work is required however to confirm the relevance of the responses shown in the cell lines.

The major outcomes from the clinical trials of bevacizumab in newly diagnosed glioblastoma (RTOG 0825 and AVAglio) were that bevacizumab increases PFS and not overall survival. If a signature is to be viable, the main aim of that signature is to categorise patients according to overall survival. The signature I have designed in this study predicts overall survival with the most power. The prediction of PFS was not significant in the test set, possibly because it is only a small group of 37 patients, and further splitting of this into two smaller groups reduces the chance of gaining significance. However, when the test and training sets were combined, PFS is significantly better in the responder group. Another indicator that a signature is performing well is to test whether an increase in treatment length is associated with a longer survival within the responder group but not the non-responder group. Again this was not significant in the test group of patients but when the two groups were combined there was a significant improvement in survival with longer treatment in the responder group. As expected there was no effect in the non-responder group.

Of the eight microRNAs identified in the signature, seven are involved in angiogenesis. The microRNAs that have a positive weight in the signature (negative microRNAs), and therefore are lower in responders (who have a risk score less than the cut-off), are likely to be anti-angiogenic. This is deduced because responders should have more angiogenesis than non-responders, due to their response to an anti-angiogenic drug. Of these positively-weighted microRNAs, miR-124a has been shown to transcriptionally decrease VEGF (vascular endothelial growth factor) through RAS (rat sarcoma) signalling (Shi et al. 2014b) and miR-222 is considered one of three most important anti-angiogenic microRNAs in coronary artery disease (Zhang et al. 2011).

Overexpression of miR-7 in a neuroblastoma mouse model significantly reduced angiogenesis and overexpression in endothelial cell lines decreased tube formation and sprouting (Babae et al. 2014). Of the microRNAs that were negatively weighted, miR-363 and miR-663 are reported to have pro-angiogenic properties; miR-363 has been shown to improve endothelial cell angiogenic properties and endothelial interaction with haematopoietic precursors (Costa et al. 2013) and miR-663 indirectly increases VEGF and promotes angiogenesis.

The reports on these microRNAs are concordant with their effect in the signature. However, miR-204 and miR-630 show anti-angiogenic properties and are negatively weighted microRNAs; miR-204 directly decreases VEGF and also targets angiopoietin-1 (Zhao et al. 2014; Kather et al. 2014), and miR-630 has been shown to be induced by the anti-angiogenic protein angiopoietin-like protein 1 (Kuo et al. 2013).

Previous studies have suggested that patients with glioblastoma of the mesenchymal subtype respond less well to bevacizumab (Sulman et al. 2013). There are fewer patients with tumours of mesenchymal molecular glioblastoma subtype in the responder category, which is in agreement with previous findings. However, of all the mesenchymal tumours, 38% still stratified to the responder group, which suggests that the signature is not simply predicting a mesenchymal subtype.

A number of biomarkers have been proposed for the response to bevacizumab, including protein biomarkers and imaging features, these could be incorporated into a signature such as this for a highly robust stratification (Field et al. 2014). To drive this forward, access would be required to samples and data from patients treated with bevacizumab, as indicators that these biomarkers may be successful before embarking on a prospective clinical trial. This, as well as a consensus on strategies for biomarkers for the drug, has been urged upon by a number of professionals in the field and is a continuing need (Fine 2014; Prados 2014).

6. Discussion.

'Whoever is careless with the truth in small matters cannot be trusted with important matters' A. Einstein.

6.1 MicroRNA expression and its association with prognosis.

This work began from a standpoint where it was accepted that microRNAs have roles in gliomagenesis. MicroRNAs are also recognised to have a prognostic, and potentially predictive value in glioma. Despite this, few studies have robustly demonstrated that microRNAs with prognostic potential affect biological processes in the tumour. When microRNAs have been identified to be prognostic through simple correlations with survival it is entirely plausible that the prognostic ability is due to a factor that is not associated with the tumour biology. Thus, in some cases the microRNAs may be predictive of clinical factors that could be important, but are not associated with the actual biology of tumour cells. For example, the presence of inflammatory cells that may have different endogenous microRNA expression, and patient age that may also be a confounding factor (Zhu et al., 2012b; Bozdag et al., 2013). The presence of these cells is obviously an important aspect of the tumour biology and may have a bearing on prognosis. On the other hand it is likely that some microRNAs are prognostic due to their direct role in glioma biology. This is likely since it has previously been shown that microRNAs have roles in differentiation, development, stem cells and response to the environment (Letzen et al., 2010; Godlewski et al., 2010a; Kim et al., 2011b).

I have shown in Chapter two that microRNA expression levels are associated with glioma prognosis, and when combined the expression of these microRNAs robustly predicts survival. The combined microRNA expression score is predictive of age, but also independently predictive of prognosis. Other factors were also assessed, such as KPS, extent of resection and important molecular features, some of which were associated with survival but in each case, the combined expression score was independently prognostic. Chapter three revealed that the microRNA expression in tumours reflects different stages in differentiation of the cells altered in glioma, and that the more the microRNA expression mirrors certain of these cells, the poorer the patient outcome. In this way, the microRNAs appear to provide information on the differentiation state of

the tumour sample as a whole, or provide an indication of the number of less differentiated cell types present in a tumour sample. The prognostic data for miR-9 in Chapter four initially conflicted with the literature on this microRNA, however there are plausible clinical and biological explanations for this (Schraivogel et al., 2011; Munoz et al., 2013; Munoz et al., 2014). I have confirmed that miR-9 does indeed have oncogenic potential in glioblastoma cells, by increasing their migration, and a possible reason for the conflict with prognostic data is that miR-9 is expressed at a lower level in tumours with more infiltration, hypoxia and necrosis. These are all features of mesenchymal glioblastoma, a particularly aggressive form of the disease (Verhaak et al., 2010; Mao et al., 2013). These results reflect the fact that the study of glioblastoma using *in vitro* assays, with cell lines, does not fully depict the tumour, and other factors must be taken into consideration. These factors include extreme conditions in the tumour environment, or infiltration with other cells that are lost when cells are expanded in culture. In this regard, microRNAs can provide biologically relevant information on a tumour sample. This is potentially easier to do with microRNAs than mRNAs, since microRNAs have multiple targets often involved in similar pathways.

6.2. The use of microRNAs to determine the ultimate 'aim' of the cell/tumour.

Pathway analysis of the microRNAs involved in prognosis was performed in Chapters two, three and four of this thesis and all pointed to pathways with prominent roles in gliomagenesis. In Chapter two, microRNAs involved in prediction of prognosis in glioblastoma include MAPK and WNT signalling as well as adhesion and polarity. In Chapter three prognostic microRNAs in the whole of malignant glioma pointed towards AKT and IGFR signalling as well as adhesion and differentiation. In Chapter four, pathways predicted to be targeted by prognostic miR-9 involved VEGFR, TGF, PKA and WNT signalling as well as cell adhesion pathways. It is important to note that pathway analysis can be biased towards well studied pathways, and this coupled with the high false positive rate of target prediction analysis suggests these results should be interpreted with caution. However, in combination with the overexpression studies of miR-9, and the alignment of microRNA expression with cell types in

the differentiation pathway of astrocytes, I have provided evidence that they are biologically relevant as well as prognostic.

6.3. MicroRNAs in the clinical setting.

The reason I chose microRNAs for these prognostic studies was, besides the fact they offer an alternative to the more studied mRNAs, they are more stable in biological samples (Hall et al., 2012). This not only allows more samples to be tested (even samples older than ten years, or those poorly handled) but it also indicates that quantitation of their expression is likely to be more robust across samples. One drawback of the study of microRNAs is that the methods used for quantification are less well developed. Microarray analysis still seems to be a popular choice over the multiple sequencing methods, which can be complicated and time-consuming for both the library preparation and bioinformatics (Pradervand et al., 2010; Ach et al., 2008). Sequencing methods are more expensive and labour intensive for microRNAs than mRNAs. The principle involves enrichment of the small RNA fraction with a 5' monophosphate (suggesting it has been processed by Dicer) to ensure sufficient read depth of the microRNA sequences is obtained in relation to all other RNA in the sample (Hafner et al., 2012). In my experience, this enrichment was poor, and only a small fraction (5-30%) aligned to microRNA precursor sequences in a microRNA sequencing trial using the NEBNext[®] microRNA library preparation method in the Next Generation Sequencing facility in our institution. Custom library methods have shown that the ligases used in these preparations are also critical and some may neglect or have a preference for microRNAs of a certain secondary structure (Sorefan et al., 2012). Additionally the RNA integrity value (RIN), which has been generated using neural networks and adaptive learning on total RNA samples, is not necessarily useful for the integrity of small RNA, which degrades less than mRNA (Schroeder et al., 2006). The RIN is still of importance in microRNA profiling studies because the samples with a low RIN may have more mRNA sequences in the small RNA fraction due to degradation.

6.4. How I have adapted my work for eventual introduction to the clinic.

Profiling of microRNAs for mining of prognostic microRNAs requires different methods to those for clinical assessment. Profiling requires a cost-effective assay which is accurate to quantify as many microRNAs as possible in the sample, whereas clinical assessment requires a less labour intensive, economical test with fewer microRNAs profiled. In Chapters two and five, where I have attempted to generate signatures of microRNA expression for clinical assessment I used the LASSO algorithm because it performed a selection on the microRNAs and so only few (nine in the prognostic signature in Chapter two and eight in the predictive signature in Chapter five) were included in the final algorithm. This means that assays that profile a small number microRNAs with fewer transfer steps (less chance of mix-ups in a clinical setting) and less labour requirement can be used. The LASSO algorithm allowed all microRNAs to be used in generation of the model, in contrast to Srinivasan *et al* and Sana *et al* who used preselected microRNAs for generation of the model. Additionally, I included all glioblastoma samples, rather than generating different models for different molecular subtypes, as performed by Li *et al* (S. Srinivasan et al., 2011; R. Li et al., 2014b). Both these studies, as well as the signature generated by Sana *et al* (Sana et al., 2014), used univariate Cox regression to assess the microRNAs for association with survival. This does not take into account the effects of microRNAs collectively. It is also likely that a large number of microRNAs would be identified with an expression pattern that correlates with patient outcome leaving too many predictors for a valid clinical signature. This was reduced in the Li *et al* study by permutation tests. The LASSO method therefore generates a very different model than those previously created, where not all prognostic microRNAs are included, yet the ones that are included reflect all the microRNAs with prognostic potential. This is shown by the fact that alone some of the microRNAs in my signature are not associated with survival by univariate cox regression.

I attempted to validate the signatures using methods that could be used in a diagnostic laboratory, and in Chapter two this was done using qRT-PCR. Other methods that could be used may be Luminex, or Nanostring nCounter, both of

which require little labour requirement. This was a small cohort of only 20 glioblastoma samples and a larger cohort would have been more optimal. This patient group collectively had a poorer survival than the TCGA dataset, and as a different platform had been used it caused problems as to where to cut-off the score into groups of high and low risk patients. It was not appropriate to use the median, because it would be expected that more patients should fall into the high-risk group. I decided to evaluate the groups based on how many patients survived in relation to the conventional median of 450 days. As there were 10% more patients surviving less than the conventional median in the validation cohort compared to the cohort I used to generate the model, the 60th percentile was used. Despite the differences and small number of patients, the signature had the ability to separate patients into high and low risk groups, which is encouraging. This therefore represents a validation using an independent sample set, taken from a different population, using a different platform. These are all suitable validation models for validating a prognostic signature (Altman et al., 2009). Taken with the fact that in the lower grade TCGA dataset of sequencing data the signature also stratified patients into significant high and low risk groups, this is highly promising. The signature stratified patients from all molecular subtypes of glioblastoma except the G-CIMP negative proneural group. This group of patients was defined by Brennan *et al* to be the group with the poorest prognosis of all glioblastoma subtypes, and in the original study by Verhaak *et al* the proneural group is particularly refractory to treatment (Verhaak et al., 2010; Brennan et al., 2013). It would be expected that most of these patients should be stratified to the high-risk subgroup of patients but there were still 39% of proneural G-CIMP negative patients in the good prognosis group. Stem cell signatures for glioblastoma have indicated that the proneural G-CIMP negative group have the most similar expression patterns to stem cells compared to other glioblastoma subtypes (Patel et al., 2014). Also, the presence of heterogeneity at the single cell level within the proneural G-CIMP negative tumours confers a worse prognosis (Patel et al., 2014). In the case of a stem cell signature, you would expect that the more closely a tumour mirrors a stem cell expression pattern (due to the presence of more stem cells) then a patient would have a worse outcome. However, in the case of proneural non-G-CIMP tumours it seems the opposite is true. This may be a reason for the

conflicting results. Interestingly, when the microRNA expression patterns of the glioma samples, rather than single cells, are correlated with oligodendrocyte precursors as in Chapter three, the mesenchymal subtype shows a higher correlation than the proneural subtype. It is likely that the 'stem cell' is different in a proneural sample to a mesenchymal sample, as is shown when miR-9 was quantified in proneural and mesenchymal GSCs in Chapter four.

My nine-microRNA signature is clinically relevant in the fact that it predicts prognosis similar to, or better than, *MGMT* promoter methylation, and also has the added advantage that it stratifies patients regardless of their treatment regime. Stratification of patients will allow the high-risk patients to be more closely monitored, and if eligible, may be suitable for clinical trials for novel treatments. The cut-off used in my study was arbitrary, defined by the median, to show proof of the algorithm, but in practice cut-offs may be decided based on availability of clinical resources.

LASSO regression was again used for generation of a signature predicting response to bevacizumab in Chapter five. The data I used in this chapter was based on a merged value for each microRNA based on all the probes for that microRNA signature, which differed from Chapter two where I used actual probe values. This was because in Chapter two, the aim was to produce a signature suitable for a clinical test, and this can in theory be extrapolated to any platform based on the sequences of that probe. The aim of Chapter five, was to determine in principle whether microRNAs can predict the response of a drug, and whilst the necessities of a clinical test were taken in to consideration, the microRNA expression levels themselves were considered more appropriate here. Bevacizumab is not an easy drug to evaluate, as was shown by the trial designs. The drug was administered at different times during treatment in my dataset, each patient received different numbers of cycles of the drug and determination of response is difficult without imaging data. My aim was to determine whether, during the course of a patient's treatment, bevacizumab administration could be of benefit. This could be at any time, for an unspecified number of cycles. With this in mind it was considered optimal to use the entire patient treatment time (date of diagnosis to death) as the overall outcome. These microRNAs, when combined in a similar way to the prognostic microRNAs in Chapter two, stratified patients significantly into 'responders' and

'non-responders' in both test and training sets. Ideally, validation would be performed in another, independent dataset. I found it difficult to obtain further samples for validation due to relatively few patients having been treated with bevacizumab, which is why the dataset was split into training and test sets for generation and testing of the model. Samples for an independent validation are still being sourced, and ideally this will be performed using a test suitable for a clinical setting.

I chose to study microRNA expression-based prediction of response to bevacizumab in particular because it is clear that some patients do respond and therefore the drug has clinical utility despite its lack of success at clinical trial. A ten-gene signature was tested on the patients in the bevacizumab treated arm in the RTOG 0825 trial and results were reported at ASCO (American Society of Clinical Oncology) in 2013 (Sulman et al., 2013). These showed that the ten-gene signature that was linked to a more mesenchymal genotype predicted poor response to the drug. It is difficult to compare their signature with my eight-microRNA signature because they didn't provide survival information on groups defined by a cut-off, only for the proportional hazard of the score. Their ten-gene signature was generated using preselected genes, which are associated with a mesenchymal subtype of glioblastoma, and may suggest that this subtype has a poor response to the drug. The association of mesenchymal glioblastoma and poor bevacizumab response is supported to an extent by my results as more patients with mesenchymal glioblastoma were stratified to the non-responder group. There are however, still 11 of 29 patients with a mesenchymal subtype of glioblastoma who were stratified to the responder group using my signature. Compared to the Sulman *et al* study, which included 234 cases, my study is relatively small and separation into test and training sets made it even smaller (Sulman et al., 2013). It is possible that with a larger cohort more significance could be gained. My study also includes patients treated at recurrence and progression, which are not included in their model as it is based on data from the trial for newly diagnosed glioblastoma. This may generate more noise in my data as patients may have been treated with different regimens at initial diagnosis. The inclusion of patients treated at diagnosis and recurrence may be an advantage however, as my signature has the ability to predict response to bevacizumab without taking these factors into

account and may offer information that proves useful to patient management at any stage in a patient's treatment course.

6.5. The role of prognostic microRNAs in malignant glioma.

The aim of Chapter three was to directly determine what the role of prognostic microRNAs in malignant glioma might be. Unlike Chapters two and five where the number of prognostic microRNAs was minimised for a signature, this study required all prognostic microRNAs to be identified. Following this, target pathways of the microRNAs were inferred and a pathway was chosen to determine whether the pathway as a whole might have a bearing on survival. The study of prognostic microRNAs in the entire group of malignant glioma generated some problems. The data for grade III glioma was expression data from Illumina sequencing; yet the expression data for glioblastoma was microarray data. I decided to initially perform differential expression analysis separately in both groups prior to merging to minimise the effects of artefacts for each platform. This prior analysis however included small subsets of patients with extremely poor or good survival. It could be argued that these extremes are not representative of the grade III glioma and glioblastoma patients as a whole. Because of this, I validated the findings from the pathway analysis by using the whole dataset for correlation of the expression patterns of the tumours with the cell stages in the OP differentiation pathway. The other pathways were not followed up further, and although interesting, were therefore not validated. It could also be argued that it is not clinically relevant to define a group of microRNAs of poor patient outcome in grade III glioma and glioblastoma collectively because the literature implicates different molecular pathways in their aetiology and therefore their prognostic features would also be different. Brennan *et al* showed that a proportion of glioblastoma (the G-CIMP positive) are reminiscent of a lower grade tumour, and these may have similar molecular expression patterns to lower grade tumours, suggesting that a common pathway may exist, in some subtypes at least. Whether glioblastomas with similarity to lower grade tumours should be included as grade IV gliomas or a separate entity in the classification of glioma remains to be decided. Ozawa *et al* indicate that all non-G-CIMP glioblastoma is derived from a proneural-like precursor and this may be an argument that the non-G-CIMP gliomas should be

a separate classification from the G-CIMP ones (Ozawa et al., 2014). In this event, the signature in Chapter two should also exclude the G-CIMP proneural cohort. Revisions of the WHO classifications for glioma have been suggested, but until then, in my opinion, a signature should incorporate all patients in a particular WHO classification for ease of clinical testing (Louis et al., 2014). As part of the correlation analysis of the oligodendrocyte differentiation pathway cell types with the grade II gliomas and glioblastoma, 150 microRNAs were included. Not all microRNAs were included because extremely high correlations were generated when all microRNA expression was taken into account. This is likely because a large proportion of microRNAs may be at zero expression in both cell types, and some microRNAs may be expressed at a baseline level and not be involved in differences between these cell types at all. It is also difficult to compare the expression of microRNAs with few transcripts quantified by microarray in the OP cell types and quantified by sequencing in the grade III gliomas. These microRNAs may not have any reads detected at all in the sequencing experiment which could be due to both an alignment issue or because they were too low to detect at that read depth. Therefore these were omitted. It is recognised however, that some of these microRNAs may be important. Additionally in Chapter three, tumour sample microRNA expression is correlated to cell lines for the cell stages in the OP pathway. MicroRNA expression patterns can be reflective of the cellular microenvironment, and the OP cells were grown in an optimal environment whereas a proportion of the cells in the tumour sample would have been exposed to extreme conditions. This is recognised as a limitation of this study. Despite these limitations, the prognostic microRNAs in Chapter three pointed to the targeting of pathways of OP involvement, and when the changes in microRNA expression across one of these pathways was further studied, it showed that these had links to the survival of patients with glioma.

Whilst it is recognised that using a combination of microRNAs to predict prognosis or drug response is more optimal, some microRNAs alone have very powerful prognostic potential. In the prognostic signature in Chapter two, miR-222 and miR-148a have the coefficients furthest from zero and the highest median increase in the high risk group compared to low risk group. It was miR-9 however that was more consistently prognostic in the lower grades of glioma,

and also showed more connectors when predicted targets that are also prognostic were identified. This may suggest that, although the fold change per monthly increase in survival is lower for miR-9 than that for miR-222 and miR-148a, it is more consistently associated with prognosis than these other microRNAs.

This microRNA is most highly expressed in the brain and therefore you would expect that normal brain tissue would have high levels of miR-9 (Sempere et al., 2004). Additionally, miR-9 increases with differentiation and so you would expect more differentiated brain cells to have more miR-9 (Krichevsky et al., 2006). In glioblastoma, infiltration of lymphocytes and the presence of GSCs in the tumour would suggest that the levels of miR-9 in a tumour sample should be lower than that of a non-tumour sample. You would also expect that the more infiltrating lymphocytes and GSCs present in the tumour sample the poorer the patient outcome. My findings could not confirm that the levels of miR-9 are higher in non-tumours compared to glioblastoma, most likely due to few samples that are unmatched, in the analysis. Also, the levels of miR-9 are high in both tissues, and the changes between the two groups are minimal. However, I have shown that miR-9 levels are indeed lower in samples of glioblastoma tumours that are from patients with a poorer outcome. These levels also correlate with the estimated percentage of tumour cells in a sample. This suggests miR-9 levels in a sample are reflective of the cellular composition of the tumour.

To add to this, miR-9 also appears to have an oncogenic role. It has already been shown that miR-9 inhibits STAT3 and suppresses mesenchymal differentiation (Krichevsky et al., 2006; Kim et al., 2011b). I have proven that miR-9 is lower in the mesenchymal group of glioblastomas, both in patient samples and in cultured GSCs. This group of tumours are also known to have more infiltration of immune cells (Engler et al., 2012) and therefore the cellular expression of miR-9 and the multicellular composition result in a magnified effect that miR-9 is lower in samples of this tumour subtype.

When the levels of miR-9 are assessed in proneural and mesenchymal tumours, there is still some overlap (Fig. 4.2). This may indicate that miR-9 cannot perfectly delineate the mesenchymal tumour type. This could be because some tumours have both subtypes present, as shown from single cell

studies (Sottoriva et al., 2013). These could also be tumours that are proneural in nature and some cells are transforming to a mesenchymal subtype, as all non-G-CIMP glioblastomas have been shown to arise from a proneural precursor (Ozawa et al., 2014). This group of proneural non-G-CIMP tumours with high heterogeneity have a poorer prognosis to those without high heterogeneity as shown by the Patel *et al* stem cell signature (Patel et al., 2014). If miR-9 could be used as a measure of the presence and number of cells that deviate from normal brain tissue, this would provide information for patient management. To add to this, miR-9 has been shown through successive generations of *Drosophila melanogaster* that it dampens down selective pressure in proneural networks. The reduction of miR-9 in these organisms was shown to 'free up' the genomic landscape to exert greater phenotypic influence (Cassidy et al., 2013). This suggests it is possible that lower levels of miR-9 may increase the ability of the tumour to adapt to its extreme environment. Such a hypothesis is reflected in my findings where miR-9 decreased when mesenchymal GSC cells were exposed to hypoxia and also tumours with a more necrotic environment had lower miR-9 expression levels. I demonstrated the effect of hypoxia on miR-9 levels using a mesenchymal GSC line, which has amplified EGFR and the increased EGFR activity may be of consequence. The interaction between EGFR and the microRNA RISC loading component AGO2 is enhanced in hypoxia and prevents its binding to Dicer resulting in less mature microRNA present in the cell (Shen et al., 2014). The decreased miR-9 therefore may be as a result of a decrease in the global microRNA level. There are conflicting findings as to the effect of hypoxia on miR-9 expression. miR-9-1 and miR-9-3 have been shown to be under the control of HIF1A (hypoxia inducible factor 1A) which can drive the levels of this precursor up under hypoxic conditions in pulmonary artery smooth muscle cells (Shan et al., 2014). This may increase the level of miR-9 precursor but if the ability to process these precursors is compromised due to the EGFR binding of AGO2, this point is irrelevant. Further to this, mesenchymal cells migrate more when exposed to hypoxia (Joseph et al., 2015). Low miR-9 in hypoxia does not seem to support this. miR-9 does not explain all migration in glioblastoma however, as mesenchymal cells are highly mobile and invasive compared to proneural cells (Mao et al., 2013), yet have low miR-9 expression levels. I have shown that

overexpression of miR-9 increases migration but that does not exclude that other mechanisms may also alter migration, particularly in mesenchymal glioblastoma. Future study could assess the migration patterns of proneural and mesenchymal cells. It has been shown in paediatric glioma that different cell lines migrate with different patterns such as a cogwheel pattern with protrusions from the core as opposed to a sheet like migration by the extension of flattened protrusions more laterally (Cockle et al., 2015). This may shed more light on why mesenchymal GSCs, which have low endogenous miR-9 expression levels, are still migratory and invasive in nature (Mao et al., 2013). It would also be appropriate to assess the effect of miR-9 knockdown on migration of a proneural GSC cell line (which have high endogenous levels of miR-9) *in vivo*, ideally in an immunocompetent mouse model (due to the involvement of miR-9 in predicting infiltration). The hypothesis for this would be that reducing miR-9 levels may result in a tumour with mesenchymal hallmarks of infiltration and necrosis due to increase in factors such as NF- κ B and E-cadherin. Should miR-9 be evaluated further as a marker of heterogeneity, single cell sequencing should be performed as in Patel *et al* with correlation of the levels of heterogeneity with the levels of miR-9 in the tumours (Patel et al., 2014).

The null hypothesis of this study was that microRNAs involved in prognosis are not involved in glioma biology. I have shown here, in Chapters two to four, that these microRNAs target mRNAs involved in glioma biology, they have functional roles in the disease and they are involved in differentiation pathways of the likely cell of origin in glioma. The combined expression of nine microRNAs, eight of which have defined roles in glioma biology, predicted prognosis in glioblastoma better than the current predictive biomarker of *MGMT* promoter methylation. In Chapter three, upon the discovery that the prognostic microRNAs in malignant glioma are predicted to target pathways enriched for the differentiation and survival of oligodendrocyte precursors I proceeded to show that the microRNA expression patterns of poorer prognostic gliomas align with those of oligodendrocyte precursors. Following this I showed that, for each individual glioma, a score can be generated as to how closely a tumour mirrors an oligodendrocyte precursor microRNA expression pattern, and that this score is itself prognostic. miR-9 has been a large part of this thesis, and in Chapter two and four I showed that this microRNA has the ability to reduce the

expression of mRNAs involved in glioma biology. In Chapter four I demonstrated that miR-9 has a functional role in glioma in the ability to increase migration and is also altered in response to the extreme changes in a glioma tumour on a more broad level. This microRNA is prognostic in grade II, III and IV glioma and is highly connected to prognostic mRNAs in Chapter two. These findings suggest that alone, or in combination, the prognostic microRNAs in glioma have a role in glioma biology. The reasons why a microRNA may be prognostic are many. It may be a marker of a particular phenotype of tumour such as its aggressiveness, invasiveness and migration capacity, or it could be due to the tumours response to current treatments, by targeting drug transporters for example. The microRNAs may be markers of particular types of cells in the tumour, e.g., immune cells, lymphocytes and stem cells that confer a survival advantage to the tumour. They may also mark the cells responding to a particular environmental feature, such as high pressure in the cranium, hypoxia and necrotic surroundings. Additionally, and already mentioned, the microRNAs may be providing information on features of the patient, such as their age or gender.

6.6. Limitations of this work.

The results generated in this thesis relied on the availability of large amounts of data on both microRNAs and mRNAs, in large groups of patients with comprehensive clinical information. Most of the data I have used has been publically available from TCGA. This resource has large numbers of patients, with many different data types as well as varying levels of normalisation of the data. It is clear that this portal is invaluable for the study and validation of genetic and protein studies but it does have its limitations. The huge volumes of data, with multiple samples for each patient in some cases, mean that the possibility of inaccurate information is high. The histo-pathological data, such as the percentage of necrosis, tumour cells and stromal cells in a sample is subjective therefore correlations involving this data should be interpreted with caution. The genetic data is also subject to human error- the departments providing this high throughput analysis are meticulous in their delivery but it is inevitable with this level of processing that some mistakes will be made. This is what makes validations in other sample sets important.

A further limitation involves sampling, and in a disease named 'glioblastoma multiforme' because of its extensive heterogeneity, this is even further exaggerated. It has already been proven that multiple molecular subtypes may be present in one tumour and therefore taking a sample of a tumour for molecular profiling may be confounded by the sampling technique (Sottoriva et al., 2013). The data I used to characterise each of the tumours according to molecular subtype was also defined using TCGA data (Brennan et al., 2013). It is likely that the samples taken for both mRNA and microRNA profiling were from a similar portion of the tumour, and therefore correlations of the microRNAs expression levels with their tumour subtypes may only be applicable to that small part of the tumour, without taking in to consideration the substructure of the tumour. Additional features that may not be taken into consideration because of this, such as extensive vasculature, necrosis and the invasive edge, may be of importance in patient outcome. Improvements on this may include the involvement of metabolic imaging, so some information on the region of the sampling site could be included, or only samples from extremely aggressive (high Choline to N-acetyl aspartate ratio) are included. Future work for both signatures could include a sample set using these imaging parameters, with samples that have the most clinical utility, such as FFPE sections.

6.7. Conclusion and future perspectives.

MicroRNAs represent an alternative, potentially more optimal, set of genetic markers that can be used to provide information on a tumour. In cases such as glioblastoma, where there are relatively few molecular markers used in the clinic, these offer an expanded predictor set to exploit for clinical management. They are also useful in cases where a patient subset desperately needs to be determined, such as with bevacizumab efficacy. Where mRNA and protein based markers fail, both directly involved with the biology and identified using correlated methods, microRNAs offer an alternative.

In the next 5-10 years glioma clinical management is set to change due to the availability of new drugs for treatment. These may include immunotherapy and different combinatorial therapies. In a disease as heterogeneous as glioblastoma, personalised medicine is extremely important, and the definition of patient subsets for these new therapies is imperative. For each prospective

trial of a new drug, patients should be selected on the most appropriate markers and it is clear from this thesis that microRNAs should be included in this repertoire. Ideally these microRNAs should be identified at the outset of a trial to avoid the problems realised in the introduction of bevacizumab for newly diagnosed glioblastoma.

This thesis confirms the previous work, which suggests microRNAs are predictive of prognosis in glioblastoma and moves the area further into more clinically relevant questions such as prediction of drug response. I have shown, on a global microRNA expression level, and also for miR-9 alone, that these associations are relevant in glioma biology. The new questions are how to include these findings in current patient management. The best protocol should include the information from the microRNA prediction algorithms together with clinical, histo-pathological and other molecular data. Either way, the addition of microRNAs into the current scope for clinical testing in glioma is of considerable advantage.

Appendices

Appendix 2.1. Patient sample collection at UCLan



8 January 2013

Lisa Shaw / Charlotte Bellamy
School of Pharmacy & Biomedical Sciences
University of Central Lancashire

Dear Lisa / Charlotte

Re: STEM Ethics Committee Application
Unique reference Number: STEM 041

The STEM ethics committee has granted approval of your proposal application '**Identification of Biomarkers for Glioma Progression**'.

Please note that approval is granted up to the end of project date or for 5 years, whichever is the longer. This is on the assumption that the project does not significantly change in which case, you should check whether further ethical clearance is required.

We shall e-mail you a copy of the end-of-project report form to complete within a month of the anticipated date of project completion you specified on your application form. This should be completed, within 3 months, to complete the ethics governance procedures or, alternatively, an amended end-of-project date forwarded to roffice@uclan.ac.uk together with reason for the extension.

Please also note that it is the responsibility of the applicant to ensure that the ethics committee that has already approved this application is either run under the auspices of the National Research Ethics Service or is a fully constituted ethics committee, including at least one member independent of the organisation or professional group.

Yours sincerely

Tal Simmons
Chair
STEM Ethics Committee



National Research Ethics Service

Cambridgeshire 1 Research Ethics Committee

Victoria House
Capital Park
Fulbourn
Cambridge
CB21 5XB

Telephone: 01223 597653
Facsimile: 01223 597645

23 February 2010

Dr Timothy P Dawson
Neuropathology
Royal Preston Hospital
Sharoe Green Lane
Fulwood
Preston PR2 9HT

Dear Dr Dawson

Title of the Research Tissue Bank: Archival nervous system tissue and blood collection
REC reference: 09/H0304/88
Designated Individual: Dr Timothy P Dawson

Thank you for your letter of 09 February 2010, responding to the Committee's request for further information on the above research tissue bank and submitting revised documentation.

The further information has been considered on behalf of the Committee by the Sub-Committee.

Confirmation of ethical opinion

On behalf of the Committee, I am pleased to confirm a favourable ethical opinion of the above research tissue bank on the basis described in the application form and supporting documentation as revised.

The Committee has also confirmed that the favourable ethical opinion applies to all research projects conducted in the UK using tissue or data supplied by the tissue bank, provided that the release of tissue or data complies with the attached conditions. It will not be necessary for these researchers to make project-based applications for ethical approval. They will be deemed to have ethical approval from this committee. You should provide the researcher with a copy of this letter as confirmation of this. The Committee should be notified of all projects receiving tissue and data from this tissue bank by means of an annual report.

Duration of ethical opinion

The favourable opinion is given for a period of five years from the date of this letter and provided that you comply with the conditions set out in the attached document. You are advised to study the conditions carefully. The opinion may be renewed for a further period of up to five years on receipt of a fresh application. It is suggested that the fresh application is made 3-6 months before the 5 years expires, to ensure continuous approval for the research tissue bank.

NHS researchers undertaking specific research projects using tissue or data supplied by a research tissue bank must apply for permission to R&D offices at all organisations where the research is conducted, whether or not the research tissue bank has ethical approval.

Site-specific assessment (SSA) is not a requirement for ethical review of research tissue banks. There is no need to inform Local Research Ethics Committees.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

Here you will find links to the following:

- a) Providing feedback. You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.
- b) Annual Reports. Please refer to the attached conditions of approval.
- c) Amendments. Please refer to the attached conditions of approval.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk

09/H0304/88

Please quote this number on all correspondence

Yours sincerely



Dr Daryl Rees
Chair

E-mail: susan.davies@eoe.nhs.uk

Enclosures: Standard approval conditions

Copy to: R&D Office
Royal Preston Hospital
Sharoe Green Lane
Fulwood
Preston PR2 9HT

Approved documents

The documents reviewed and approved at the meeting were:

<i>Document</i>	<i>Version</i>	<i>Date</i>
Covering Letter		17 November 2009
REC application	Submission Code 37356/76939/3/694	
Applicant's checklist		
Human Tissue Authority Licence	Number 12056	14 February 2008
Protocol for Management of the Tissue Bank	I	17 November 2009
Participant Information Sheet	Version IV	15 February 2010
Participant Information Sheet: Consultee Information Sheet	Version IV	15 February 2010
Participant Consent Form: Collection of Brain Tissue and/or Blood from patients without Brain Tumours	Version IV	15 February 2010
Participant Consent Form: Collection of Nervous System Tumour Tissue and/or Blood for Future use	Version IV	15 February 2010
Participant Consent Form: Consultee Declaration Form - Collection of Nervous System Tumour Tissue and/or Blood for Future use	Version IV	15 February 2010
Participant Consent Form: Consultee Declaration Form - Collection of Brain Tissue and/or Blood from patients without Brain Tumours	Version IV	15 February 2010
Response to Request for Further Information, letter from Dr T P Dawson, Consultant Neuropathologist and Honorary Senior Lecturer		09 February 2010
Email from Dr T P Dawson attaching response to Committee		15 February 2010

Licence from the Human Tissue Authority

Thank you for providing a copy of the above licence.

Research governance

A copy of this letter is being sent to the R&D office responsible for the Royal Preston Hospital. You are advised to check their requirements for approval of the research tissue bank.

Under the Research Governance Framework (RGF), there is no requirement for NHS research permission for the establishment of research tissue banks in the NHS. Applications to NHS R&D offices through IRAS are not required as all NHS organisations are expected to have included management review in the process of establishing the research tissue bank.

Research permission is also not required by collaborators at tissue collection centres (TCCs) who provide tissue or data under the terms of a supply agreement between the organisation and the research tissue bank. TCCs are not research sites for the purposes of the RGF.

Research tissue bank managers are advised to provide R&D offices at all TCCs with a copy of the REC application for information, together with a copy of the favourable opinion letter when available. All TCCs should be listed in Part C of the REC application.

This Research Ethics Committee is an advisory committee to East of England Strategic Health Authority
*The National Research Ethics Service (NRES) represents the NRES Directorate within
the National Patient Safety Agency and Research Ethics Committees in England*

NHS researchers undertaking specific research projects using tissue or data supplied by a research tissue bank must apply for permission to R&D offices at all organisations where the research is conducted, whether or not the research tissue bank has ethical approval.

Site-specific assessment (SSA) is not a requirement for ethical review of research tissue banks. There is no need to inform Local Research Ethics Committees.

Statement of compliance

The Committee is constituted in accordance with the Governance Arrangements for Research Ethics Committees (July 2001) and complies fully with the Standard Operating Procedures for Research Ethics Committees in the UK.

After ethical review

Now that you have completed the application process please visit the National Research Ethics Service website > After Review

Here you will find links to the following:

- a) Providing feedback. You are invited to give your view of the service that you have received from the National Research Ethics Service and the application procedure. If you wish to make your views known please use the feedback form available on the website.
- b) Annual Reports. Please refer to the attached conditions of approval.
- c) Amendments. Please refer to the attached conditions of approval.

We would also like to inform you that we consult regularly with stakeholders to improve our service. If you would like to join our Reference Group please email referencegroup@nres.npsa.nhs.uk

09/H0304/88

Please quote this number on all correspondence

Yours sincerely



Dr Daryl Rees
Chair

E-mail: susan.davies@eoe.nhs.uk

Enclosures: Standard approval conditions

Copy to: R&D Office
Royal Preston Hospital
Sharoe Green Lane
Fulwood
Preston PR2 9HT

This Research Ethics Committee is an advisory committee to East of England Strategic Health Authority
The National Research Ethics Service (NRES) represents the NRES Directorate within the National Patient Safety Agency and Research Ethics Committees in England

CONDITIONS OF ETHICAL APPROVAL

Research Ethics Committee:	Cambridgeshire 1
Research Tissue Bank:	Archival nervous system tissue and blood collection
REC reference number:	09/H0304/88
Name of applicant:	Dr Timothy P Dawson
Date of approval:	23 February 2010

Ethical approval is given to the Research Tissue Bank ("the Bank") by the Research Ethics Committee ("the Committee") subject to the following conditions.

1. Further communications with the Committee
 - 1.1 Further communications with the Committee are the personal responsibility of the applicant.

2. Duration of approval
 - 2.1 Approval is given for a period of 5 years, which may be renewed on consideration of a new application by the Committee, taking account of developments in legislation, policy and guidance in the interim. New applications should include relevant changes of policy or practice made by the Bank since the original approval together with any proposed new developments.

3. Licensing
 - 3.1 A copy of the Licence from the Human Tissue Authority (HTA) should be provided when available (if not already submitted).
 - 3.2 The Committee should be notified if the Authority renews the licence, varies the licensing conditions or revokes the Licence, or of any change of Designated Individual. If the Licence is revoked, ethical approval would be terminated.

4. Generic ethical approval for projects receiving tissue
 - 4.1 Samples of human tissue or other biological material may be supplied and used in research projects to be conducted within the establishment responsible for the Bank and/or by researchers and research institutions external to the Bank within the UK and in other countries in accordance with the following conditions.
 - 4.1.1 The research project should be within the fields of medical or biomedical research described in the approved application form.
 - 4.1.2 The Bank should be satisfied that the research has been subject to scientific critique, is appropriately designed in relation to its objectives and (with the exception of student research below doctoral level) is likely to add something useful to existing knowledge.
 - 4.1.3 Where tissue samples have been donated with informed consent for use in future research ("broad consent"), the Bank should be satisfied that the use of the samples complies with the terms of the donor consent.
 - 4.1.4 All samples and any associated clinical information must be non-identifiable to the researcher at the point of release (i.e. anonymised or linked anonymised).
 - 4.1.5 Samples will not be released to any project requiring further data or tissue from donors or involving any other research procedures. Any contact with donors must be confined to ethically approved arrangements for the feedback of clinically significant information.
 - 4.1.6 A supply agreement must be in place with the researcher to ensure storage, use and disposal of the samples in accordance with the HTA Codes of Practice, the terms of the ethical approval and any other conditions required by the Bank.
 - 4.2 A research project in the UK using tissue provided by a Bank in accordance with these conditions will be considered to have ethical approval from the Committee under the terms of this approval. In England, Wales and Northern Ireland this means that the researcher will not require a licence from the Human Tissue Authority for storage of the tissue for use in relation to this project.
 - 4.3 The Bank may require any researcher to seek specific ethical approval for their project. Such applications should normally be made to the Committee and booked via the NRES Central Allocation System.
 - 4.4 A Notice of Amendment form should be submitted to seek the Committee's agreement to change the conditions of generic approval.
5. Records
 - 5.1 The Bank should maintain a record of all research projects to which tissue has been supplied. The record should contain at least the full title of the project, a summary of its purpose, the name of the Chief Investigator, the sponsor, the location of the research, the date on which the project was approved by the Bank, details of the tissue released and any relevant reference numbers.

5.2 The Committee may request access to these records at any time.

6. Annual reports

6.1 An annual report should be provided to the Committee listing all projects for which tissue has been released in the previous year. The list should give the full title of each project, the name of the Chief Investigator, the sponsor, the location of the research and the date of approval by the Bank. The report is due on the anniversary of the date on which ethical approval for the Bank was given.

6.2 The Committee may request additional reports on the management of the Bank at any time.

7. Substantial amendments

7.1 Substantial amendments should be notified to the Committee and ethical approval sought before implementing the amendment. A substantial amendment generally means any significant change to the arrangements for the management of the Bank as described in the application to the Committee and supporting documentation.

7.2 The NRES Notice of Amendment form should be used to seek approval. The form is available at <http://www.nres.npsa.nhs.uk/applicants/review/after/amendments.htm#notice>.

7.3 The following changes should always be notified as substantial amendments:

7.3.1 Any significant change to the policy for use of the tissue in research, including changes to the types of research to be undertaken or supported by the Bank.

7.3.2 Any significant change to the types of biological material to be collected and stored, or the circumstances of collection.

7.3.3 Any significant change to informed consent arrangements, including new/modified information sheets and consent forms.

7.3.4 A change to the conditions of generic approval.

7.3.5 Any other significant change to the governance of the RTB.

8. Serious adverse events

8.1 The Committee should be notified as soon as possible of any serious adverse event or reaction, any serious breach of security or confidentiality, or any other incident that could undermine public confidence in the ethical management of the tissue. The criteria for notifying the Committee will be the same as those for notifying the Human Tissue Authority in the case of research tissue banks in England, Wales and Northern Ireland.

9. Other information to be notified
 - 9.1 The Committee should be notified of any change in the contact details for the applicant or where the applicant hands over responsibility for communication with the Committee to another person at the establishment.
10. Closure of the Bank
 - 10.1 Any plans to close the Bank should be notified to the Committee as early as possible and at least two months before closure. The Committee should be informed what arrangements are to be made for disposal of the tissue or transfer to another research tissue bank.
 - 10.2 Where tissue is transferred to another research tissue bank, the ethical approval for the Bank is not transferable. Where the second bank is ethically approved, it should notify the responsible Research Ethics Committee. The terms of its own ethical approval would apply to any tissue it receives.
11. Breaches of approval conditions
 - 11.1 The Committee should be notified as soon as possible of any breach of these approval conditions.
 - 11.2 Where serious breaches occur, the Committee may review its ethical approval and may, exceptionally, suspend or terminate the approval.

Appendix 2.2. The top 100 pathways for predicted targets of the microRNAs from Metacore.

Pathway	Pathway ID	Number of Genes	-ln(p-value)	Genes
Adherens junction	hsa04520	40	22.32	CTNNA1, PTPRM, ACTB, PTPN1, LEF1, ACTN2, TGFB1, IGF1R, MET, WASF1, VCL, PTPRF, WASL, PARD3, YES1, ENSG00000158195, SMAD2, CSNK2A1, CSNK2A1P, TCF7, CTNNA2, BAIAP2, MAPK1, MAP3K7, PTPRJ, RAC1, SRC, PVRL2, MLLT4, ACVR1B, SSX2IP, NLK, TGFB2, PVRL1, ACTG1, EP300, ACTN3, IQGAP1, SMAD4, SNAI2, SMAD3, MAP4K4, MAP2K3, FGF12, RPS6KA1, FOS, NTRK2, NFATC2, MAPK8, PRKCA, TRAF6, CSDE1, NTF3, PDGFRA, EVI1, MAP3K7IP1, TGFB1, NFKB1, MAP2K1, STMN1, GADD45A, MRAS, MAP3K4, MAP3K1, FLNB, ENSG0000091436, PPM1B, PRKACA, MAP3K7IP2, MAP4K2, CACNB1, CACNA2D1, ATF2, CACNB2, MEF2C, ARRB2, FGF23, BDNF, PLA2G2F, CRKL, TAOK1, MAP3K10, FGFR2, HSPA1A, HSPA1B, NR4A1, STK4, RASA2, MAP2K4, RPS6KA4, SOS1, FGF9, DUSP16, PRKX, MAP3K3, RPS6KA5, CRK, MAPK1, MAP3K7, RAC1, SRF, CDC25B, FGF7, GRB2, PAK2, MAP2K7, FGF18, RAP1B, ENSG0000187446, PPP3CA, MAPKAPK2, ACVR1B, PLA2G3, CACNA2D2, NLK, NF1, DUSP1, TGFB2, FGF5, DUSP3, DUSP6, RASA1, MAPK14, PTPRR, PRKCB1, MAX, MAP3K14, RPS6KA3, MAP3K5, RRAS, GNG12, MAPK10, PDGFRB, PRKACB, AKT3, CACNB3
MAPK signaling pathway	hsa04010	94	17.93	

Focal adhesion	hsa04510	76	17.48	<p>ITGB4, BCL2, ACTB, PIK3CA, MAPK8, PRKCA, PGF, SHC2, ITGB1, PDGFRA, ACTN2, FN1, IGF1R, TNC, MYLK2, MAP2K1, MET, DIAPH1, TNR, RELN, VCL, COL5A1, FIGF, FLNB, ROCK1, PDPK1, CAV2, ITGB8, FLT1, ITGA9, CRKL, ITGA6, PIK3R1, TNN, COL4A4, SOS1, VAV3, VEGFA, PTEN, PTENP1, CRK, SHC1, MAPK1, THBS1, RAC1, SRC, KDR, PAK4, GRB2, PAK2, PAK6, THBS2, CAV1, RAP1B, ARHGAP5, PPP1CC, ITGA11, CCND2, ACTG1, PRKCB1, COL2A1, LAMC1, IGF1, CCND3, PAK7, ILK, MAPK10, ACTN3, ITGA7, PDGFRB, SHC4, PIK3R3, COL4A1, AKT3, PDGFC, ITGA5, LAMA4</p> <p>EFNB2, SRGAP3, SEMA6A, EPHA3, SRGAP1, GNAI2, NTN4, DPYSL2, SLIT1, NFATC2, CSDE1, ABLIM3, CXCL12, ITGB1, SEMA6C, EFNA3, MET, GNAI3, SEMA6D, SRGAP2, ROCK1, L1CAM, CFL2, EFNA5, EPHA7, EPHA4, NFATC3, NRP1, SEMA4C, UNC5D, CFL1, MAPK1, RAC1, NFATC1, PAK4, PAK2, PAK6, ENSG00000187446, PPP3CA, EPHA8, RASA1, PLXNA3, SEMA3A, EFNB3, NFAT5, PAK7, EPHB1, GNAI1, DCC, ROBO2, SEMA4F, EPHB4, EFNB1, EFNA1</p>
Axon guidance	hsa04360	54	16.71	<p>FZD7, CTNNBIP1, AXIN2, CXXC4, CAMK2D, WNT4, NFATC2, MAPK8, PRKCA, LEF1, FZD5, LRP6, TBL1X, ROCK1, PRKACA, FZD8, VANGL2, SMAD2, FZD3, FOSL1, BTRC, CSNK2A1, CSNK2A1P, SKP1A, TCF7, NFATC3, PRKX, DAAM1, MAP3K7, VANGL1, WNT1, RAC1, FBXW11, NFATC1, SIAH1, FRAT2, SFRP2, ENSG00000187446, PPP3CA, CHD8, NLK, PSEN1, PPARD, PLCB1, CCND2, DKK2, PRKCB1, PLCB4, FZD4, SENP2, NFAT5, PRICKLE2, CCND3, MAPK10, EP300, PRKACB, CAMK2B, SMAD4, SMAD3</p>
Wnt signaling pathway	hsa04310	58	13.48	

Oxidative phosphorylation	hsa00190	5	13.47	ATP6V0E1, NDUFC2, SDHC, COX7A2, ATP6V0A2
Regulation of actin cytoskeleton	hsa04810	76	13.41	ITGB4, ARPC5, PFN1, ACTB, MYH9, PIK3CA, FGF12, CSDE1, TIAM1, SCIN, ITGB1, PDGFRA, ACTN2, FN1, IQGAP2, MYLK2, MAP2K1, WASF1, DIAPH1, SSH2, MRAS, VCL, ROCK1, WASL, ITGB8, ENSG00000158195, ARPC1A, ITGA9, FGF23, ARHGEF7, NCKAP1, PIP5K1B, PIP4K2B, CRKL, FGFR2, CFL2, ITGA6, PIK3R1, SOS1, VAV3, BAIAP2, FGF9, CRK, CFL1, MAPK1, RAC1, PIP5K1A, FGF7, DIAPH2, PAK4, ARHGEF4, PIP4K2C, PAK2, PAK6, FGF18, FGF5, PPP1CC, ITGA11, GNA13, ACTG1, GSN, SLC9A1, PIP5K3, MYH10, RRAS, PAK7, GNG12, ACTN3, ITGA7, PDGFRB, PIK3R3, RDX, IQGAP1, PFN2, ITGA5, GIT1
ErbB signaling pathway	hsa04012	39	12.47	CAMK2D, CDKN1B, PIK3CA, MAPK8, HBEGF, PRKCA, CSDE1, SHC2, NRG3, CDKN1A, MAP2K1, ERBB3, CBL, CRKL, PIK3R1, MAP2K4, SOS1, CRK, SHC1, MAPK1, SRC, PAK4, GRB2, PAK2, PAK6, MAP2K7, NRG1, PRKCB1, ERBB4, PAK7, MAPK10, PLCG1, SHC4, RPS6KB1, ENSG00000109321, PIK3R3, AKT3, EREG, CAMK2B
Colorectal cancer	hsa05210	37	10.88	FZD7, AXIN2, BCL2, PIK3CA, RALGDS, FOS, MAPK8, LEF1, PDGFRA, FZD5, TGFBR1, IGF1R, MAP2K1, MET, MSH2, APPL1, FZD8, SMAD2, FZD3, MLH1, PIK3R1, SOS1, TCF7, MAPK1, RAC1, GRB2, ACVR1B, TGFBR2, BAX, FZD4, MAPK10, DCC, PDGFRB, PIK3R3, AKT3, SMAD4, SMAD3
Long-term potentiation	hsa04720	29	10.24	CAMK2D, RPS6KA1, PRKCA, CSDE1, GRIA1, GRM1, MAP2K1, ADCY1, PRKACA, ITPR1, GRIN1, ITPR3, ENSG00000198668, PRKX, MAPK1, GRIN2A, GRIA2, RAP1B, ENSG00000187446, PPP3CA, PPP1CC, PLCB1, PRKCB1, PLCB4, RPS6KA3, EP300, PRKACB, GNAQ, CAMK2B

Melanogenesis	hsa04916	39	8.54	FZD7, CAMK2D, GNAI2, CREB3L1, GNAO1, WNT4, PRKCA, CSDE1, LEF1, FZD5, EDNRB, MAP2K1, GNAI3, ADCY1, ADCY5, PRKACA, FZD8, KIT, FZD3, TCF7, ENSG00000198668, PRKX, MITF, MAPK1, KITLG, WNT1, ADCY9, PLCB1, PRKCB1, PLCB4, FZD4, CREB1, EP300, GNAI1, PRKACB, GNAQ, CAMK2B, ADCY6, CREB3L2
TGF-beta signaling pathway	hsa04350	36	8.39	E2F5, FST, TGFBR1, ID4, LTBP1, BMP6, ROCK1, SMURF1, ZFYVE9, SMAD7, SMAD2, RBL1, SMAD5, ACVR1, SKP1A, ACVR2A, NOG, INHBB, GDF6, MAPK1, THBS1, THBS2, SMAD1, BMP2, ACVR2B, ACVR1B, SMURF2, TGFBR2, BMPR1B, SP1, ZFYVE16, PITX2, EP300, RPS6KB1, SMAD4, SMAD3
GnRH signaling pathway	hsa04912	37	7.84	CAMK2D, MAP2K3, PLD1, MAPK8, HBEGF, PRKCA, CSDE1, MAP2K1, ADCY1, ADCY5, MAP3K4, MAP3K1, PRKACA, PLA2G2F, ITPR1, MAP2K4, SOS1, ITPR3, ENSG00000198668, PRKX, MAP3K3, MAPK1, SRC, GRB2, ADCY9, MAP2K7, PLA2G3, PLD2, MAPK14, PLCB1, PRKCB1, PLCB4, MAPK10, PRKACB, GNAQ, CAMK2B, ADCY6
Glycan structures - biosynthesis 1	hsa01030	43	6.94	B3GNT1, GALNT2, HS6ST2, UST, GALNT7, GALNT1, CHST3, EXT1, ENSG00000174473, NDST1, CHST14, GALNT3, ST3GAL3, STT3B, ENSG00000147408, GALNT13, C1GALT1, FUT8, ALG9, MGAT5B, CHST1, ENSG00000182022, B4GALT2, B3GNT2, GALNT9, GALNT10, HS3ST3B1, MGAT4A, GCNT1, EXTL2, B4GALT5, EXTL3, MAN1A2, GALNT12, OGT, GALNTL1, GCNT4, MGAT1, ALG2, CHSY1, B4GALT1, XYLT1, ST6GAL1
Renal cell carcinoma	hsa05211	28	6.77	PIK3CA, PGF, CSDE1, MAP2K1, MET, SLC2A1, FIGF, EGLN3, EPAS1, CRKL, PIK3R1, SOS1, PTPN11, VEGFA, CRK, MAPK1, RAC1, PAK4, GRB2, PAK2, PAK6, RAP1B, ARNT, ETS1, PAK7, EP300, PIK3R3, AKT3

Chronic myeloid leukemia	hsa05220	30	6.56	CDKN1B, PIK3CA, E2F3, CSDE1, SHC2, EVI1, TGFBR1, CDKN1A, NFKB1, MAP2K1, CDK6, CBL, CRKL, PIK3R1, RELA, SOS1, PTPN11, CRK, SHC1, MAPK1, RUNX1, GRB2, ACVR1B, TGFBR2, SHC4, PIK3R3, BCL2L1, AKT3, SMAD4, SMAD3
Pancreatic cancer	hsa05212	29	6.51	PIK3CA, RALGDS, RALBP1, E2F3, PLD1, MAPK8, PGF, STAT3, TGFBR1, NFKB1, MAP2K1, RALB, FIGF, CDK6, SMAD2, PIK3R1, RELA, VEGFA, MAPK1, RAC1, ACVR1B, TGFBR2, MAPK10, RALA, PIK3R3, BCL2L1, AKT3, SMAD4, SMAD3
O-Glycan biosynthesis	hsa00512	15	6.48	GALNT2, GALNT7, GALNT1, ENSG00000174473, GALNT3, GALNT13, C1GALT1, GALNT9, GALNT10, GCNT1, B4GALT5, GALNT12, OGT, GALNTL1, GCNT4
Glioma	hsa05214	26	6.37	CAMK2D, PIK3CA, E2F3, PRKCA, CSDE1, SHC2, PDGFRA, IGF1R, CDKN1A, MAP2K1, CDK6, PIK3R1, SOS1, PTEN, PTENP1, ENSG00000198668, SHC1, MAPK1, GRB2, PRKCB1, IGF1, PLCG1, PDGFRB, SHC4, PIK3R3, AKT3, CAMK2B
Ubiquitin mediated proteolysis	hsa04120	45	6.21	UBE2D1, CDC27, UBE4A, TRAF6, UBE3A, UBE2Z, UBE3C, UBE2R2, UBE1, MAP3K1, UBE2Q2, BIRC6, UBE2E3, UBE2E4P, UBE2J1, CUL3, UBE2I, SMURF1, UBE2W, CBL, MID1, PML, UBE2Q1, UBE2B, BTRC, SKP1A, UBE2L3, CUL4A, FBXW7, UBR5, FBXW11, SIAH1, UBE4B, TRIM32, UBE1L2, KLHL13, SMURF2, NEDD4L, PIAS3, UBE2O, ANAPC5, CUL5, NEDD4, UBE2H, HUWE1, UBE2D3
Tight junction	hsa04530	46	6.09	CTNNA1, ACTB, MYH9, GNAI2, MAGI2, PRKCA, CSDE1, CLDN2, ACTN2, CLDN14, F11R, IGSF5, GNAI3, MRAS, VAPA, ENSG00000091436, PARD3, EPB41L2, YES1, PRKCE, PPP2R3A, CSNK2A1, CSNK2A1P, CLDN11, CTNNA2, PTEN, PTENP1, MPP5, EPB41L1, AMOTL1, CLDN18, SRC, MYH11, CTTN, CASK, MLLT4, CGN, ACTG1,

				PRKCB1, CSDA, MYH10, RRAS, ACTN3, EPB41L3, GNAI1, AKT3, MYH1, MAGI1
Arachidonic acid metabolism	hsa00590	3	5.93	CYP2E1, PLA2G2F, PLA2G3
				CCNE2, BCL2, CDKN1B, PIK3CA, E2F3, CREB3L1, CSDE1, LEF1, PDGFRA, IGF1R, CDKN1A, NFKB1, MAP2K1, PDPK1, FGFR2, PIK3R1, RELA, SOS1, TCF7, HSP90B1, PTEN, PTENP1, FOXO1, MAPK1, GRB2, CREB5, IGF1, CREB1, EP300, PDGFRB, PIK3R3, AKT3, PDGFC, CREB3L2
Prostate cancer	hsa05215	33	5.93	
Androgen and estrogen metabolism	hsa00150	2	5.71	CARM1, HSD17B7
Metabolism of xenobiotics by cytochrome P450	hsa00980	4	5.55	ADH1B, ADH1C, CYP2E1, ADHFE1, ADH1A
Glutathione metabolism	hsa00480	1	5.54	GCLM
Starch and sucrose metabolism	hsa00500	6	5.21	ENPP3, HK1, EP400, GYS1, HK2, PGM1
Pyruvate metabolism	hsa00620	2	5.05	ACACB, PC
Glycine, serine and threonine metabolism	hsa00260	3	4.65	PSPH, PSPHL, RDH11, SHMT2
Hedgehog signaling pathway	hsa04340	21	4.28	WNT4, RAB23, SMO, GLI3, BMP6, ZIC2, CSNK1G2, PRKACA, BTRC, PRKX, WNT1, FBXW11, PTCH1, GAS1, BMP2, CSNK1G3, CSNK1G1, GLI2, STK36, PRKACB, LRP2
Aminoacyl-tRNA biosynthesis	hsa00970	2	4.23	WARS2, RARS
Insulin signaling pathway	hsa04910	44	4.11	PPARGC1A, ACACB, PRKAG2, PIK3CA, PTPN1, MAPK8, PHKA2, CSDE1, SHC2, PFKP, MAP2K1, LIPE, PTPRF, SOCS4, PDPK1, PRKACA, CBL, GYS1, CRKL, PIK3R1, FLOT2, SOS1, ENSG00000198668, PRKAR2A, PRKX, FOXO1, CRK, SHC1, MAPK1, PRKAA1, GRB2, RHOQ, FLOT1, PPP1CC,

PPP1R3A, PRKAR1A, MAPK10,
TRIP10, SHC4, RPS6KB1,
PRKACB, PIK3R3, AKT3, EIF4E

Histidine metabolism	hsa00340	3	4.01	ASPA, PRPS1, PRPS1L1, CARM1
Acute myeloid leukemia	hsa05221	21	4.01	PIK3CA, CSDE1, LEF1, STAT3, NFKB1, MAP2K1, PIM1, KIT, PML, PIK3R1, RELA, SOS1, TCF7, MAPK1, RUNX1, GRB2, JUP, PPARD, RPS6KB1, PIK3R3, AKT3
Polyunsaturated fatty acid biosynthesis	hsa01040	10	3.88	BAAT, ELOVL2, SCD, ACOT7, SCD5, ELOVL5, YOD1, PECR, HADHA, ACOT8
VEGF signaling pathway	hsa04370	25	3.86	PIK3CA, NFATC2, PRKCA, CSDE1, SHC2, MAP2K1, PLA2G2F, PIK3R1, VEGFA, NFATC3, MAPK1, RAC1, SRC, NFATC1, KDR, ENSG00000187446, PPP3CA, MAPKAPK2, PLA2G3, MAPK14, PRKCB1, NFAT5, PLCG1, PIK3R3, AKT3
Long-term depression	hsa04730	26	3.73	GNAI2, GNAO1, PRKCA, CSDE1, GRIA1, GRM1, IGF1R, MAP2K1, GNAI3, PLA2G2F, ITPR1, PRKG1, ITPR3, MAPK1, GRIA2, PLA2G3, NOS1, RYR1, PLCB1, GNA13, GRID2, PRKCB1, PLCB4, IGF1, GNAI1, GNAQ
Base excision repair	hsa03410	2	3.57	POLL, TDG
Porphyrin and chlorophyll metabolism	hsa00860	2	3.57	UROD, ALAD
Amyotrophic lateral sclerosis (ALS)	hsa05030	9	3.45	BCL2, CAT, NEFL, NEFH, RAC1, PPP3CA, BAX, SLC1A2, BCL2L1
Non-small cell lung cancer	hsa05223	20	3.43	RARB, PIK3CA, RASSF5, E2F3, PRKCA, CSDE1, FOXO3, MAP2K1, PDPK1, CDK6, STK4, PIK3R1, SOS1, MAPK1, GRB2, PRKCB1, PLCG1, PIK3R3, AKT3, RXRA
Glycolysis / Gluconeogenesis	hsa00010	7	3.39	ADH1B, ADH1C, PFKP, HK1, HK2, ADHFE1, PGM1, ADH1A
Complement and coagulation cascades	hsa04610	8	3.32	FGA, CD46, CR2, THBD, F3, SERPINE1, SERPINF2, F13A1

Endometrial cancer	hsa05213	19	3.11	CTNNA1, AXIN2, PIK3CA, CSDE1, LEF1, FOXO3, MAP2K1, PDPK1, MLH1, PIK3R1, SOS1, TCF7, CTNNA2, PTEN,PTENP1, MAPK1, GRB2, ILK, PIK3R3, AKT3
Neuroactive ligand-receptor interaction	hsa04080	45	3.03	EDG1, NR3C1, CCKBR, LEP, GRIA4, PRLR, GABRA1, GRIN3A, EDG4, GRIA1, GRM1, GLRA3, EDNRB, GRM3, ADRB3, PTGFR, PTGER2, NPY2R, GRID1, NPFFR2, PARD3, GABRB2, HRH1, HTR2C, ADRA2A, GRIN1, PTGER3, HTR7, CNR1, F2RL2, GRIN2A, RXFP2, GRIA2, GPR23, GABRG2, GRM7, GABBR2, GABRB3, THRB, GHRHR, GRID2, CALCR, GABRQ, DRD2, TBXA2R
Autoimmune thyroid disease	hsa05320	5	2.98	CD40, ENSG00000179344, HLA-DPB1, IL10, HLA-DRB5
Phosphatidylinositol signaling system	hsa04070	24	2.95	SYNJ1, PIK3CA, PRKCA, OCRL, PIP5K1B, PIP4K2B, ITPR1, PIK3R1, ITPR3, PTEN,PTENP1, ENSG00000198668, PIP5K1A, PIP4K2C, DGKQ, DGKI, ITPK1, CARKL, PLCB1, PRKCB1, PLCB4, PIP5K3, PLCG1, DGKZ, PIK3R3
Cell adhesion molecules (CAMs)	hsa04514	38	2.83	CD4, MPZL1, PTPRM, CD40, NRCAM, SDC2, CLDN2, ITGB1, CLDN14, CDH2, F11R, VCAN, PTPRF, ENSG00000179344, NEGR1, L1CAM, ITGB8, CNTNAP2, ITGA9, SDC1, NFASC, ITGA6, CLDN11, SPN, ALCAM, HLA-DPB1, CLDN18, PVRL2, PTPRC, NLGN2, NLGN1, CNTNAP1, CADM1, PVRL1, CNTN1, CNTN2, NLGN3, HLA-DRB5
Gap junction	hsa04540	30	2.82	GNAI2, PRKCA, CSDE1, GRM1, PDGFRA, MAP2K1, GNAI3, ADCY1, ADCY5, PRKACA, HTR2C, ITPR1, PRKG1, SOS1, ITPR3, PRKX, MAPK1, SRC, GRB2, ADCY9, PLCB1, PRKCB1, PLCB4, DRD2, GNAI1, PDGFRB, PRKACB, GNAQ, PDGFC, ADCY6
Melanoma	hsa05218	24	2.76	PIK3CA, FGF12, E2F3, CSDE1, PDGFRA, IGF1R, CDKN1A, MAP2K1, MET, CDK6, FGF23, PIK3R1, FGF9, PTEN,PTENP1, MITF, MAPK1, FGF7, FGF18, FGF5, IGF1, PDGFRB, PIK3R3, AKT3, PDGFC

Glycan structures - degradation	hsa01032	2	2.75	GNS, BRUNOL6 PIK3CA, RPS6KA1, PGF, FIGF, PDPK1, PIK3R1, VEGFA, MAPK1, ENSG00000164327, PRKAA1, ULK2, RPS6KA3, IGF1, RPS6KB1, PIK3R3, AKT3, EIF4E
mTOR signaling pathway	hsa04150	17	2.7	CCL2, TNFRSF11B, LEP, CD40, PRLR, TNFSF4, IL6ST, CXCL12, CNTFR, PDGFRA, EDA, TGFBR1, MET, LIFR, CCL22, CXCL11, FLT1, TNFSF11, KIT, TNFRSF1B, IL7R, ACVR1, VEGFA, ACVR2A, INHBB, CSF1R, KITLG, TNFRSF21, KDR, CSF1, IL10, IL10RA, CCL21, BMP2, ACVR2B, ACVR1B, IL15, TGFBR2, BMPR1B, EDAR, IL28RA, PDGFRB, IL11, IL9R, TNFRSF8, PDGFC
Cytokine-cytokine receptor interaction	hsa04060	46	2.68	ITGB4, FNDC5, SDC2, ITGB1, FN1, TNC, TNR, RELN, COL5A1, ITGB8, ITGA9, SDC1, DAG1, ITGA6, TNN, COL4A4, FNDC3A, THBS1, THBS2, ITGA11, COL2A1, LAMC1, ITGA7, COL4A1, ITGA5, LAMA4
ECM-receptor interaction	hsa04512	26	2.61	B3GNT1, ST3GAL3, FUT8, CHST1, B4GALT2, B3GNT2, B4GALT1
Keratan sulfate biosynthesis	hsa00533	7	2.59	CTNNA1, ACTB, GNAI2, PIK3CA, RASSF5, PRKCA, CXCL12, CLDN2, ITGB1, ACTN2, CLDN14, F11R, GNAI3, VCL, ROCK1, PIK3R1, VAV3, CLDN11, CTNNA2, PTPN11, RAC1, CLDN18, CYBB, RAP1B, MLLT4, ARHGAP5, MAPK14, ACTG1, PRKCB1, PLCG1, ACTN3, GNAI1, PIK3R3
Leukocyte transendothelial migration	hsa04670	33	2.25	PER3, NR1D1, BHLHB3, PER2, NPAS2, CLOCK
Circadian rhythm	hsa04710	6	2.15	
Urea cycle and metabolism of amino groups	hsa00220	3	2.13	ARG2, ODC1, SMS PTPRZ1, MAPK8, HBEGF, F11R, ADAM10, IGSF5, NFKB1, MET, ATP6V0E1, MAP2K4, RELA, PTPN11, RAC1, SRC, ADAM17, ATP6V0A2, MAPK14, MAP3K14, MAPK10, PLCG1, GIT1
Epithelial cell signaling in Helicobacter pylori	hsa05120	21	2.13	

Notch signaling pathway	hsa04330	14	1.98	APH1A, NUMB, NOTCH2, NUMBL, DLL1, HES1, ADAM17, PSEN1, DTX4, JAG1, NCSTN, PCAF, EP300, JAG2
Calcium signaling pathway	hsa04020	46	1.93	CCKBR, CAMK2D, PDE1C, PLN, PHKA2, PRKCA, GRM1, PDGFRA, EDNRB, MYLK2, ADRB3, PTGFR, ADCY1, ATP2B4, ERBB3, PRKACA, ATP2A2, VDAC2, SLC8A1, HRH1, HTR2C, ITPR1, GRIN1, PTGER3, ITPR3, HTR7, ENSG00000198668, PRKX, GRIN2A, ADCY9, ENSG00000187446, PPP3CA, NOS1, RYR1, PLCB1, PRKCB1, PLCB4, ERBB4, GNAL, PLCG1, PDGFRB, PRKACB, TBXA2R, GNAQ, VDAC3, CAMK2B
Cysteine metabolism	hsa00272	1	1.92	GOT1
Aminophosphonate metabolism	hsa00440	1	1.92	CARM1
Olfactory transduction	hsa04740	11	1.91	CAMK2D, PDE1C, PRKACA, ARRB2, PRKG1, ENSG00000198668, PRKX, GNAL, PRKACB, CLCA2, CAMK2B
Propanoate metabolism	hsa00640	4	1.91	ACACB, SUCLG2, HADHA, PCCA
Small cell lung cancer	hsa05222	26	1.87	CCNE2, RARB, BCL2, CDKN1B, PIK3CA, E2F3, TRAF6, ITGB1, FN1, NFKB1, CDK6, ITGA6, PIK3R1, COL4A4, RELA, PTEN, PTENP1, NOS1, PIAS3, MAX, LAMC1, PIK3R3, BCL2L1, COL4A1, AKT3, RXRA, LAMA4
Regulation of autophagy	hsa04140	3	1.83	GABARAPL1, GABARAPL3, PRKAA1, ULK2
Fc epsilon RI signaling pathway	hsa04664	23	1.73	MAP2K3, PIK3CA, MAPK8, PRKCA, CSDE1, MAP2K1, PRKCE, PLA2G2F, PIK3R1, MAP2K4, SOS1, VAV3, MAPK1, RAC1, GRB2, MAP2K7, PLA2G3, MAPK14, PRKCB1, MAPK10, PLCG1, PIK3R3, AKT3
Valine, leucine and isoleucine degradation	hsa00280	6	1.71	HADH, ACAA2, BCAT2, HADHA, PCCA, DBT
Taste transduction	hsa04742	6	1.71	ACCN1, PRKACA, ITPR3, PRKX, PRKACB, ADCY6
Tryptophan metabolism	hsa00380	9	1.7	CAT, WARS2, CARM1, HADH, OGDHL, NFX1, GCDH, HADHA, OGDH

p53 signaling pathway	hsa04115	21	1.67	CCNE2, CCNG1, IGFBP3, CDKN1A, GADD45A, RRM2, CDK6, EI24, STEAP3, BID, PTEN, PTENP1, THBS1, SERPINE1, SIAH1, MDM4, CCND2, BAX, PMAIP1, IGF1, CCND3, BBC3
Folate biosynthesis	hsa00790	5	1.6	QDPR, FPGS, EP400, GCH1, ALPL
Thyroid cancer	hsa05216	10	1.58	CSDE1, LEF1, MAP2K1, RET, NCOA4, TCF7, MAPK1, TFG, CCDC6, RXRA
N-Glycan degradation	hsa00511	1	1.56	BRUNOL6
Linoleic acid metabolism	hsa00591	4	1.49	CYP2E1, RDH11, PLA2G2F, PLA2G3
SNARE interactions in vesicular transport	hsa04130	12	1.39	SNAP23, STX11, SYBL1, VAMP3, VAMP1, STX1B, STX17, VAMP2, VAMP4, SNAP25, STX1A, SNAP29
T cell receptor signaling pathway	hsa04660	26	1.39	CD4, PIK3CA, FOS, NFATC2, CSDE1, NFKB1, CBL, PIK3R1, SOS1, VAV3, NFATC3, NFATC1, IL10, PAK4, GRB2, PAK2, PAK6, PTPRC, ENSG00000187446, PPP3CA, MAP3K14, NFAT5, PAK7, PLCG1, PIK3R3, AKT3
Inositol phosphate metabolism	hsa00562	15	1.37	SYNJ1, PIK3CA, OCRL, MINPP1, PIP5K1B, PIP4K2B, PTEN, PTENP1, PIP5K1A, PIP4K2C, ITPK1, CARKL, PLCB1, PLCB4, PIP5K3, PLCG1
Chondroitin sulfate biosynthesis	hsa00532	8	1.36	UST, CHST3, DSE, CHST14, ENSG00000147408, ENSG00000182022, CHSY1, XYLT1
1- and 2-Methylnaphthalene degradation	hsa00624	8	1.36	ADH1B, ADH1C, DHRS1, MYST3, ADHFE1, LYCAT, NAT5, MYST4, ADH1A
Selenoamino acid metabolism	hsa00450	4	1.35	CARM1, PAPSS2, SEPHS2, SEPHS1
Glycerophospholipid metabolism	hsa00564	20	1.31	GPD2, PTDSS1, ETNK2, PLD1, PPAP2B, GPAM, MYST3, PLA2G2F, AGPAT3, DGKQ, DGKI, PLA2G3, LYCAT, ETNK1, PHOSPHO1, PLD2, NAT5, MYST4, DGKZ, ACHE
Naphthalene and anthracene degradation	hsa00626	2	1.28	CARM1, DHRS1
Heparan sulfate biosynthesis	hsa00534	7	1.25	HS6ST2, EXT1, NDST1, HS3ST3B1, GLCE, EXTL2, EXTL3

Novobiocin biosynthesis	hsa00401	2	1.25	TAT, GOT1
Dorso-ventral axis formation	hsa04320	9	1.25	MAP2K1, NOTCH2, SPIRE1, SOS1, MAPK1, GRB2, ETS2, ERBB4, ETS1
Taurine and hypotaurine metabolism	hsa00430	4	1.24	GAD1, BAAT, GAD2, C10orf22
Nitrogen metabolism	hsa00910	3	1.24	GLS, GLUL, CA7
Butanoate metabolism	hsa00650	7	1.23	GAD1, HADH, DDHD1, RDH11, GAD2, L2HGDH, HADHA
Pyrimidine metabolism	hsa00240	15	1.22	UPRT, NME4, POLR1B, POLR1C, RRM2, DCK, AK3, NT5C2, POLR3H, ENTPD5, ENTPD1, NT5C3, POLR2D, CMPK, POLR3G
Sulfur metabolism	hsa00920	1	1.2	PAPSS2
Methane metabolism	hsa00680	3	1.19	CAT, SHMT2, LPO
Cell Communication	hsa01430	26	1.17	ITGB4, ACTB, LMNB1, LMNA, FN1, TNC, TNR, RELN, COL5A1, INA, VIM, KRT74, GJA5, ITGA6, TNN, COL4A4, THBS1, THBS2, KRT38, GJA3, ACTG1, COL2A1, LAMC1, COL4A1, GJA7, LAMA4
N-Glycan biosynthesis	hsa00510	13	1.16	DHDDS, DOLPP1, STT3B, FUT8, ALG9, MGAT5B, B4GALT2, MGAT4A, MAN1A2, MGAT1, ALG2, B4GALT1, ST6GAL1
Carbon fixation	hsa00710	3	1.09	GPT2, RPIA, GOT1
gamma-Hexachlorocyclohexane degradation	hsa00361	3	1.09	DHRS1, ACP6, ALPL
Neurodegenerative Diseases	hsa01510	12	1.06	VAPB, BCL2, NR4A2, NEFH, PRNP, FBXW7, GRB2, PSEN1, BAX, EP300, BCL2L1, HSPA5
Benzoate degradation via CoA ligation	hsa00632	9	1.06	DHRS1, MYST3, YOD1, GCDH, LYCAT, HADHA, CARKL, NAT5, MYST4

Appendix 3.1. 150 microRNAs used in the OP signature.

1	hsa-let-7a	51	hsa-mir-184	101	hsa-mir-346
2	hsa-let-7b	52	hsa-mir-185	102	hsa-mir-34a
3	hsa-let-7c	53	hsa-mir-186	103	hsa-mir-363
4	hsa-let-7d	54	hsa-mir-18a	104	hsa-mir-365
5	hsa-let-7e	55	hsa-mir-190	105	hsa-mir-369-3p
6	hsa-let-7f	56	hsa-mir-191	106	hsa-mir-370
7	hsa-let-7g	57	hsa-mir-192	107	hsa-mir-375
8	hsa-let-7i	58	hsa-mir-193b	108	hsa-mir-378
9	hsa-mir-1	59	hsa-mir-194	109	hsa-mir-379
10	hsa-mir-100	60	hsa-mir-195	110	hsa-mir-381
11	hsa-mir-101	61	hsa-mir-197	111	hsa-mir-382
12	hsa-mir-103	62	hsa-mir-19a	112	hsa-mir-409-3p
13	hsa-mir-105	63	hsa-mir-19b	113	hsa-mir-410
14	hsa-mir-106a	64	hsa-mir-200c	114	hsa-mir-411
15	hsa-mir-106b	65	hsa-mir-203	115	hsa-mir-421
16	hsa-mir-107	66	hsa-mir-204	116	hsa-mir-424
17	hsa-mir-10a	67	hsa-mir-20a	117	hsa-mir-425-5p
18	hsa-mir-10b	68	hsa-mir-20b	118	hsa-mir-432
19	hsa-mir-125b	69	hsa-mir-21	119	hsa-mir-433
20	hsa-mir-126	70	hsa-mir-210	120	hsa-mir-451
21	hsa-mir-130a	71	hsa-mir-212	121	hsa-mir-484
22	hsa-mir-130b	72	hsa-mir-217	122	hsa-mir-485-5p
23	hsa-mir-132	73	hsa-mir-218	123	hsa-mir-487b
24	hsa-mir-133a	74	hsa-mir-22	124	hsa-mir-488
25	hsa-mir-134	75	hsa-mir-221	125	hsa-mir-495
26	hsa-mir-135a	76	hsa-mir-222	126	hsa-mir-497
27	hsa-mir-136	77	hsa-mir-223	127	hsa-mir-500
28	hsa-mir-137	78	hsa-mir-23a	128	hsa-mir-505
29	hsa-mir-138	79	hsa-mir-23b	129	hsa-mir-539
30	hsa-mir-142-5p	80	hsa-mir-24	130	hsa-mir-542-3p
31	hsa-mir-143	81	hsa-mir-25	131	hsa-mir-551b
32	hsa-mir-144	82	hsa-mir-26a	132	hsa-mir-577
33	hsa-mir-145	83	hsa-mir-26b	133	hsa-mir-584
34	hsa-mir-146a	84	hsa-mir-27a	134	hsa-mir-589
35	hsa-mir-148a	85	hsa-mir-27b	135	hsa-mir-592
36	hsa-mir-148b	86	hsa-mir-29a	136	hsa-mir-598
37	hsa-mir-149	87	hsa-mir-29b	137	hsa-mir-629
38	hsa-mir-150	88	hsa-mir-29c	138	hsa-mir-652
39	hsa-mir-152	89	hsa-mir-30b	139	hsa-mir-660
40	hsa-mir-153	90	hsa-mir-30c	140	hsa-mir-758
41	hsa-mir-155	91	hsa-mir-30d	141	hsa-mir-766

42	hsa-mir-15a	92	hsa-mir-32	142	hsa-mir-767-5p
43	hsa-mir-15b	93	hsa-mir-320	143	hsa-mir-769-5p
44	hsa-mir-16	94	hsa-mir-320	144	hsa-mir-9
45	hsa-mir-181a	95	hsa-mir-324-5p	145	hsa-mir-92b
46	hsa-mir-181b	96	hsa-mir-326	146	hsa-mir-93
47	hsa-mir-181c	97	hsa-mir-328	147	hsa-mir-95
48	hsa-mir-181d	98	hsa-mir-335	148	hsa-mir-98
49	hsa-mir-182	99	hsa-mir-340	149	hsa-mir-99a
50	hsa-mir-183	100	hsa-mir-345	150	hsa-mir-99b

Appendix 4.1. Consent and transfer for the samples.



BRIGHAM AND
WOMEN'S HOSPITAL



HARVARD
MEDICAL SCHOOL

75 Francis Street
Boston, Massachusetts 02115
Tel: (617) 732-6600
Fax: (617) 734-8342

Department of Neurosurgery

Josie Hayes,
LICAP
Leeds University Medical School,
Leeds, UK

30 April 2015

Dear Josie,

I am writing to confirm that the brain tumor samples that were used in your studies were collected under a protocol approved by The Ohio State University Institutional Review Board. All samples were de-identified and approved for experimental use. These samples were collected when I was a faculty member at Ohio State University. The protocol is IRB 2005C0075 "Investigating novel therapeutic strategies for brain tumor treatment".

Please contact me if you have any further questions.

Yours Sincerely,

A handwritten signature in black ink, appearing to read 'SJC'.

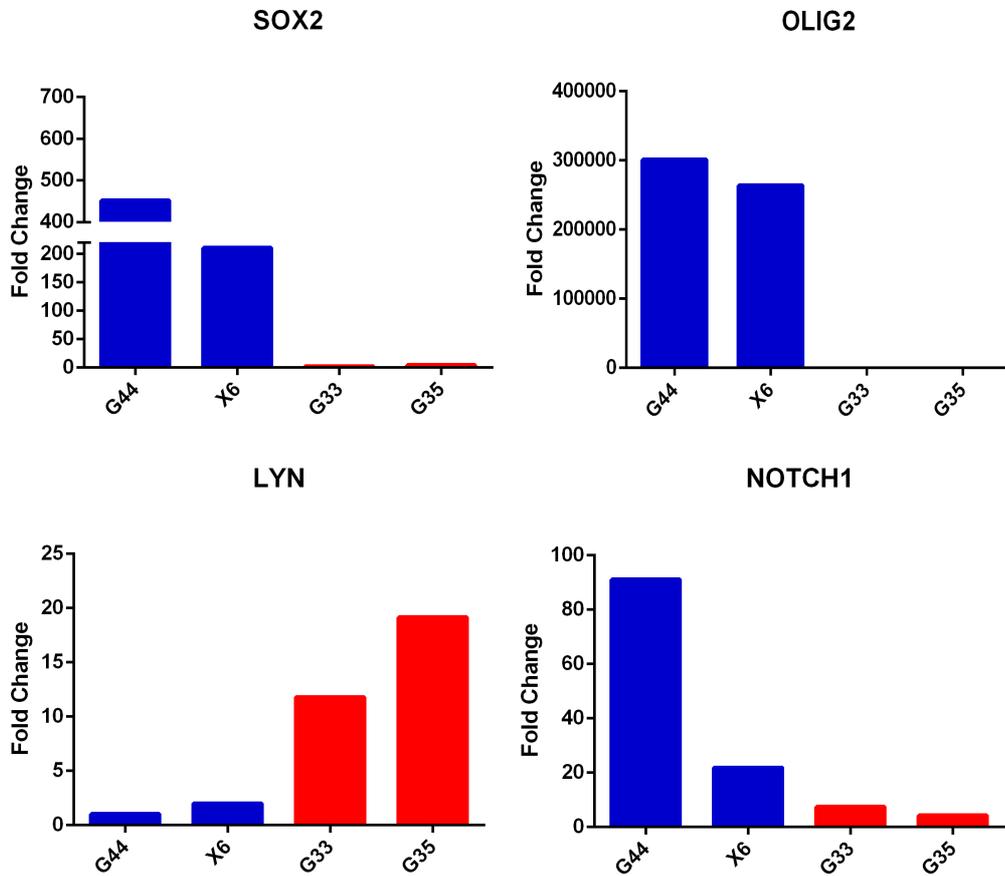
Sean Lawler, Ph.D.
Principal InvestigatorHarvey Cushing Neuro-Oncology Laboratories, Department of Neurosurgery
Brigham and Women's Hospital, Harvard Medical School, 4 Blackfan Circle, H.I.M. 930A, Boston, MA 02115.
slawler@partners.org

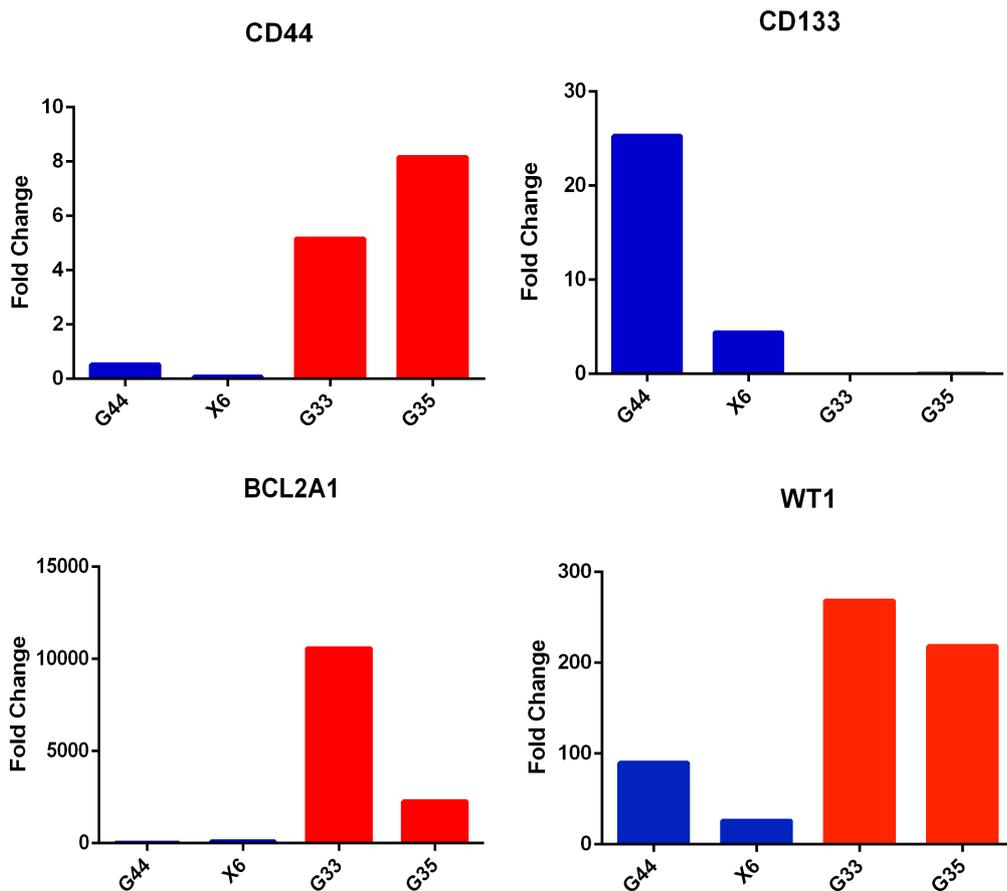
Appendix 4.2. Primer Sequences.

CD133 forward: GTTTCCGACTCCTTTTG
CD133 reverse: ATCTCCCTGTTGGTGAT
Olig2 forward: GTTCTCCCTGAGGCTTTTC
Olig2 reverse: GGAAGATAGTCGTCGCAGCTT
Sox2 forward: CCTGATTCCAGTTTGCCTCT
Sox2 reverse: CAGCTCCGTCTCCATCATCATGT
Notch1 forward: TCACGCTGACGGAGTACAAG
Notch1 reverse: GGCAGTGGCAGATGTAGGAG
CD44 forward: AAGGTGGAGCAAACACAACC
CD44 reverse: CTTCTGCCACACCTTCTTC
LYN forward: TTCCCTACCCAGGGAGAACT
LYN reverse: CTGCCTTTTCTTCCAGCAC
WT1 forward: ACTCTTGTACGGTCGGCATC
WT1 reverse: TCTCACCAGTGTGCTTCTTG
BCL2A1 forward: ATGGATAAGGCAAACGGAG
BCL2A1 reverse: TGGAGTGTCTTTCTGGTCA
GAPDH forward: GAAGGTGAAGGTCGGAGTCA
GAPDH reverse: TTGAGGTCAATGAAGGGGTC
18S reverse: CTTGGATGTGGTAGCCGTTT
18S forward: AACTTTCGATGGTAGTCGCCG
SHC1 forward: CTCAGGAACCCACCAAAC
SHC1 reverse: GATGGTCAGGTGGCTCTTC
SLC25A24 forward: AGAAATTGTCCAGTCTCTCCAG
SLC25A24 reverse: AAGTAGTCTCTCCATTCAATCCA
P4HA2 forward: CCCAGGCACAATTTCCAGAG
P4HA2 reverse: TCCACAACACCGTATGATAATAGT
SLC31A2 forward: CGGTGCTTCTGTTTGATTTCTG
SLC31A2 reverse: TTGCCAACCTTGATGCCTTC
FBN1 forward: GCATTTGCCAGAACAACCTCT
FBN1 reverse: TTACCCTCACACTCGTCCAC
WNT4 forward: GCGGGAGAGAAGCAAGGG
WNT4 reverse: GCATTCCACCCGCATGTG
LMNA forward: TCACCCGCTCCTACCTCCT
LMNA reverse: GGCAGGTCCCAGATTACATGAT
FNDC3B forward: ACAATGATGATGACCGACAAA
FNDC3B reverse: GGATTAACCTTGAACGAGAATAACCT
TGFB1 forward: GGCAATCATCTCTCTGGAAGT
TGFB1 reverse: AATTATGTGGTTCCGAAGCAAAT
GLUT1 forward: AACTCTTCAGCCAGGGTCCAC
GLUT1 reverse: CACAGTGAAGATGATGAAGAC

Appendix 4.3. Characterisation of the GSC cell lines (performed by Dr. Marco Mineo, Harvard Medical School)

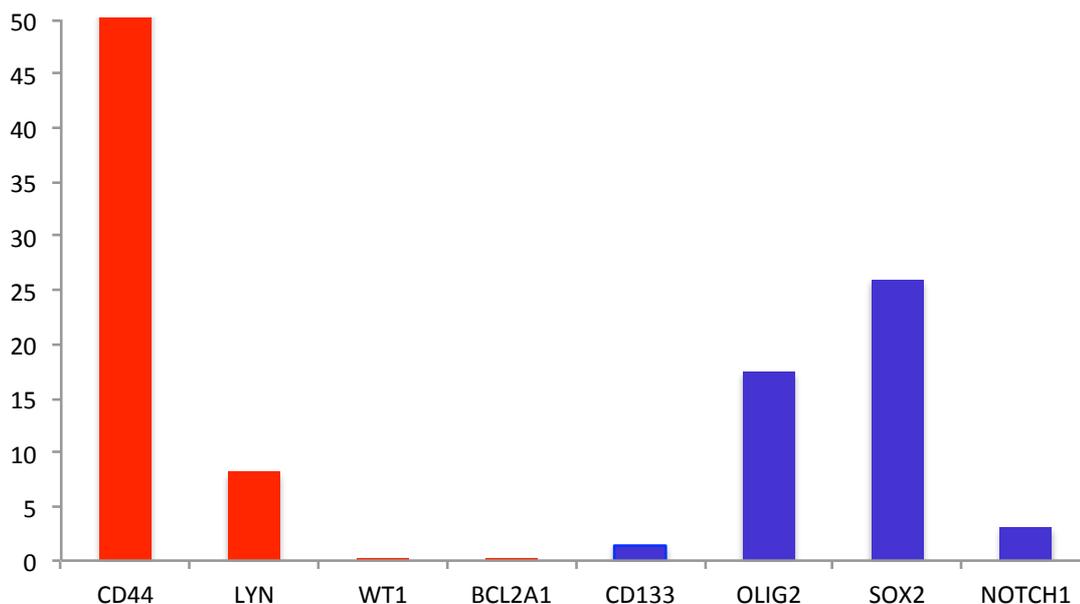
Blue are proneural markers and red are mesenchymal markers according to Mao *et al* (Mao *et al.*, 2013).

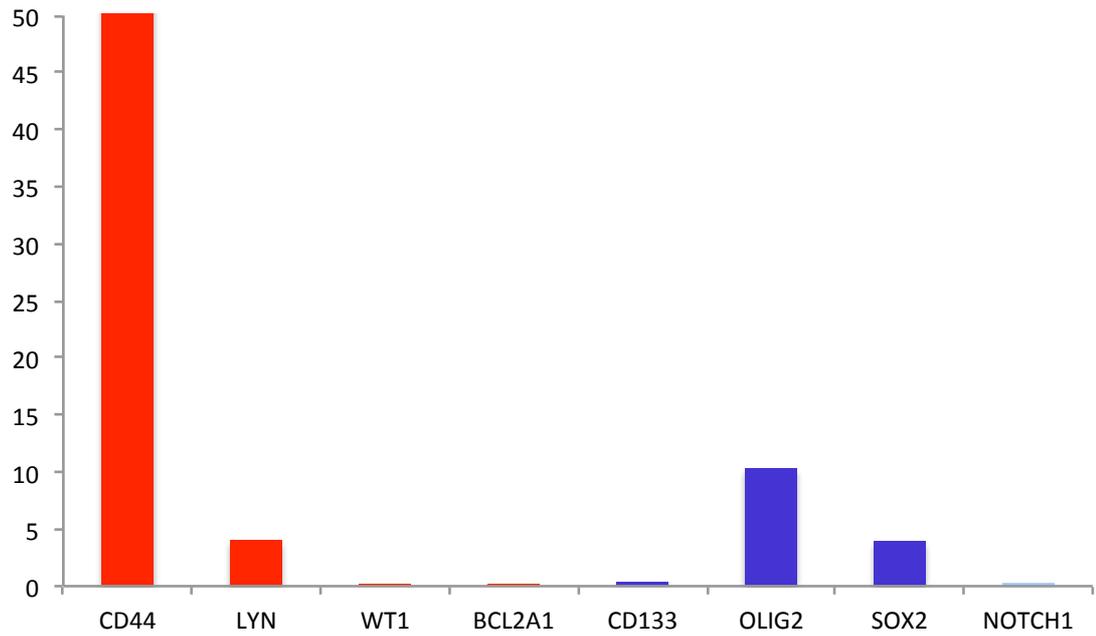




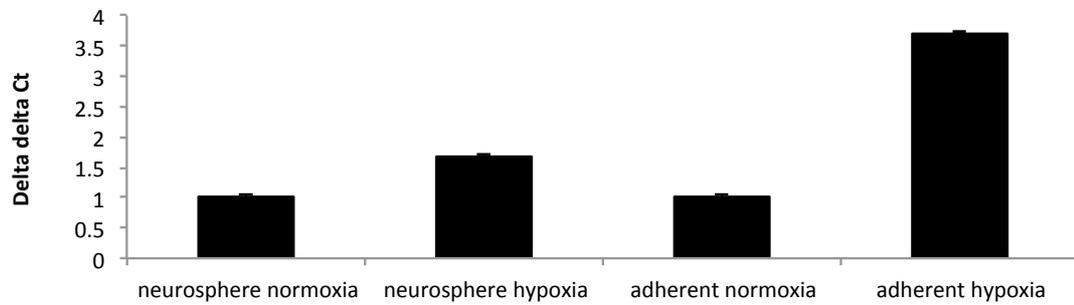
The following characterisations were performed using sequencing data produced by Dr. Sally Harrison.

GBM1



GBM4

Appendix 4.4. Glut1 (target of HIF1A) mRNA levels to show cellular response to hypoxia.



References

- Ach, R.A. et al. 2008. Measuring microRNAs: comparisons of microarray and quantitative PCR measurements, and of different total RNA prep methods. *BMC Biotechnology*. **8**,pp.1–16.
- Alcantara Llaguno, S. et al. 2009. Malignant astrocytomas originate from neural stem/progenitor cells in a somatic tumor suppressor mouse model. *Cancer cell*. **15**(1),pp.45–56.
- Aldaz, B. et al. 2013. Involvement of miRNAs in the differentiation of human glioblastoma multiforme stem-like cells. *PLoS ONE*. **8**(10),p.e77098.
- Alencar, A.J. et al. 2011. MicroRNAs are independent predictors of outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Clinical cancer research*. **17**(12),pp.4125–4135.
- Altman, D.G. and Bland, J.M. 1998. Time to event (survival) data. *BMJ*. **317**(7156),pp.468–469.
- Altman, D.G. et al. 2009. Prognosis and prognostic research: validating a prognostic model. *BMJ*. **338**,pp.b605–b605.
- Andorfer, C.A. et al. 2011. MicroRNA signatures: clinical biomarkers for the diagnosis and treatment of breast cancer. *Trends in Molecular Medicine*. **17**(6),pp.313–319.
- Andreasen, D. et al. 2010. Improved microRNA quantification in total RNA from clinical samples. *Methods*. **50**(4),pp.S6–S9.
- The Cancer Genome Atlas - Data Portal. USA: National Institute of Health. Available from: <http://tcga-data.nci.nih.gov/> [Accessed April 24, 2013].
- Arroyo, J.D. et al. 2011. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proceedings of the National Academy of Sciences*. **108**(12),pp.5003–5008.
- Arzt, L. et al. 2011. Evaluation of formalin-free tissue fixation for RNA and microRNA studies. *Experimental and molecular pathology*. **91**(2),pp.490–495.
- Auffinger, B. et al. 2014. Conversion of differentiated cancer cells into cancer stem-like cells in a glioblastoma model after primary chemotherapy. *Cell Death and Differentiation*. **21**(7),pp.1119–1131.
- Auvinen, K.Q.E.A.D.G.P. et al. 2012. miRSeqNovel: An R based workflow for analyzing miRNA sequencing data. *Molecular and Cellular Probes*. **26**(5),pp.208–211.

- Auyeung, V.C. et al. 2013. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*. **152**(4),pp.844–858.
- Babae, N. et al. 2014. Systemic miRNA-7 delivery inhibits tumor angiogenesis and growth in murine xenograft glioblastoma. *Oncotarget*. **5**(16),pp.6687–6700.
- Bachoo, R.M. et al. 2002. Epidermal growth factor receptor and Ink4a/Arf: convergent mechanisms governing terminal differentiation and transformation along the neural stem cell to astrocyte axis. *Cancer cell*. **1**(3),pp.269–277.
- Backen, A. et al. 2014. The combination of circulating Ang1 and Tie2 levels predicts progression-free survival advantage in bevacizumab-treated patients with ovarian cancer. *Clinical cancer research*. **20**(17),pp.4549–4558.
- Bader, A.G. 2012. miR-34 - a microRNA replacement therapy is headed to the clinic. *Frontiers in genetics*. **3**,p.120.
- Bady, P. et al. 2012. MGMT methylation analysis of glioblastoma on the Infinium methylation BeadChip identifies two distinct CpG regions associated with gene silencing and outcome, yielding a prediction model for comparisons across datasets, tumor grades, and CIMP-status. *Acta Neuropathologica*. **124**(4),pp.547–560.
- Ballotti, R. et al. 1989. Insulin receptor: tyrosine kinase activity and insulin action. *Reproduction, nutrition, development*. **29**(6),pp.653–661.
- Ban, E. et al. 2013. Enhanced extraction efficiency of miRNA from cells by addition of Triton X-100. *Analytical and bioanalytical chemistry*. **405**(23),pp.7535–7539.
- Bao, S. et al. 2006. Stem cell-like glioma cells promote tumor angiogenesis through vascular endothelial growth factor. *Cancer research*. **66**(16),pp.7843–7848.
- Barani, I.J. and Larson, D.A. 2015. Radiation therapy of glioblastoma. *Cancer treatment and research*. **163**,pp.49–73.
- Bates, S. et al. 1998. p14ARF links the tumour suppressors RB and p53. *Nature*. **395**(6698),pp.124–125.
- Bazzoni, F. et al. 2009. Induction and regulatory function of miR-9 in human monocytes and neutrophils exposed to proinflammatory signals. *Proceedings of the National Academy of Sciences*. **106**(13),pp.5282–5287.
- Beier, C.P. et al. 2012. The Cancer Stem Cell Subtype Determines Immune Infiltration of Glioblastoma. *Stem Cells and Development*. **21**(15),pp.2753–2761.
- Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological*. **57**(1),pp.289–300.

- Berghoff, A.S. et al. 2013. Clinical Neuropathology Practice Guide 3-2013: levels of evidence and clinical utility of prognostic and predictive candidate brain tumor biomarkers. *Clinical Neuropathology*. **32**(05),pp.148–158.
- Bhat, K.P.L. et al. 2013. Mesenchymal Differentiation Mediated by NF-kB Promotes Radiation Resistance in Glioblastoma. *Cancer cell*. **24**(3),pp.331-46.
- Bidlingmaier, S. et al. 2008. The utility and limitations of glycosylated human CD133 epitopes in defining cancer stem cells. *Journal of molecular medicine*. **86**(9),pp.1025–1032.
- Bienkowski, M. et al. 2013. Screening for EGFR amplifications with a novel method and their significance for the outcome of glioblastoma patients. *PLoS ONE*. **8**(6),p.e65444.
- BIPM 2006. The International System of Units (SI).pp.1–88.
- Bleeker, F.E. et al. 2012. Recent advances in the molecular understanding of glioblastoma. *Journal of neuro-oncology*. **108**(1),pp.11–27.
- Blower, P.E. et al. 2008. MicroRNAs modulate the chemosensitivity of tumor cells. *Molecular Cancer Therapeutics*. **7**(1),pp.1–9.
- Bouchie, A. 2013. First microRNA mimic enters clinic. *Nature Biotechnology*. **31**(7),p.577.
- Boudreau, R.L. et al. 2014. Transcriptome-wide discovery of microRNA binding sites in human brain. *Neuron*. **81**(2),pp.294–305.
- Bozdag, S. et al. 2013. Age-specific signatures of glioblastoma at the genomic, genetic, and epigenetic levels. *PLoS ONE*. **8**(4),p.e62982.
- Brennan, C.W. et al. 2013. The Somatic Genomic Landscape of Glioblastoma. *Cell*. **155**(2),pp.462–477.
- Brescia, P. et al. 2012. Current strategies for identification of glioma stem cells: adequate or unsatisfactory? *Journal of Oncology*. **2012**,p.376894.
- Bruhn, M.A. et al. 2014. Proangiogenic tumor proteins as potential predictive or prognostic biomarkers for bevacizumab therapy in metastatic colorectal cancer. *International Journal of Cancer*. **135**(3),pp.731–741.
- Brummer, A. and Hausser, J. 2014. MicroRNA binding sites in the coding region of mRNAs: extending the repertoire of post-transcriptional gene regulation. *BioEssays : news and reviews in molecular, cellular and developmental biology*. **36**(6),pp.617–626.
- Calin, G.A. and Croce, C.M. 2006. MicroRNA signatures in human cancers. *Nature Reviews Cancer*. **6**(11),pp.857–866.
- Calin, G.A. et al. 2002. Frequent deletions and down-regulation of micro- RNA

- genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proc Natl Acad Sci USA*. **99**(24),pp.15524–15529.
- Cancer Genome Atlas Research Network 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. **455**(7216),pp.1061–1068.
- Cantor, A.B. and Shuster, J.J. 1992. Parametric versus non-parametric methods for estimating cure rates based on censored survival data. *Statistics in medicine*. **11**(7),pp.931–937.
- Carroll, D. and Schaefer, A. 2012. General Principles of miRNA Biogenesis and Regulation in the Brain. *Neuropsychopharmacology*,pp.1–16.
- Cassidy, J.J. et al. 2013. miR-9a Minimizes the Phenotypic Impact of Genomic Diversity by Buffering a Transcription Factor. *Cell*. **155**(7),pp.1556–1567.
- Chambers, P.A. et al. 2012. Mutation Detection by Clonal Sequencing of PCR Amplicons and Grouped Read Typing is Applicable to Clinical Diagnostics. *Human Mutation*. **34**(1),pp.248–254.
- Chan, J.K. et al. 2014. MiR-378 as a biomarker for response to anti-angiogenic treatment in ovarian cancer. *Gynecologic oncology*. **133**(3),pp.568-74.
- Chang, S.-C. et al. 2013. Comparison of KRAS Genotype: Therascreen Assay vs. LNA-Mediated qPCR Clamping Assay. *Clinical colorectal cancer*. **12**(3),pp.195–203.
- Cheang, M.C.U. et al. 2009. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *Journal of the National Cancer Institute*. **101**(10),pp.736–750.
- Chen, J. et al. 2014. miR-125b inhibitor enhance the chemosensitivity of glioblastoma stem cells to temozolomide by targeting Bak1. *Tumour biology*. **35**(7),pp.6293–6302.
- Chen, L., Zhang, J., et al. 2012a. Downregulation of miR-221/222 sensitizes glioma cells to temozolomide by regulating apoptosis independently of p53 status. *Oncology reports*. **27**(3),pp.854–860.
- Chen, L., Zhang, W., et al. 2012b. The Putative Tumor Suppressor miR-524-5p Directly Targets Jagged-1 and Hes-1 in Glioma. *Carcinogenesis*. **33**(11),pp.2276-82.
- Chendrimada, T.P. et al. 2005. TRBP recruits the Dicer complex to Ago2 for microRNA processing and gene silencing. *Nature Cell Biology*. **436**(7051),pp.740–744.
- Chi, S.W. et al. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature*. **460**(7254),pp.479–486.
- Chiarugi, P. and Giannoni, E. 2008. Anoikis: a necessary death program for

- anchorage-dependent cells. *Biochemical Pharmacology*. **76**(11),pp.1352–1364.
- Chin, L.J. et al. 2008. A SNP in a let-7 microRNA Complementary Site in the KRAS 3' Untranslated Region Increases Non-Small Cell Lung Cancer Risk. *Cancer research*. **68**(20),pp.8535–8540.
- Chinot, O.L. et al. 2014. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *The New England journal of medicine*. **370**(8),pp.709–722.
- Chong, C.R. and Janne, P.A. 2013. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nature Medicine*. **19**(11),pp.1389–1400.
- Choudhury, Y., Tay, F.C., Lam, D.H., Sandanaraj, E., Tang, C., Ang, B.-T. and Wang, S. 2012a. Attenuated adenosine-to-inosine editing of microRNA-376a* promotes invasiveness of glioblastoma cells. *Journal of Clinical Investigation*. **122**(11),pp.4059-76.
- Choudhury, Y., Tay, F.C., Lam, D.H., Sandanaraj, E., Tang, C., Ang, B.-T. and Wang, S. 2012b. Epigenetics and genetics. MicroRNAs en route to the clinic: progress in validating and targeting microRNAs for cancer therapy. *Journal of Clinical Investigation*. **122**(11),pp.4059–4076.
- Chronaiou, I. et al. 2014. Impacts of MR spectroscopic imaging on glioma patient management. *Acta Oncologica*. **53**(5),pp.580–589.
- Claus, E.B. et al. 2015. Survival and low-grade glioma: the emergence of genetic information. *Neurosurgical Focus*. **38**(1),p.E6.
- Cockle, J.V. et al. 2015. Cell migration in paediatric glioma;characterisation and potential therapeutic targeting. *British journal of cancer*. **112**(4),pp.693–703.
- Cohen, M.H. et al. 2009. FDA drug approval summary: bevacizumab (Avastin) as treatment of recurrent glioblastoma multiforme. *The oncologist*. **14**(11),pp.1131–1138.
- Cohen, M.H. et al. 2003. FDA drug approval summary: gefitinib (ZD1839) (Iressa) tablets. *The oncologist*. **8**(4),pp.303–306.
- Colman, H. et al. 2010. A multigene predictor of outcome in glioblastoma. *Neuro-oncology*. **12**(1),pp.49–57.
- Conde, E. et al. 2013. The ALK translocation in advanced non-small-cell lung carcinomas: preapproval testing experience at a single cancer centre. *Histopathology*. **62**(4),pp.609–616.
- Cooper, D.N. 1983. Eukaryotic DNA methylation. *Human genetics*. **64**(4),pp.315–333.
- Cooper, Z.A. et al. 2014. Response to BRAF inhibition in melanoma is enhanced when combined with immune checkpoint blockade. *Cancer Immunology*

- Research*. **2**(7),pp.643–654.
- Costa, A. et al., 2013. miR-363-5p regulates endothelial cell properties and their communication with hematopoietic precursor cells. *Journal of hematology & oncology*, **6**(1), p.87.
- Cox, D. 1972. Regression models and life tables. *Journal of the Royal Statistical Society Series B-Methodological*. **34**(2),pp.187–220.
- Cushman-Vokoun, A.M. et al. 2013. Comparison study of the performance of the QIAGEN EGFR RGQ and EGFR pyro assays for mutation analysis in non-small cell lung cancer. *American journal of clinical pathology*. **140**(1),pp.7–19.
- Dagan, L.N. et al. 2012. miR-155 regulates HGAL expression and increases lymphoma cell motility. *Blood*. **119**(2),pp.513–520.
- de Larrea, C.F. et al. 2012. Impact of MiRSNPs on survival and progression in patients with multiple myeloma undergoing autologous stem cell transplantation. *Clinical cancer research*. **18**(13),pp.3697–3704.
- de Vries, N.A. et al. 2012. Restricted brain penetration of the tyrosine kinase inhibitor erlotinib due to the drug transporters P-gp and BCRP. *Investigational new drugs*. **30**(2),pp.443–449.
- DeAngelis, L.M. 2001. Brain Tumors. *The New England journal of medicine*. **344**(2),pp.114–123.
- Deeken, J.F. and Loscher, W. 2007. The Blood-Brain Barrier and Cancer: Transporters, Treatment, and Trojan Horses. *Clinical Cancer Research*. **13**(6),pp.1663–1674.
- Del Vecchio, C.A. and Wong, A.J. 2010. Rindopepimut, a 14-mer injectable peptide vaccine against EGFRvIII for the potential treatment of glioblastoma multiforme. *Current opinion in molecular therapeutics*. **12**(6),pp.741–754.
- Deleyrolle, L.P. et al. 2011. Evidence for label-retaining tumour-initiating cells in human glioblastoma. *Brain*. **134**,pp.1331–1343.
- Di Stefano, A.L. et al. 2014. VEGFA SNP rs2010963 is associated with vascular toxicity in recurrent glioblastomas and longer response to bevacizumab. *Journal of neuro-oncology*. **121**(3),pp.499-504.
- Dick, J.E. 1991. Immune-deficient mice as models for human hematopoietic disease. *Molecular genetic medicine*. **1**,pp.77–115.
- Ding, C. et al. 2013. A miR-SNP of the XPO5 gene is associated with advanced non-small-cell lung cancer. *OncoTargets and therapy*. **6**,pp.877–881.
- Dirks, P.B. 2010. Brain tumor stem cells: the cancer stem cell hypothesis writ large. *Molecular Oncology*. **4**(5),pp.420–430.

- Díaz-Martín, J. et al. 2014. A core microRNA signature associated with inducers of the epithelial-to-mesenchymal transition. *The Journal of Pathology*. **232**(3),pp.319–329.
- Dosanjh, M.K. et al. 1991. Kinetics of extension of O6-methylguanine paired with cytosine or thymine in defined oligonucleotide sequences. *Biochemistry*. **30**(49),pp.11595–11599.
- Douville, J. et al. 2009. ALDH1 as a functional marker of cancer stem and progenitor cells. *Stem Cells and Development*. **18**(1),pp.17–25.
- Doyle, F.H. et al. 1981. Imaging of the brain by nuclear magnetic resonance. *Lancet*. **2**(8237),pp.53–57.
- Druker, B.J. et al. 1996. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nature Medicine*. **2**(5),pp.561–566.
- Ducray, F. et al. 2009. Diagnostic and prognostic markers in gliomas. *Current Opinion in Oncology*. **21**(6),pp.537–542.
- Ebert, M.S. and Sharp, P.A. 2012. Roles for MicroRNAs in Conferring Robustness to Biological Processes. *Cell*. **149**(3),pp.515–524.
- Engler, J.R. et al. 2012. Increased microglia/macrophage gene expression in a subset of adult and pediatric astrocytomas. *PLoS ONE*. **7**(8),p.e43339.
- Eraslan, L.E. et al. 2014. Identification of patients with recurrent GBM that benefit from Bevacizumab. *Neuro-Oncology*. **16**(3),pp.iii8–iii9.
- Esteller, M. 2002. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene*. **21**(35),pp.5427–5440.
- Evans, S.M. et al. 2004. Hypoxia is important in the biology and aggression of human glial brain tumors. *Clinical cancer research*. **10**(24),pp.8177–8184.
- Fabbri, M. et al. 2012. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. *Proceedings of the National Academy of Sciences*. **109**(31),pp.e2110–e2116.
- Farazi, T.A. et al. 2012. Bioinformatic analysis of barcoded cDNA libraries for small RNA profiling by next-generation sequencing. *Methods*. **58**(2),pp.171–187.
- Farazi, T.A. et al. 2014. Identification of distinct miRNA target regulation between breast cancer molecular subtypes using AGO2-PAR-CLIP and patient datasets. *Genome Biology*. **15**(1),p.R9.
- Fassl, A. et al. 2012. Notch1 signaling promotes survival of glioblastoma cells via EGFR-mediated induction of anti-apoptotic Mcl-1. *Oncogene*. **31**(44),pp.4698–4708.
- Favero, F. RmiR: Package to work with miRNAs and miRNA targets with R. 1st ed.

- Field, K.M. et al. 2015. Bevacizumab and glioblastoma: Scientific review, newly reported updates, and ongoing controversies. *Cancer*. **121**(7),pp.997-1007.
- Filshtein, T.J. et al. 2012. OrbiD: Origin-based identification of microRNA targets. *Mobile Genetic Elements*. **2**(4),pp.184-192.
- Fine, H.A., 2014. Bevacizumab in Glioblastoma — Still Much to Learn. *The New England journal of medicine*, **370**(8), pp.764-765.
- Fowler, A. et al. 2011. miR-124a is frequently down-regulated in glioblastoma and is involved in migration and invasion. *European journal of cancer*. **47**(6),pp.953-963.
- Friedman, R.C. et al. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research*. **19**(1),pp.92-105.
- Fuchs, E. and Segre, J.A. 2000. Stem cells: a new lease on life. *Cell*. **100**(1),pp.143-155.
- Gabriely, G. et al. 2011. Human glioma growth is controlled by microRNA-10b. *Cancer research*. **71**(10),pp.3563-3572.
- Gabriely, G. et al. 2008. MicroRNA 21 promotes glioma invasion by targeting matrix metalloproteinase regulators. *Molecular and Cellular Biology*. **28**(17),pp.5369-5380.
- Gabrusiewicz, K. et al. 2014. Anti-vascular endothelial growth factor therapy-induced glioma invasion is associated with accumulation of Tie2-expressing monocytes. *Oncotarget*. **5**(8),pp.2208-2220.
- Gan, H.K. et al. 2009. The EGFRvIII variant in glioblastoma multiforme. *Journal of Clinical Neuroscience*. **16**(6),pp.748-754.
- Garcia, D.M. et al. 2011. Weak seed-pairing stability and high target-site abundance decrease the proficiency of. *Nature structural & molecular biology*. **18**(10),pp.1139-1146.
- Garrido, W. et al. 2014. Chemoresistance in high-grade gliomas: relevance of adenosine signalling in stem-like cells of glioblastoma multiforme. *Current drug targets*. **15**(10),pp.931-942.
- Garzon, R. et al. 2008. MicroRNA signatures associated with cytogenetics and prognosis in acute myeloid leukemia. *Blood*. **111**(6),pp.3183-3189.
- Gaspar, N. et al. 2010. MGMT-Independent Temozolomide Resistance in Pediatric Glioblastoma Cells Associated with a PI3-Kinase-Mediated HOX/Stem Cell Gene Signature. *Cancer research*. **70**(22),pp.9243-9252.
- Gebert, L.F.R. et al. 2014. Miravirsen (SPC3649) can inhibit the biogenesis of miR-122. *Nucleic Acids Research*. **42**(1),pp.609-621.

- Geisbrecht, B.V. and Gould, S.J. 1999. The human PICD gene encodes a cytoplasmic and peroxisomal NADP(+)-dependent isocitrate dehydrogenase. *The Journal of biological chemistry*. **274**(43),pp.30527–30533.
- Geiss, G.K. et al. 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature biotechnology*. **26**(3),pp.317–325.
- Genovese, G. et al. 2012. microRNA Regulatory Network Inference Identifies miR-34a as a Novel Regulator of TGF- Signaling in Glioblastoma. *Cancer Discovery*. **2**(8),pp.736–749.
- Gilbert, M.R. et al. 2014. A randomized trial of bevacizumab for newly diagnosed glioblastoma. *The New England journal of medicine*. **370**(8),pp.699–708.
- Giunti, L. et al. 2015. Anti-miR21 oligonucleotide enhances chemosensitivity of T98G cell line to doxorubicin by inducing apoptosis. *American journal of cancer research*. **5**(1),pp.231–242.
- Godlewski, J., et al. 2010. microRNA-451: A conditional switch controlling glioma cell proliferation and migration. *Cell cycle*. **9**(14),pp.2742–2748.
- Godlewski, J., et al. 2010. MicroRNA-451 Regulates LKB1/AMPK Signaling and Allows Adaptation to Metabolic Stress in Glioma Cells. *Molecular Cell*. **37**(5),pp.620–632.
- Goeman, J.J. 2010. L1 penalized estimation in the Cox proportional hazards model. *Biometrics journal*. **52**(1),pp.70-84.
- Goff, L.A. et al. 2009. Ago2 immunoprecipitation identifies predicted microRNAs in human embryonic stem cells and neural precursors. *PLoS ONE*. **4**(9),p.e7192.
- Gokhale, A. et al. 2010. Distinctive microRNA signature of medulloblastomas associated with the WNT signaling pathway. *Journal of cancer research and therapeutics*. **6**(4),pp.521–529.
- Gomez, G.G. et al. 2014. Suppression of MicroRNA-9 by Mutant EGFR Signaling Upregulates FOXP1 to Enhance Glioblastoma Tumorigenicity. *Cancer research*. **74**(5),pp.1429–1439.
- Gonda, D.D. et al. 2014. The Cancer Genome Atlas expression profiles of low-grade gliomas. *Neurosurgical Focus*. **36**(4),p.e23.
- Goodenberger, M.L. and Jenkins, R.B. 2012. Genetics of adult glioma. *Cancer Genetics*. **205**(12),pp.613–621.
- Gregory, R.I. 2005. MicroRNA Biogenesis and Cancer. *Cancer research*. **65**(9),pp.3509–3512.
- Griffiths-Jones, S. et al. 2008. miRBase: Tools for microRNA genomics. *Nucleic Acids Research*. **36**(1),pp.D154–D158.

- Grimson, A. et al. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Molecular Cell*. **27**(1),pp.91–105.
- Guessous, F. et al. 2013. Oncogenic effects of miR-10b in glioblastoma stem cells. *Journal of neuro-oncology*. **112**(2),pp.153–163.
- Guo, C.-W. et al. 2013. Culture under low physiological oxygen conditions improves the stemness and quality of induced pluripotent stem cells. *Journal of Cellular Physiology*. **228**(11),pp.2159–2166.
- Gwak, J.M. et al. 2014. MicroRNA-9 is associated with epithelial-mesenchymal transition, breast cancer stem cell phenotype, and tumor progression in breast cancer. *Breast cancer research and treatment*. **147**(1),pp.39–49.
- Habig, J.W. et al. 2007. miRNA Editing—We Should Have Inosine This Coming. *Molecular Cell*. **25**(6),pp.792–793.
- Haemmig, S. et al. 2014. miR-125b controls apoptosis and temozolomide resistance by targeting TNFAIP3 and NKIRAS2 in glioblastomas. *Cell Death and Disease*. **5**,p.e1279.
- Hafner, M. et al. 2012. Barcoded cDNA library preparation for small RNA profiling by next-generation sequencing. *Methods*. **58**(2),pp.164–170.
- Hafner, M. et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell*. **141**(1),pp.129–141.
- Hall, J.S. et al. 2012. Enhanced stability of microRNA expression facilitates classification of FFPE tumour samples exhibiting near total mRNA degradation. *British journal of cancer*. **107**(4),pp.684–694.
- Hanahan, D. and Weinberg, R.A. 2011. Hallmarks of Cancer: The Next Generation. *Cell*. **144**(5),pp.646–674.
- Hao, J. et al. 2014. MicroRNA control of epithelial-mesenchymal transition in cancer stem cells. *International Journal of Cancer*. **135**(5),pp.1019–27.
- Harris, S.L. and Levine, A.J. 2005. The p53 pathway: positive and negative feedback loops. *Oncogene*. **24**(17),pp.2899–2908.
- Hasselbalch, B. et al. 2010. Cetuximab, bevacizumab, and irinotecan for patients with primary glioblastoma and progression after radiation therapy and temozolomide: a phase II trial. *Neuro-oncology*. **12**(5),pp.508–516.
- Havens, M.A. et al. 2012. Biogenesis of mammalian microRNAs by a non-canonical processing pathway. *Nucleic Acids Research*. **40**(10),pp.4626–4640.
- Havrda, M.C. et al. 2014. Id2 mediates oligodendrocyte precursor cell maturation arrest and is tumorigenic in a PDGF-rich microenvironment. *Cancer research*. **74**(6),pp.1822–1832.

- Hayes, J.L. et al. 2013. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics*. **102**(3),pp.174–181.
- Hegde, M. et al. 2014. Novel approaches and mechanisms of immunotherapy for glioblastoma. *Discovery medicine*. **17**(93),pp.145–154.
- Hegi, M.E. et al. 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. *The New England journal of medicine*. **352**(10),pp.997–1003.
- Hein, M. and Graver, S. 2013. Tumor cell response to bevacizumab single agent therapy in vitro. *Cancer Cell International*. **13**(1),pp.1–1.
- Hu, C.-W. et al. 2013. Quantitative Proteomics Reveals Diverse Roles of miR-148a from Gastric Cancer Progression to Neurological Development. *Journal of Proteome Research*. **12**(9),pp.3993–4004.
- Hu, J. et al. 2014. Human miR-1228 as a stable endogenous control for the quantification of circulating microRNAs in cancer patients. *International Journal of Cancer*. **135**(5),pp.1187–94.
- Hu, X. et al. 2010. A microRNA expression signature for cervical cancer prognosis. *Cancer research*. **70**(4),pp.1441–1448.
- Huang, Y. et al. 2014. Oligodendrocyte Progenitor Cells Promote Neovascularization in Glioma by Disrupting the Blood-Brain Barrier. *Cancer research*. **74**(4),pp.1011–1021.
- Hubbard, S.R. and Till, J.H. 2000. Protein tyrosine kinase structure and function. *Annual review of biochemistry*. **69**,pp.373–398.
- Hummel, R. et al. 2011. Mir-148a improves response to chemotherapy in sensitive and resistant oesophageal adenocarcinoma and squamous cell carcinoma cells. *Journal of gastrointestinal surgery*. **15**(3),pp.429–438.
- Hwang, W.-L. et al. 2014. MicroRNA-146a directs the symmetric division of Snail-dominant colorectal cancer stem cells. *Nature Cell Biology*. **16**(3),pp.268–280.
- Ignatova, T.N. et al. 2002. Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro. *Glia*. **39**(3),pp.193–206.
- Inskip, P.D. et al. 2010. Brain cancer incidence trends in relation to cellular telephone use in the United States. *Neuro-oncology*. **12**(11),pp.1147–1151.
- Inskip, P.D. et al. 1995. Etiology of brain tumors in adults. *Epidemiologic reviews*. **17**(2),pp.382–414.
- Jacob, F. et al. 2013. Careful selection of reference genes is required for reliable performance of RT-qPCR in human normal and cancer cell lines. *PLoS ONE*. **8**(3),p.e59180.

- Janssen, H.L.A. et al. 2013. Treatment of HCV infection by targeting microRNA. *The New England journal of medicine*. **368**(18),pp.1685-94.
- Jarjour, A.A., Kennedy, T.E., 2004. Oligodendrocyte Precursors on the Move: Mechanisms Directing Migration. *Neuroscientist*. **10**,pp.99–105.
- Jarry, J. et al. 2014. The validity of circulating microRNAs in oncology: Five years of challenges and contradictions. *Molecular Oncology*. **8**(4),pp.819-29.
- Jeon, H.M. et al. 2011. ID4 Imparts Chemoresistance and Cancer Stemness to Glioma Cells by Derepressing miR-9*-Mediated Suppression of SOX2. *Cancer research*. **71**(9),pp.3410–3421.
- Jeuken, J.W.M. et al. 2004. Molecular pathogenesis of oligodendroglial tumors. *Journal of neuro-oncology*. **70**(2),pp.161–181.
- Jiao, Y. et al. 2012. Frequent ATRX, CIC, and FUBP1 mutations refine the classification of malignant gliomas. *Oncotarget*. **3**(7),pp.709–722.
- Jin, Y. et al. 2013. Evaluating the microRNA targeting sites by luciferase reporter gene assay. *Methods in molecular biology*. **936**,pp.117–127.
- John, B. et al. 2004. Human MicroRNA targets. *PLoS biology*. **2**(11),p.e363.
- Johnson, B.E. et al. 2014. Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science*. **343**(6167),pp.189–193.
- Joseph, J.V. et al. 2015. Hypoxia enhances migration and invasion in glioblastoma by promoting a mesenchymal shift mediated by the HIF1alpha-ZEB1 axis. *Cancer Letters*. **359**(1),pp.107-16.
- Kaminska, B. et al. 2013. TGF beta signaling and its role in glioma pathogenesis. *Advances in experimental medicine and biology*. **986**,pp.171–187.
- Kather, J.N. et al., 2014. Angiopoietin-1 is regulated by miR-204 and contributes to corneal neovascularization in KLEIP-deficient mice. *Investigative ophthalmology & visual science*, **55**(7), pp.4295–4303.
- Kefas, B. et al. 2013. A miR-297/hypoxia/DGK- axis regulating glioblastoma survival. *Neuro-oncology*. **15**(12),pp.1652–1663.
- Kelnar, K. et al. 2014. Quantification of therapeutic miRNA mimics in whole blood from nonhuman primates. *Analytical Chemistry*. **86**(3),pp.1534–1542.
- Killela, P.J. et al. 2013. The genetic landscape of anaplastic astrocytoma. *Oncotarget*. **5**(6),pp.1452-7.
- Kim, J. et al. 2014. microRNA-148a is a prognostic oncomiR that targets MIG6 and BIM to regulate EGFR and apoptosis in glioblastoma. *Cancer research*. **74**(5),pp.1541-53.

- Kim, N.H. et al. 2011. A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition. *The Journal of cell biology*. **195**(3),pp.417–433.
- Kim, T.-M. et al. 2011. A Developmental Taxonomy of Glioblastoma Defined and Maintained by MicroRNAs. *Cancer research*. **71**(9),pp.3387–3399.
- Kim, Y.-H. et al. 2011. TET2 promoter methylation in low-grade diffuse gliomas lacking IDH1/2 mutations. *Journal of Clinical Pathology*. **64**(10),pp.850–852.
- Kim, Y.-K. et al. 2012. Short Structured RNAs with Low GC Content Are Selectively Lost during Extraction from a Small Number of Cells. *Molecular Cell*. **46**(6),pp.893–895.
- Kloosterhof, N.K. et al. 2013. Molecular subtypes of glioma identified by genome-wide methylation profiling. *Genes, Chromosomes and Cancer*. **52**(7),pp.665–674.
- Koboldt, D.C. et al. 2012. Comprehensive molecular portraits of human breast tumours. *Nature*. **490**(7418),pp.61–70.
- Kolbert, C.P. et al. 2013. Multi-Platform Analysis of MicroRNA Expression Measurements in RNA from Fresh Frozen and FFPE Tissues. *PLoS ONE*. **8**(1),p.e52517.
- Kong, W. et al. 2008. MicroRNA-155 is regulated by the transforming growth factor beta/Smad pathway and contributes to epithelial cell plasticity by targeting RhoA. *Molecular and Cellular Biology*. **28**(22),pp.6773–6784.
- Koshy, M. et al. 2011. Improved survival time trends for glioblastoma using the SEER 17 population-based registries. *Journal of neuro-oncology*. **107**(1),pp.207–212.
- Kotorashvili, A. et al. 2012. Effective DNA/RNA co-extraction for analysis of microRNAs, mRNAs, and genomic DNA from formalin-fixed paraffin-embedded specimens. *PLoS ONE*. **7**(4),p.e34683.
- Krek, A. et al. 2005. Combinatorial microRNA target predictions. *Nature Genetics*. **37**(5),pp.495–500.
- Krex, D. et al. 2007. Long-term survival with glioblastoma multiforme. *Brain*. **130**(10),pp.2596–2606.
- Krichevsky, A.M. et al. 2006. Specific microRNAs modulate embryonic stem cell-derived neurogenesis. *Stem cells*. **24**(4),pp.857–864.
- Krol, J. et al. 2010. The widespread regulation of microRNA biogenesis, function and decay. *Nature reviews. Genetics*. **11**(9),pp.597–610.
- Kuang, Y. et al. 2013. Repression of Dicer is associated with invasive phenotype and chemoresistance in ovarian cancer. *Oncology Letters*. **5**(4),pp.1149–1154.

- Kwak, E.L. et al. 2010. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *The New England journal of medicine*. **363**(18),pp.1693–1703.
- Kuo, T.-C. et al., 2013. Angiopoietin-like protein 1 suppresses SLUG to inhibit cancer cell motility. *Journal of Clinical Investigation*, **123**(3), pp.1082–1095.
- Lai, A. et al. 2011. Evidence for sequenced molecular evolution of IDH1 mutant glioblastoma from a distinct cell of origin. *Journal of Clinical Oncology*. **29**(34),pp.4482–4490.
- Lakomy, R. et al. 2011. MiR-195, miR-196b, miR-181c, miR-21 expression levels and O-6-methylguanine-DNA methyltransferase methylation status are associated with clinical outcome in glioblastoma patients. *Cancer science*. **102**(12),pp.2186–2190.
- Lamouille, S. et al. 2014. Molecular mechanisms of epithelial– mesenchymal transition. *Nature Publishing Group*. **15**(3),pp.178–196.
- Landgraf, P. et al. 2007. A Mammalian microRNA Expression Atlas Based on Small RNA Library Sequencing. *Cell*. **129**(7),pp.1401–1414.
- Larjavaara, S. et al. 2007. Incidence of gliomas by anatomic location. *Neuro-oncology*. **9**(3),pp.319–325.
- Laws, E.R. et al. 2003. Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the Glioma Outcomes Project. *Journal of Neurosurgery*. **99**(3),pp.467–473.
- Lee, H.K. et al. 2013. Mesenchymal stem cells deliver synthetic microRNA mimics to glioma cells and glioma stem cells and inhibit their cell migration and self-renewal. *Oncotarget*. **4**(2),pp.346–361.
- Lee, J. et al. 2006. Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines. *Cancer cell*. **9**(5),pp.391–403.
- Lee, R.C. et al. 1993. The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell*. **75**(5),pp.843–854.
- Lei, L. et al. 2011. Glioblastoma Models Reveal the Connection between Adult Glial Progenitors and the Proneural Phenotype M. S. Lesniak, ed. *PLoS ONE*. **6**(5),p.e20041.
- Lerner, R.G. and Petritsch, C. 2014. A microRNA-operated switch of asymmetric-to-symmetric cancer stem cell divisions. *Nature*. **16**(3),pp.212–214.
- Leshkowitz, D. et al. 2013. Differences in microRNA detection levels are technology and sequence dependent. *RNA*. **19**(4),pp.527–538.
- Letzen, B.S. et al. 2010. MicroRNA Expression Profiling of Oligodendrocyte Differentiation from Human Embryonic Stem Cells R. Linden, ed. *PLoS ONE*.

5(5),p.e10480.

- Lewis, B.P. et al. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell*. **120**(1),pp.15–20.
- Lewis, C.M. et al. 2012. A phase II study of gefitinib for aggressive cutaneous squamous cell carcinoma of the head and neck. *Clinical cancer research*. **18**(5),pp.1435–1446.
- Li, J. et al. 2007. Comparison of miRNA expression patterns using total RNA extracted from matched samples of formalin-fixed paraffin-embedded (FFPE) cells and snap frozen cells. *BMC Biotechnology*. **7**,p.36.
- Li, Q. et al. 2014a. Gene amplification of EGFR and its clinical significance in various cervical (lesions) lesions using cytology and FISH. *International journal of clinical and experimental pathology*. **7**(5),pp.2477–2483.
- Li, R. et al. 2014b. Identification of intrinsic subtype-specific prognostic microRNAs in primary glioblastoma. *Journal of Experimental & Clinical Cancer Research*. **33**(1),pp.1–6.
- Li, Yu and Kowdley, K.V. 2012. Method for microRNA isolation from clinical serum samples. *Analytical Biochemistry*. **431**(1),pp.69–75.
- Li, Yue et al. 2012. Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. *Nucleic Acids Research*. **40**(10),pp.4298–4305.
- Liang, Y. et al. 2007. Characterization of microRNA expression profiles in normal human tissues. *BMC genomics*. **8**(1),p.166.
- Lin, J. et al. 2012. MicroRNA-10b pleiotropically regulates invasion, angiogenicity and apoptosis of tumor cells resembling mesenchymal subtype of glioblastoma multiforme. *Cell Death and Disease*. **3**(10),p.e398–12.
- Lindberg, N. et al. 2014. Oncogenic signaling is dominant to cell of origin and dictates astrocytic or oligodendroglial tumor development from oligodendrocyte precursor cells. *The Journal of neuroscience*. **34**(44),pp.14644–14651.
- Liu, Can et al. 2012. Distinct microRNA expression profiles in prostate cancer stem/progenitor cells and tumor-suppressive functions of let-7. *Cancer research*. **72**(13),pp.3393–3404.
- Liu, Chong et al. 2011a. Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell*. **146**(2),pp.209–221.
- Liu, Michelle X et al. 2014. Epigenetic silencing of microRNA-199b-5p is associated with acquired chemoresistance via activation of JAG1-Notch1 signaling in ovarian cancer. *Oncotarget*. **5**(4),pp.944–958.

- Liu, Pixu et al. 2009. Targeting the phosphoinositide 3-kinase pathway in cancer. *Nature reviews. Drug discovery*. **8**(8),pp.627–644.
- Liu, Xiqiang et al. 2011b. MicroRNA-138 suppresses epithelial-mesenchymal transition in squamous cell carcinoma cell lines. *The Biochemical journal*. **440**(1),pp.23–31.
- Liu, Yuangang et al. 2012. NF-kappaB repression by PIAS3 mediated RelA SUMOylation. *PLoS ONE*. **7**(5),p.e37636.
- Lo, S.-S. et al. 2012. Overexpression of miR-370 and downregulation of its novel target TGFbeta-RII contribute to the progression of gastric carcinoma. *Oncogene*. **31**(2),pp.226–237.
- Lopotova, T. et al. 2011. Leukemia Research. *Leukemia Research*. **35**(7),pp.974–977.
- Louis, D.N. et al. 2014. International Society of Neuropathology-Haarlem Consensus Guidelines for Nervous System Tumor Classification and Grading. *Brain Pathology*. **24**(5),pp.429–435.
- Louis, D.N., et al. 2007a. *WHO Classification of Tumours of the Central Nervous System*. 4 ed. Lyon: International Agency for Research on Cancer.
- Louis, D.N. et al. 2007b. The 2007 WHO classification of tumours of the central nervous system. *Acta Neuropathologica*. **114**(2),pp.97–109.
- Lu, J. et al. 2005. MicroRNA expression profiles classify human cancers. *Nature Cell Biology*. **435**(7043),pp.834–838.
- Lu, Y. et al. 2012. MicroRNA profiling and prediction of recurrence/relapse-free survival in stage I lung cancer. *Carcinogenesis*. **33**(5),pp.1046–1054.
- Luo, S. 2012. MicroRNA expression analysis using the Illumina microRNA-Seq Platform. *Methods in molecular biology*. **822**,pp.183–188.
- Lv, S. et al. 2012. Correlation of EGFR, IDH1 and PTEN status with the outcome of patients with recurrent glioblastoma treated in a phase II clinical trial with the EGFR-blocking monoclonal antibody cetuximab. *International journal of oncology*. **41**(3),pp.1029–1035.
- Ma, L. et al. 2010. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nature cell Biology*. **2**(3),pp.247–56.
- Manterola, L. et al. 2014. A small noncoding RNA signature found in exosomes of GBM patient serum as a diagnostic tool. *Neuro-oncology*. **16**(4),pp.520–527.
- Mao, P. et al. 2013. Mesenchymal glioma stem cells are maintained by activated glycolytic metabolism involving aldehyde dehydrogenase 1A3. *Proceedings of the National Academy of Sciences*. **110**(21),pp.8644–8649.

- Marchio, C. et al. 2015. Predictive Diagnostic Pathology in the Target Therapy Era in Breast Cancer. *Current drug targets*. [Epub ahead of print].
- Margison G. P. and Santibáñez-Koref M. F. 2002. O6-alkylguanine-DNA alkyltransferase: role in carcinogenesis and chemotherapy. *BioEssays*. **24**(3), pp.255–266.
- Martin, E.R. et al. 2010. SeqEM: an adaptive genotype-calling approach for next-generation sequencing studies. *Bioinformatics*. **26**(22),pp.2803–2810.
- Mauro, M.J. and Druker, B.J. 2001. STI571: targeting BCR-ABL as therapy for CML. *The oncologist*. **6**(3),pp.233–238.
- Mavrakis, K.J. et al. 2011. A cooperative microRNA-tumor suppressor gene network in acute T-cell lymphoblastic leukemia (T-ALL). *Nature Genetics*. **43**(7),pp.673–678.
- McCord, A.M. et al. 2009. Physiologic oxygen concentration enhances the stem-like properties of CD133+ human glioblastoma cells in vitro. *Molecular cancer research*. **7**(4),pp.489–497.
- McKinney, P.A. 2004. Brain tumours: incidence, survival, and aetiology. *Journal of Neurology, Neurosurgery & Psychiatry*. **75**(2),pp.ii12–ii17.
- McShane, L.M. et al. 2013. Criteria for the use of omics-based predictors in clinical trials. *Nature*. **502**(7471),pp.317–320.
- Mellinghoff, I.K. et al. 2005. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *The New England journal of medicine*. **353**(19),pp.2012–2024.
- Melo, S.A. et al. 2010. A Genetic Defect in Exportin-5 Traps Precursor MicroRNAs in the Nucleus of Cancer Cells. *Cancer cell*. **18**(4),pp.303–315.
- Merritt, W.M. et al. 2008. Dicer, Drosha, and outcomes in patients with ovarian cancer. *The New England journal of medicine*. **359**(25),pp.2641–2650.
- Meyer, M. et al. 2015. Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity. *Proceedings of the National Academy of Sciences*. **112**(3),pp.851–856.
- Mikheeva, S.A. et al. 2010. TWIST1 promotes invasion through mesenchymal change in human glioblastoma. *Molecular cancer*. **9**(1),p.194.
- Milner, R., Edwards, G., Streuli, C., Ffrench-Constant, C., 1996. A role in migration for the alpha V beta 1 integrin expressed on oligodendrocyte precursors. *J Neurosci*. **16**,pp.7240–7252.

- Miranda, K.C. et al. 2006. A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell*. **126**(6),pp.1203–1217.
- Mitchell, P.S. et al. 2008. Circulating microRNAs as stable blood-based markers for cancer detection. *Proceedings of the National Academy of Sciences*. **105**(30),pp.10513–10518.
- Mizoguchi, M. et al. 2012. MicroRNAs in Human Malignant Gliomas. *Journal of Oncology*. **2012**(10),pp.1–7.
- Montano, N. et al. 2011. Expression of EGFRvIII in glioblastoma: prognostic significance revisited. *Neoplasia*. **13**(12),pp.1113–1121.
- Morgan, J.E. et al. 2010. Genetic diagnosis of familial breast cancer using clonal sequencing. *Human Mutation*. **31**(4),pp.484–491.
- Morin, R.D. et al. 2010. Preparation and analysis of microRNA libraries using the Illumina massively parallel sequencing technology. *Methods in molecular biology*. **650**,pp.173–199.
- Moskwa, P. et al. 2014. A functional screen identifies miRs that induce radioresistance in glioblastomas. *Molecular cancer research*. **12**(12),pp.1767–1778.
- Mountz, J. et al. 2014. Malignant gliomas: current perspectives in diagnosis, treatment, and early response assessment using advanced quantitative imaging methods. *Cancer Management and Research*. **6**,pp.149-70.
- Mucaj, V. et al. 2014. MicroRNA-124 expression counteracts pro-survival stress responses in glioblastoma. *Oncogene*. **34**(17),pp.2204-14.
- Muller, P.A.J. et al. 2011. p53 and its mutants in tumor cell migration and invasion. *The Journal of cell biology*. **192**(2),pp.209–218.
- Munoz, J.L. et al. 2013. Delivery of Functional Anti-miR-9 by Mesenchymal Stem Cell-derived Exosomes to Glioblastoma Multiforme Cells Conferred Chemosensitivity. *Molecular Therapy Nucleic Acids*. **2**p.e126.
- Munoz, J.L. et al. 2014. Temozolomide resistance in glioblastoma occurs by miRNA-9-targeted PTCH1, independent of sonic hedgehog level. *Oncotarget*. **6**(2),pp.1190-201.
- Naeini, K.M. et al. 2013. Identifying the mesenchymal molecular subtype of glioblastoma using quantitative volumetric analysis of anatomic magnetic resonance images. *Neuro-oncology*. **15**(5),pp.626–634.
- Nana-Sinkam, S.P. and Croce, C.M. 2012. Clinical applications for microRNAs in cancer. *Clinical Pharmacology & Therapeutics*. **93**(1),pp.98–104.
- Nevins, J.R. 2001. The Rb/E2F pathway and cancer. *Human Molecular Genetics*.

- 10(7)**,pp.699–703.
- Niyazi, M. et al. 2011. MiRNA expression patterns predict survival in glioblastoma. *Radiation Oncology*. **6(1)**,p.153.
- Northcott, P.A. et al. 2012. Rapid, reliable, and reproducible molecular sub-grouping of clinical medulloblastoma samples. *Acta Neuropathologica*. **123(4)**,pp.615–626.
- Noushmehr, H. et al. 2010. Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer cell*. **17(5)**,pp.510–522.
- Obad, S. et al. 2011. Silencing of microRNA families by seed-targeting tiny LNAs. *Nature Genetics*. **43(4)**,pp.371–378.
- Ohgaki, H. and Kleihues, P. 2013. The definition of primary and secondary glioblastoma. *Clinical cancer research*. **19(4)**,pp.764–772.
- Ohgaki, H. et al. 2004. Genetic pathways to glioblastoma: a population-based study. *Cancer research*. **64(19)**,pp.6892–6899.
- Omuro, A. et al. 2014. Phase II study of bevacizumab, temozolomide, and hypofractionated stereotactic radiotherapy for newly diagnosed glioblastoma. *Clinical cancer research*. **20(19)**,pp.5023–5031.
- Ozawa, T. et al. 2014. Most Human Non-GCIMP Glioblastoma Subtypes Evolve from a Common Proneural-like Precursor Glioma. *Cancer cell*. **26(2)**,pp.288–300.
- Ozawa, T. et al. 2010. PDGFRA gene rearrangements are frequent genetic events in PDGFRA-amplified glioblastomas. *Genes & development*. **24(19)**,pp.2205–2218.
- Papadopoulos, G.L. et al. 2009. The database of experimentally supported targets: a functional update of TarBase. *Nucleic Acids Research*. **37**,pp.155–8.
- Papagiannakopoulos, T. et al. 2008. MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. *Cancer research*. **68(19)**,pp.8164–8172.
- Paraskevopoulou, M.D. et al. 2013. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Research*. **41**,pp.W169–73.
- Pardini, B. et al. 2013. Variation within 3'-UTRs of base excision repair genes and response to therapy in colorectal cancer patients: A potential modulation of microRNAs binding. *Clinical cancer research*. **19(21)**,pp.6044–6056.
- Parsons, D.W. et al. 2008. An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science*. **321(5897)**,pp.1807–1812.
- Patel, A.P. et al. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*. **344(6190)**,pp.1396–1401.

- Paul, M.S. and Bass, B.L. 1998. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *The EMBO journal*. **17**(4),pp.1120–1127.
- Peltier, H.J. and Latham, G.J. 2008. Normalization of microRNA expression levels in quantitative RT-PCR assays: identification of suitable reference RNA targets in normal and cancerous human solid tissues. *RNA*. **14**(5),pp.844–852.
- Pencheva, N. and Tavazoie, S.F. 2013. Control of metastatic progression by microRNA regulatory networks. *Nature Cell Biology*. **15**(6),pp.546–554.
- Peruzzi, P. et al. 2013. MicroRNA-128 coordinately targets Polycomb Repressor Complexes in glioma stem cells. *Neuro-oncology*. **15**(9),pp.1212–24.
- Phillips, H.S. et al. 2006. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer cell*. **9**(3),pp.157–173.
- Piao, Y. et al. 2013. Acquired resistance to anti-VEGF therapy in glioblastoma is associated with a mesenchymal transition. *Clinical cancer research*. **19**(16),pp.4392–4403.
- Piao, Y. et al. 2012. Glioblastoma resistance to anti-VEGF therapy is associated with myeloid cell infiltration, stem cell accumulation, and a mesenchymal phenotype. *Neuro-oncology*. **14**(11),pp.1379–1392.
- Pirzkall, A. et al. 2002. Metabolic imaging of low-grade gliomas with three-dimensional magnetic resonance spectroscopy. *International journal of radiation oncology, biology, physics*. **53**(5),pp.1254–1264.
- Ponten, J. and Macintyre, E.H. 1968. Long term culture of normal and neoplastic human glia. *Acta Pathologica Microbiologica Scandinavica*. **74**(4),pp.465–486.
- Popovici-Muller, J. et al. 2012. Discovery of the First Potent Inhibitors of Mutant IDH1 That Lower Tumor 2-HG in Vivo. *ACS Medicinal Chemistry Letters*. **3**(10),pp.850–855.
- Pradervand, S. et al. 2010. Concordance among digital gene expression, microarrays, and qPCR when measuring differential expression of microRNAs. *BioTechniques*. **48**(3),pp.219–222.
- Prados, M.D. 2014. ‘We will know it when we see it;’ bevacizumab and glioblastoma. *Neuro-oncology*. **16**(4),pp.469–470.
- Pritchard, C.C. et al. 2012. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. *Cancer prevention research*. **5**(3),pp.492–497.
- Qiu, Z. and Dai, Y. 2014. Roadmap of miR-122-related clinical application from bench to bedside. *Expert opinion on investigational drugs*. **23**(3),pp.347–355.
- Quail, M.A. et al. 2008. A large genome center's improvements to the Illumina

- sequencing system. *Nature Methods*. **5**(12),pp.1005–1010.
- Quintavalle, C., Donnarumma, E., et al. 2012a. Effect of miR-21 and miR-30b/c on TRAIL-induced apoptosis in glioma cells. *Oncogene*. **32**(34),pp.4001-8.
- Quintavalle, C. et al. 2012b. miR-221/222 overexpression in human glioblastoma increases invasiveness by targeting the protein phosphate PTPmu. *Oncogene*. **31**(7),pp.858–868.
- Rani, S.B. et al. 2013. MiR-145 functions as a tumor-suppressive RNA by targeting Sox9 and adducin 3 in human glioma cells. *Neuro-oncology*. **15**(10),pp.1302–1316.
- Reardon, D.A. and Wen, P.Y. 2014. Targeted therapies: Further delineating bevacizumab's response spectrum. *Nature Reviews Clinical Oncology*. **11**(5),pp.243–244.
- Reardon, D.A., et al. 2014a. Immunotherapy advances for glioblastoma. *Neuro-oncology*. **16**(11),pp.1441–1458.
- Reardon, D.A., et al. 2014b. Targeted molecular therapies against epidermal growth factor receptor: Past experiences and challenges. *Neuro-oncology*. **16**(8),pp.viii7–viii13.
- Reguart, N. and Remon, J. 2015. Common EGFR-mutated subgroups (Del19/L858R) in advanced non-small-cell lung cancer: chasing better outcomes with tyrosine-kinase inhibitors. *Future oncology*. **11**(8):1245-57.
- Riemenschneider, M.J. et al. 2010. Molecular diagnostics of gliomas: state of the art. *Acta Neuropathologica*. **120**(5),pp.567–584.
- Risueño, R.M. et al. 2012. Inability of human induced pluripotent stem cell-hematopoietic derivatives to downregulate microRNAs in vivo reveals a block in xenograft hematopoietic regeneration. *Stem cells*. **30**(2),pp.131–139.
- Rivera-Diaz, M. et al. 2015. MicroRNA-27a distinguishes glioblastoma multiforme from diffuse and anaplastic astrocytomas and has prognostic value. *American journal of cancer research*. **5**(1),pp.201–218.
- Robinson, M.D. et al. 2009. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)*. **26**(1),pp.139–140.
- Rohle, D. et al. 2013. An Inhibitor of Mutant IDH1 Delays Growth and Promotes Differentiation of Glioma Cells. *Science*. **340**(6132),pp.626–630.
- Rosenfeld, N. et al. 2008. MicroRNAs accurately identify cancer tissue origin. *Nature Biotechnology*. **26**(4),pp.462–469.
- Rukov, J.L. et al. 2013. Pharmaco-miR: linking microRNAs and drug effects. *Briefings in Bioinformatics*. **15**(4),pp.648-59.

- Sana, J. et al. 2014. Risk score based on microRNA expression signature is independent prognostic classifier of glioblastoma patients. *Carcinogenesis*. **35**(12),pp.2756-62.
- Sandhu, S.K. et al. 2013. B-cell malignancies in microRNA Emu-miR-17~92 transgenic mice. *Proceedings of the National Academy of Sciences*. **110**(45),pp.18208–18213.
- Scholl, V. et al. 2012. miRNA-451: A putative predictor marker of Imatinib therapy response in chronic myeloid leukemia. *Leukemia Research*. **36**(1),pp.119–121.
- Schraivogel, D. et al. 2011. CAMTA1 is a novel tumour suppressor regulated by miR-9/9* in glioblastoma stem cells. *The EMBO journal*. **30**(20),pp.4309–4322.
- Schroeder, A. et al. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC molecular biology*. **7**,p.3.
- Schuster, J. et al. 2015. A phase II, multicenter trial of rindopepimut (CDX-110) in newly diagnosed glioblastoma: the ACT III study. *Neuro-oncology*. [Epub ahead of print]
- Schwartzentruber, J. et al. 2012. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature*. **482**(7384),pp.226-31.
- Schwarzenbach, H. et al. 2014. Clinical relevance of circulating cell-free microRNAs in cancer. *Nature Reviews Clinical Oncology*. **11**(3),pp.145-56.
- Sebio, A. et al. 2013. The LCS6 polymorphism in the binding site of let-7 microRNA to the KRAS 3'-untranslated region: its role in the efficacy of anti-EGFR-based therapy in metastatic colorectal cancer patients. *Pharmacogenetics and genomics*. **23**(3),pp.142–147.
- Sehmer, E.A.J. et al. 2014. Incidence of glioma in a northwestern region of England, 2006-2010. *Neuro-oncology*. **16**(7),pp.971–974.
- Selvey, S. et al. 2001. Beta-actin--an unsuitable internal control for RT-PCR. *Molecular and Cellular Probes*. **15**(5),pp.307–311.
- Sempere, L.F. et al. 2004. Expression profiling of mammalian microRNAs uncovers a subset of brain-expressed microRNAs with possible roles in murine and human neuronal differentiation. *Genome Biology*. **5**(3),p.R13.
- Serna, E. et al. 2014. Correlation between EGFR amplification and the expression of microRNA-200c in primary glioblastoma multiforme. *PLoS ONE*. **9**(7),p.e102927.
- Sethi, P. and Lukiw, W.J. 2009. Micro-RNA abundance and stability in human brain: specific alterations in Alzheimer's disease temporal lobe neocortex. *Neuroscience letters*. **459**(2),pp.100–104.
- Sethupathy, P. et al. 2006. TarBase: A comprehensive database of experimentally

- supported animal microRNA targets. *RNA*. **12**(2),pp.192–197.
- Seton-Rogers, S. 2012. MicroRNAs: Editing changes the meaning. *Nature Reviews Cancer*. **12**(12),pp.797.
- Shan, F. et al. 2014. HIF-1 Alpha-Induced Up-Regulation of miR-9 Contributes to Phenotypic Modulation in Pulmonary Artery Smooth Muscle Cells During Hypoxia. *Journal of Cellular Physiology*. **229**(10),pp.1511–1520.
- Shang, C. et al. 2015. MiR-21 up-regulation mediates glioblastoma cancer stem cells apoptosis and proliferation by targeting FASLG. *Molecular biology reports*. **42**(3),pp.721–727.
- Shannon, K.M. 2002. Resistance in the land of molecular cancer therapeutics. *Cancer cell*. **2**(2),pp.99–102.
- Shay, J.W. and Wright, W.E. 2000. Hayflick, his limit, and cellular ageing. *Nature reviews*. **1**(1):72-6.
- Shen, J. et al. 2014. EGFR modulates microRNA maturation in response to hypoxia through phosphorylation of AGO2. *Nature*. **497**(7449),pp.383–387.
- Shi, L. et al. 2012. MicroRNA-125b-2 confers human glioblastoma stem cells resistance to temozolomide through the mitochondrial pathway of apoptosis. *International journal of oncology*. **40**(1),pp.119–129.
- Shi, L. et al. 2014a. miR-125b inhibitor may enhance the invasion-prevention activity of temozolomide in glioblastoma stem cells by targeting PIAS3. *BioDrugs : clinical immunotherapeutics, biopharmaceuticals and gene therapy*. **28**(1),pp.41–54.
- Shi, Z. et al., 2014b. MiR-124 governs glioma growth and angiogenesis and enhances chemosensitivity by targeting R-Ras and N-Ras. *Neuro-oncology*, **16**(10), pp.1341–1353.
- Shoemaker, R.H. 2006. The NCI60 human tumour cell line anticancer drug screen. *Nature Reviews Cancer*. **6**(10),pp.813–823.
- Siegel, T. 2015. Journal of Clinical Neuroscience. *Journal of Clinical Neuroscience*. **22**(3),pp.437–444.
- Singh, S.K. et al. 2003. Identification of a cancer stem cell in human brain tumors. *Cancer research*. **63**(18),pp.5821–5828.
- Skog, J. et al. 2008. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature Cell Biology*. **10**(12),pp.1470–1476.
- Smits, M. et al. 2010. miR-101 is down-regulated in glioblastoma resulting in EZH2-induced proliferation, migration, and angiogenesis. *Oncotarget*. **1**(8),pp.710–720.

- Smyth, G.K. 2005. *Limma: linear models for microarray data*.
- Song, L. et al. 2012. TGF-beta induces miR-182 to sustain NF-kappaB activation in glioma subsets. *Journal of Clinical Investigation*. **122**(10),pp.3563–3578.
- Song, S. et al. 2013a. Biological function of nuclear receptor tyrosine kinase action. *Cold Spring Harbor perspectives in biology*. **5**(7).
- Song, S.J. et al. 2013b. MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell*. **154**(2),pp.311–324.
- Sorefan, K. et al. 2012. Reducing sequencing bias of small RNAs. *Silence*. **3**(1),p.4.
- Sorlie, T. et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*. **98**(19),pp.10869–10874.
- Sottoriva, A. et al. 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences*. **110**(10),pp.4009–4014.
- Srinivasan, S. et al. 2011. A ten-microRNA expression signature predicts survival in glioblastoma. *PLoS ONE*. **6**(3),p.e17438.
- Stahlhut, C. and Slack, F.J. 2013. MicroRNAs and the cancer phenotype: profiling, signatures and clinical implications. *Genome Medicine*. **5**(12),p.111.
- Stappert, L. et al. 2015. The role of microRNAs in human neural stem cells, neuronal differentiation and subtype specification. *Cell and tissue research*. **359**(1),pp.47-64.
- Stouffer, Samuel A., Edward A. Suchman, Leland C. DeVinney, Shirley A. Star, Robin M. Williams, Jr. 1949. Studies in Social Psychology in World War II: The American Soldier. Vol. 1, Adjustment During Army Life. *Princeton University Press*, **1949**, 125.
- Stupp, R. et al. 2009. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncology*. **10**(5),pp.459–466.
- Sturm, D. et al. 2012. Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer cell*. **22**(4),pp.425–437.
- Subramanian, M. et al. 2015. A biochemical approach to identify direct microRNA targets. *Methods in molecular biology*. **1206**,pp.29–37.
- Sugawa, N. et al. 1990. Identical splicing of aberrant epidermal growth factor receptor transcripts from amplified rearranged genes in human glioblastomas.

- Proceedings of the National Academy of Sciences*. **87**(21),pp.8602–8606.
- Sugiarto, S. et al. 2011. Asymmetry-Defective Oligodendrocyte Progenitors Are Glioma Precursors. *Cancer cell*. **20**(3),pp.328–340.
- Sukhdeo, K. et al. 2011. Glioma Development:Where Did It All Go Wrong? *Cell*. **146**(2),pp.187–188.
- Sulman, E.P. et al. 2013. Molecular predictors of outcome and response to bevacizumab (BEV) based on analysis of RTOG 0825, a phase III trial comparing chemoradiation (CRT) with and without BEV in patients with newly diagnosed glioblastoma (GBM). *Journal of clinical oncology*. **31**(suppl; abstr LBA2010).
- Sumazin, P. et al. 2011. An Extensive MicroRNA-Mediated Network of RNA-RNA Interactions Regulates Established Oncogenic Pathways in Glioblastoma. *Cell*. **147**(2),pp.370–381.
- Sun, J. et al. 2012. Uncovering MicroRNA and Transcription Factor Mediated Regulatory Networks in Glioblastoma. *PLoS Computational Biology*. **8**(7),p.e1002488.
- Sun, T. et al. 2013. Comparative evaluation of support vector machines for computer aided diagnosis of lung cancer in CT based on a multi-dimensional data set. *Computer Methods and Programs in Biomedicine*. **111**(2),pp.519–524.
- Svolos, P. et al. 2014. The role of diffusion and perfusion weighted imaging in the differential diagnosis of cerebral tumors: a review and future perspectives. *Cancer imaging*. **14**(1),p.20.
- Taal, W. et al. 2014. Single-agent bevacizumab or lomustine versus a combination of bevacizumab plus lomustine in patients with recurrent glioblastoma (BELOB trial): a randomised controlled phase 2 trial. *Lancet Oncology*. **15**(9),pp.943–953.
- Tabouret, E. et al. 2014. Association of matrix metalloproteinase 2 plasma level with response and survival in patients treated with bevacizumab for recurrent high-grade glioma. *Neuro-oncology*. **16**(3),pp.392–399.
- Takamizawa, J. et al. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer research*. **64**(11),pp.3753–3756.
- Tan, L. et al. 2013. Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Research*. **41**(7),pp.e84–e84.
- Tan, X. et al. 2012. The CREB-miR-9 Negative Feedback Minicircuitry Coordinates the Migration and Proliferation of Glioma Cells J. Li, ed. *PLoS ONE*. **7**(11),p.e49570.

- Tang, H. et al. 2013. The miR-183/96/182 cluster regulates oxidative apoptosis and sensitizes cells to chemotherapy in gliomas. *Current cancer drug targets*. **13**(2),pp.221–231.
- Tate, M.C. 2015. Surgery for gliomas. *Cancer treatment and research*. **163**,pp.31–47.
- 't Hoen, P.A.C. et al. 2013. Reproducibility of high-throughput mRNA and small RNA sequencing across laboratories. *Nature biotechnology*. **31**,pp.1015–1022.
- Tibshirani, R. 1996. Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B-Methodological*. **58**(1),pp.267–288.
- Tomasetti, C. and Vogelstein, B. 2013. Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. *Proceedings of the national academy of sciences*. **110**(6):1999-2004.
- Topalian, S.L. et al. 2012. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *The New England journal of medicine*. **366**(26),pp.2443–2454.
- Trang, P. et al. 2011. Systemic delivery of tumor suppressor microRNA mimics using a neutral lipid emulsion inhibits lung tumors in mice. *Molecular Therapy*. **19**(6),pp.1116–1122.
- Tu, J. et al. 2012. Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis. *BMC genomics*. **13**(1),p.43.
- Turcan, S. et al. 2012. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*. **483**(7390),pp.479–483.
- Ujifuku, K. et al. 2010. miR-195, miR-455-3p and miR-10a(*) are implicated in acquired temozolomide resistance in glioblastoma multiforme cells. *Cancer Letters*. **296**(2),pp.241–248.
- Vartanian, A. et al. 2014. GBM's multifaceted landscape: highlighting regional and microenvironmental heterogeneity. *Neuro-oncology*. **16**(9),pp.1167-75.
- Venere, M. et al. 2011. Cancer stem cells in gliomas: Identifying and understanding the apex cell in cancer's hierarchy. *Glia*. **59**(8),pp.1148–1154.
- Verhaak, R.G.W. et al. 2010. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer cell*. **17**(1),pp.98–110.
- Vickers, K.C. et al. 2011. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. *Nature Cell Biology*. **13**(4),pp.423–433.
- Vinson, C. and Chatterjee, R. 2012. CG methylation. *Epigenomics*. **4**(6),pp.655–663.
- Visani, M. et al. 2013. Definition of miRNAs Expression Profile in Glioblastoma

- Samples: The Relevance of Non-Neoplastic Brain Reference D. Monleon, ed. *PLoS ONE*. **8**(1),p.e55314.
- Visone, R. et al. 2009. Karyotype-specific microRNA signature in chronic lymphocytic leukemia. *Blood*. **114**(18),pp.3872–3879.
- Vlachos, I.S. et al. 2012. DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. *Nucleic Acids Research*. **40**(W1),pp.W498–W504.
- Volinia, S. et al. 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. *Proceedings of the National Academy of Sciences*. **103**(7),pp.2257–2261.
- Waggott, D. et al. 2012. NanoStringNorm: an extensible R package for the pre-processing of NanoString mRNA and miRNA data. *Bioinformatics*. **28**(11),pp.1546–1548.
- Waghmare, I. et al. 2014. Intercellular Cooperation and Competition in Brain Cancers: Lessons From Drosophila and Human Studies. *Stem cells translational medicine*. **3**(11),p.1262-8.
- Wagner, L. et al. 2013. Smac mimetic sensitizes glioblastoma cells to Temozolomide-induced apoptosis in a. *Oncogene*. **32**(8),pp.988–997.
- Wan, G. et al. 2013. DNA-Damage-Induced Nuclear Export of Precursor MicroRNAs Is Regulated by the ATM-AKT Pathway. *Cell Reports*. **3**(6),pp.2100–2112.
- Wang, C. et al. 2014a. The concordance between RNA-seq and microarray data depends on chemical treatment and transcript abundance. *Nature biotechnology*. **32**(9),pp.926–932.
- Wang, F. et al. 2014b. MiRNA-181c inhibits EGFR-signaling-dependent MMP9 activation via suppressing Akt phosphorylation in glioblastoma. *Tumour biology*. **35**(9),pp.8653–8658.
- Wang, H. et al. 2007. Direct and sensitive miRNA profiling from low-input total RNA. *RNA*. **13**(1),pp.151–159.
- Wang, K. et al. 2010. Export of microRNAs and microRNA-protective protein by mammalian cells. *Nucleic Acids Research*. **38**(20),pp.7248–7259.
- Wang, Yan et al. 2013. Pathway-based serum microRNA profiling and survival in patients with advanced stage non-small cell lung cancer. *Cancer research*. **73**(15),pp.4801–4809.
- Wang, et al. 2014c. p53 is positively regulated by miR-542-3p. *Cancer research*. **74**(12),pp.3218-27.
- Wang Y, et al. 2011a. MicroRNAs involved in the EGFR/PTEN/AKT pathway in gliomas. *Journal of neuro-oncology*. **106**(2),pp.217–224.

- Wang, Z.-X. et al. 2011b. Prognostic significance of serum miRNA-21 expression in human non-small cell lung cancer. *Journal of Surgical Oncology*. **104**(7),pp.847–851.
- Warren, J.J. et al. 2006. The structural basis for the mutagenicity of O(6)-methyl-guanine lesions. *Proceedings of the National Academy of Sciences*. **103**(52),pp.19701–19706.
- Weber, R.G. et al. 2001. Chromosomal imbalances associated with response to chemotherapy and cytotoxic cytokines in human malignant glioma cell lines. *International Journal of Cancer*. **91**(2),pp.213–218.
- Wiltshire, R.N. et al. 2000. Comparative genetic patterns of glioblastoma multiforme: potential diagnostic tool for tumor classification. *Neuro-oncology*. **2**(3),pp.164–173.
- Wu, D. et al. 2013. The use of miRNA microarrays for the analysis of cancer samples with global miRNA decrease. *RNA*. **19**(7),pp.876–888.
- Wu, Y. et al. 2011. A miR-200b/200c/429-binding site polymorphism in the 3' untranslated region of the AP-2alpha gene is associated with cisplatin resistance. *PLoS ONE*. **6**(12),p.e29043.
- Wurdak, H. et al. 2010. An RNAi Screen Identifies TRRAP as a Regulator of Brain Tumor-Initiating Cell Differentiation. *Stem Cell*. **6**(1),pp.37–47.
- Wynendaele, J. et al. 2010. An illegitimate microRNA target site within the 3' UTR of MDM4 affects ovarian cancer progression and chemosensitivity. *Cancer research*. **70**(23),pp.9641–9649.
- Yan, H. et al. 2009. IDH1 and IDH2 Mutations in Gliomas. *The New England journal of medicine*. **360**(8),pp.765–773.
- Yang, H. et al. 2012a. IDH1 and IDH2 Mutations in Tumorigenesis: Mechanistic Insights and Clinical Perspectives. *Clinical Cancer Research*. **18**(20),pp.5562–5571.
- Yang, W. et al. 2014. Knockdown of miR-210 decreases hypoxic glioma stem cells stemness and radioresistance. *Experimental Cell Research*. **326**(1),pp.22–35.
- Yang, Y.-P. et al. 2012b. Inhibition of cancer stem cell-like properties and reduced chemoradioresistance of glioblastoma using microRNA145 with cationic polyurethane-short branch PEI. *Biomaterials*. **33**(5),pp.1462–1476.
- Ye, F. et al. 2013. Protective properties of radio-chemoresistant glioblastoma stem cell clones are associated with metabolic adaptation to reduced glucose dependence. *PLoS ONE*. **8**(11),p.e80397.
- Yi, R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes & development*. **17**(24),pp.3011–3016.

- Yin, D. et al. 2013. miR-34a functions as a tumor suppressor modulating EGFR in glioblastoma multiforme. *Oncogene*. **32**(9),pp.1155–1163.
- Yu, S.-L. et al. 2008. MicroRNA Signature Predicts Survival and Relapse in Lung Cancer. *Cancer cell*. **13**(1),pp.48–57.
- Yuan, X. et al. 2004. Isolation of cancer stem cells from adult glioblastoma multiforme. *Oncogene*. **23**(58),pp.9392–9400.
- Yue, D. et al. 2009. Survey of Computational Algorithms for MicroRNA Target Prediction. *Current genomics*. **10**(7),pp.478–492.
- Zhang, C. et al. 2013. IDH1/2 mutations target a key hallmark of cancer by deregulating cellular metabolism in glioma. *Neuro-oncology*. **15**(9),pp.1114–1126.
- Zhang, C.-Z. et al. 2010. MiR-221 and miR-222 target PUMA to induce cell survival in glioblastoma. *Molecular cancer*. **9**,p.229.
- Zhang, W. et al. 2012a. Whole-genome microRNA expression profiling identifies a 5-microRNA signature as a prognostic biomarker in Chinese patients with primary glioblastoma multiforme. *Cancer*. **119**(4),pp.814–824.
- Zhang, Y. et al. 2014. An RNA-Sequencing Transcriptome and Splicing Database of Glia, Neurons, and Vascular Cells of the Cerebral Cortex. *Journal of Neuroscience*. **34**(36),pp.11929–11947.
- Zhang, Q., Kandic, I. & Kutryk, M.J., 2011. Dysregulation of angiogenesis-related microRNAs in endothelial progenitor cells from patients with coronary artery disease. *Biochemical and biophysical research communications*, **405**(1), pp.42–46.
- Zhao, Y. et al., 2014. miRNA-directed regulation of VEGF in tilapia under hypoxia condition. *Biochemical and biophysical research communications*, **454**(1), pp.183–188.
- Zhu, D.-X. et al. 2012a. Downregulated Dicer expression predicts poor prognosis in chronic lymphocytic leukemia. *Cancer science*. **103**(5),pp.875–881.
- Zhu, W. et al. 2012b. Genomic signatures characterize leukocyte infiltration in myositis muscles. *BMC Medical Genomics*. **5**,p.53.
- Zhu, S. et al. 2008. MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell research*. **18**(3),pp.350–359.
- Ziebarth, J.D. et al. 2012. Integrative analysis of somatic mutations altering microRNA targeting in cancer genomes. *PLoS ONE*. **7**(10),p.e47137.
- Zoni, E. et al. 2014. Epithelial Plasticity in Cancer: Unmasking a MicroRNA Network for TGF- β -, Notch-, and Wnt-Mediated EMT. *Journal of Oncology*. p.e.198967. [epub ahead of print].

