DECIPHERING THE POTENTIAL INTERPLAY BETWEEN MATRIX METALLOPROTEINASE-9 (MMP9) AND THE MUCIN 4 (MUC4) IN GLIOBLASTOMA MOLECULAR MECHANISMS AND DIAGNOSTICS

Agathe Quesnel

A thesis submitted in partial fulfilment of the requirements of **Teesside University** for the award of the degree of **Doctor of Philosophy (PhD)**

December 2022

Abstract

Despite recent advances, the prognosis of glioblastoma (GBM), the most common brain cancer, remains poor. Evidence suggests that the epidermal growth factor receptor (EGFR)/matrix metalloproteinase-9 (MMP9) axis is implicated in gliomagenesis. Recent studies have shed light on EGFR interaction with the highly-glycosylated transmembrane mucin 4 (MUC4) with implication in carcinogenesis. MUC4 expression in glioma grades and its link with EGFR/MMP9-driven glioma progression have not been studied so far. Changes in the glycosylation pattern through alterations of glycoproteins, such as mucins, in brain cancer represent potential biomarkers in glioma diagnosis and monitoring.

In this thesis, EGFR, MUC4, and MMP9 tissue expression was investigated to assess their utility as a panel of biomarkers. Secondly, this thesis aimed to define the molecular mechanisms underlying their role in GBM. Lastly, Raman spectroscopy (RS) with combined bioinformatic analyses were used to monitor changes in glycosylation for glioma diagnosis and monitoring.

Immuno-staining techniques and ELISA were used in fixed tissue samples and blood serum samples. MUC4 was overexpressed in glioma grades, and especially in GBM, with MMP9. This expression was found in neoplastic cells and in the microvasculature. The two proteins combined had prognostic value. *In-vitro* assays using cellular models was employed to define the molecular mechanisms of the pathway and suggested that MUC4 might be involved in cancer cell dissemination and may interact with EGFR in glioma cells. Gene expression analysis supported the involvement of the three proteins in a common pathway.

The utility of Raman spectroscopy combined with classification learning was lastly demonstrated on cellular models. This methodology showed a promising high sensitivity potential for the detection of subtle glycoprotein patterns in cell lines. Importantly, this approach was found useful in discriminating glioma grades with high accuracy in tissue and serum samples. Discrimination was possible owing to a significant change in glycosylation.

Declaration

I declare that the work presented in this thesis is my own work and has not been submitted for any other academic award or degree.
I declare that all the data acquisitions, analyses, interpretations, figures, and illustrations of this thesis are original and have been conducted during this PhD thesis.

Publications related to this PhD thesis:

Part of **Chapter 1** was published in **Biochemica and biophysica Acta – Reviews on cancer**:

A. Quesnel, G. S. Karagiannis and P. S. Filippou (2020). "Extracellular proteolysis in glioblastoma progression and therapeutics." Biochim Biophys Acta Rev Cancer 1874(2): 188428.
https://pubmed.ncbi.nlm.nih.gov/32956761/

and in **Cancer Metastasis Reviews**:

A. Quesnel, A. Broughton, G. S. Karagiannis and P. S. Filippou (2022). "Message in the bottle: regulation of the tumor microenvironment via exosome-driven proteolysis." Cancer Metastasis Rev. 1874(2):188428.
https://pubmed.ncbi.nlm.nih.gov/35394580/

Part of **Chapter 2** results was published in **Scientific Reports**:

A. Quesnel, N. Coles, T. M. Polvikoski, G. S. Karagiannis, C. Angione, M. Islam, A. A. Khundakar and P. S. Filippou (2022). "The diagnostic and prognostic potential of the EGFR/MUC4/MMP9 axis in glioma patients." Sci Rep 12(1): 19868.
https://pubmed.ncbi.nlm.nih.gov/36400876/

Part of **Chapter 4 and 5** results was published in **BMC Cancer**:

A. Quesnel, N. Coles, C. Angione, P. Dey, T.M. Polvikoski, T.F. Outeiro, M. Islam, A.A. Khundakar, P.S. Filippou (2023). “Glycosylation spectral signatures for glioma grade discrimination using Raman spectroscopy.” BMC Cancer 23(1): 174.
https://pubmed.ncbi.nlm.nih.gov/36809974/

Table of contents

List of figures

List of tables

Acknowledgements

I would like to first express my respect and thanks to Teesside University for giving me the opportunity to conduct this fully funded PhD project.
I express my sincere gratitude to my supervisory team for their kindness and expertise. Thank you to my director of study and primary supervisor, Dr Panagiota Filippou, for her tireless enthusiastic support and guidance, that was considerably helpful over these 3 years. I would like to thank the other members of my supervisory team, Dr Ahmad Khundakar and Prof. Meez Islam for their fantastic support and kind advice.
I would like to thank all the people who kindly collaborated to this work, Prof. Claudio Angione, Dr Priyanka Dey, Dr George S Karagiannis, Dr Tuomo Polvikoski, Prof. Tiago Outeiro, and Nathan Coles. I would like to thank my PhD committee examiners, Prof. John Young (NHC, TU) and Dr Triantafyllos Stylianopoulos (University of Cyprus) for their valuable time and constructive comments.
Special thanks to the Manchester Cancer Research Centre (MCRC), the NovoPath biobanks, and Newcastle Brain Tissue Resource, as well as the individuals who kindly donated their tissue and serum for the needs of this study.
Thank you to all the PhD students and staff members at the National Horizons Centre for their practical and emotional support.
Thank you to my former teachers at the University of Paris VII who supported my PhD application.
Thank you to my Newcastle housemates who enlightened my journey in the Northeast of England.
Thank you to the people who raised me, my mother Carole, and my father Laurent, Nathalie, Adèle, Rogette, Marcel, and Roland. All your love and care gave me the strength, confidence, and open-mindedness necessary to pursue a decade of academic studies.
Thank you to my siblings, Eléonore, Manon, and Kévin, who gave me a happy childhood and life.
Thank you to my beautiful and amazing close friends from my hometown, Audrey, Apolline, Farah, Sarah, Soraya, Célia, and Maïssem, for your unconditional support.

Abbreviations

ADAM: A Disintegrin and Metalloproteinase

ADAMTS: A Disintegrin and Metalloproteinase with Thrombospondin Motifs

AMOP: Adhesion-associated Domain

ANG-2: Angiopoietin 2

AREG: Amphiregulin

ATM: Ataxia Telangiectasia Mutated

CDD: Charge-coupled Device

CIMP: CpG Island Methylator Phenotype

DAB: 3,3'-Diaminobenzidine

EANO: European Association of Neuro-Oncology

ECM: Extracellular Matrix

EGFR: Epidermal Growth Factor Receptor

ELISA: Enzyme-linked Immunosorbent Assay

EMT: Epithelial-to-mesenchymal Transition

ERK: Extracellular Signal-Regulated Kinase

FFPE: Formalin-Fixed Paraffin-Embedded

GAPDH: Glyceraldehyde-3-Phosphate Dehydrogenase

Gal-3: Galectin-3

GalNT: Acetyl-galactosaminyltransferase

GBM: Glioblastoma

GFAP: Glial Fibrillary Acidic Protein

GlcNAc: N-acetylglucosamine

HGF: Hepatocyte Growth Factor

HIF: Hypoxia-inducible Factor

Hpx: Hemopexin

HRP: Horseradish Peroxidase

ICC: Immunocytochemistry

IDH: Isocitrate Dehydrogenase

IF: Immunofluorescence

IHC: Immunohistochemistry

IL: Interleukin

MDM2: Murine Double Minute 2

MGMT: O-6-Methylguanine-DNA Methyltransferase

MMP: Matrix Metalloproteinase

MRI: Magnetic Resonance Imaging

mTOR: Mechanistic Target of Rapamycin

MUC: Mucin

MVP: Microvascular Proliferation

NFκB: Nuclear Factor Kappa B

NIDO: Nidogen-like

NRG1: Neuregulin-1

OS: Overall Survival

PARP: Poly(ADP-ribose) Polymerase

PC: Principal Component

PCA: Principal Component Analysis

PD1: Programmed Cell Death Protein 1

PDGF: Platelet-derived Factor

PDGFR: Platelet-derived Factor Receptor

PFS: Progression-free Survival

PTEN: Phosphatase and Tensin Homolog

PTM: Post-translational modification

RBC: Red Blood Cell

RS: Raman Spectroscopy

SASP: Senescence-associated Secretory Phenotype

SEA: Sperm Urchin domain

SHH: Sonic Hedgehog

SNP: Single Nucleotide Polymorphism

STAT3/5: Signal Transducer and Activator of Transcription

SVM: Support Vector Machine

SVZ: Supraventricular Zone

TAA: Tumour-associated Antigen

TAM: Tumour-associated Macrophage

TERT: Telomerase Reverse Transcriptase

TGF: Tumour Growth Factor

TIMP1: Tissue Inhibitor of Matrix Metalloproteinase

TMD: Tetramethylbenzidine

TMQ: Thymoquinone

TMZ: Temozolomide

TP53: Tumour Protein 53

uPA: Urokinase Plasminogen Activator

uPAR: Urokinase Plasminogen Activator Receptor

VEGF: Vascular Endothelial Growth Factor

VWF: Van Willebrand Factor

WHO: World Health Organisation

WT: Wild-type

XRT: Radiotherapy

Chapter 1: Introduction

1.1 Glioblastoma (GBM)

1.1.1 Disease and epidemiology

Gliomas represent a category of tumours of the central nervous system arising in the non-neural compartment of the brain occupied by glial cells. Glial cells are divided into three main cellular populations: astrocytes, the most abundant of glial cells, which play a salient role in metabolic homeostasis in the brain; oligodendrocytes, which participate in axon myelination; and microglial cells, which are involved in pathogen phagocytosis. Although gliomas may arise from any of the above-mentioned glial cell populations, the vast majority is believed to arise from astrocytes and are thereby termed 'astrocytomas'. Gliomas arising from oligodendrocytes are termed 'oligodendrogliomas' and are the least frequent type of gliomas (Larjavaara *et al.* 2007). Astrocytomas can be further categorised into different grades according to the 2016 World Health Organization (WHO) classification, based on histological characteristics: the low-grades, grade I (Pilocytic) and II (Diffuse), and the high-grades, grade III (Anaplastic) and grade IV (Glioblastoma) (Wesseling 2018).

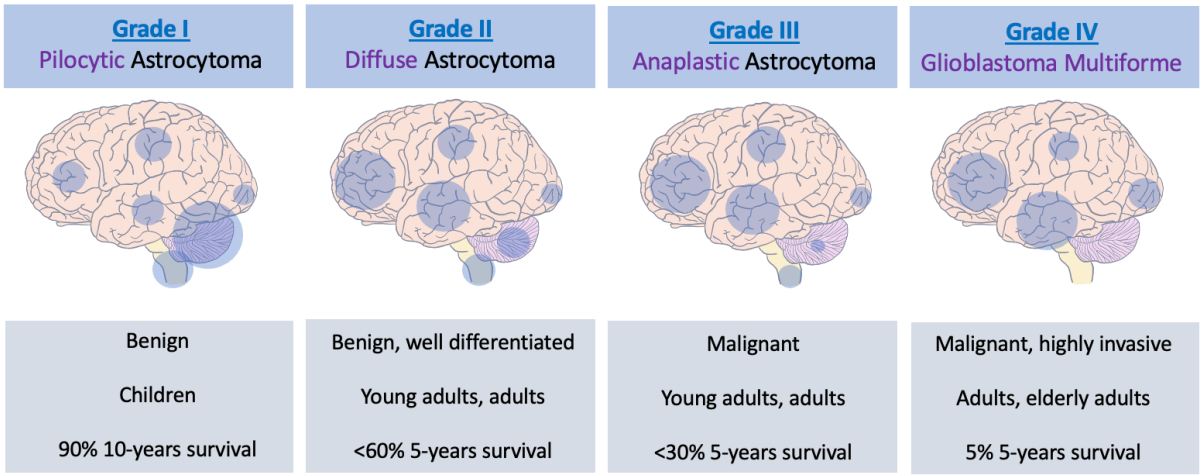


Figure 1.1 2016 WHO classification of astrocytoma. Incidence of the tumour by anatomic regions is represented for each grade (the size of the circle is proportional to incidence in each location). The degree of malignancy, the most affected category of age, and the survival rate are summarised for each grade. (Larjavaara *et al.* 2007, Tove *et al.* 2012, Collins *et al.* 2015, Davis 2016, Grimm *et al.* 2016, Wesseling 2018).

Low-grade astrocytomas encompass pilocytic (grade I) and diffuse (grade II) astrocytomas (**Fig.1.1**). Pilocytic astrocytoma is a benign tumour that typically arises in the cerebellum of young children and displays good prognosis (Collins *et al.* 2015) (**Fig.1.1**). Conversely, diffuse (Grade II) and anaplastic (Grade III) astrocytomas mainly originate from adult hemispheres and display much lower survival rates (Larjavaara *et al.* 2007, Tove *et al.* 2012) (**Fig.1.1**). Although qualified as 'low-grade', high mortality is still observed among grade II astrocytomas as it is characterised by a progressive growth that often leads to grade III. This type of tumour is referred to as 'diffuse' as it usually does not contain a solid tumour mass but instead comprises isolated single tumour cells. After malignant transformation, a solid mass can appear in this type of tumour (Pallud *et al.* 2011).

High-grade astrocytomas encompass anaplastic (grade III) and glioblastoma (grade IV). Anaplastic astrocytoma (grade III) is a malignant tumour with dedifferentiated cells, that displays poor prognosis and can progress towards glioblastoma (GBM) (Grimm *et al.* 2016). GBM represents the most advanced stage of gliomas and is both the most common and most lethal of all gliomas and brain tumours (Paw *et al.* 2015). GBM is diagnosed in older adult hemispheres and displays very poor prognosis, with a median survival of 15 months and a 5-year survival rate of 5% (Davis 2016, Dong *et al.* 2016) (**Fig.1.1**). This type of tumour is characterised by a solid central core surrounded by highly invasive cells. Most GBM cases are primary tumours, meaning they arise rapidly *de novo*, without pre-malignant lesions or evolution from lower-grades (Ohgaki *et al.* 2013). Secondary GBMs, which are associated with better survival, arise from a lower grade glioma (Grade II or III). Primary and secondary GBM exhibit the same histological features but have different genetic profiles and are believed to have different precursor cells of origin (Ohgaki *et al.* 2013).

1.1.2 Clinical features of GBM

GBMs are often diagnosed at late stage, when patients present severe symptoms associated to the tumour size and ensuing intra-cranial pressure. These symptoms include speech and motor deficits, seizures, and cognitive dysfunctions such as personality changes and memory deficits. The delayed diagnosis can be attributable to the non-specificity of the early symptoms (e.g., fatigue, headaches). GBM is diagnosed with the help of imaging techniques such as computerized tomography (CT)

or magnetic resonance imaging (MRI) (Davis 2016, Weller *et al.* 2021). However, histology characterisation of the tissue obtained after resection has been crucial so far for grading of gliomas (Taal *et al.* 2015). Histological hallmarks of GBM include cellular atypia and anaplasia (i.e., cellular abnormality and dedifferentiation), increased cell density, presence of a necrotic core, microvascular proliferation (MVP), and importantly, diffused widespread infiltration of isolated GBM cells into the otherwise healthy surrounding parenchyma (Shergalis *et al.* 2018). Of note, GBM cells rarely penetrate the blood circulation and therefore hardly ever metastasise to other parts of the body (Pasquier *et al.* 1980).

1.1.3 Molecular features of GBM

GBM displays wide morphological and molecular heterogeneity and is believed to derive from various cell types including mature glial cells and neurons, as well as neuron stem cells and progenitor populations (Friedmann-Morvinski *et al.* 2012). GBM stem cells and/or dedifferentiated astrocytes are the most commonly accepted cells of origin for glioma (Beiriger *et al.* 2022) . GBM can be classified into three main subtypes according to the molecular characteristics of the cells: classical, pro-neural, and mesenchymal.

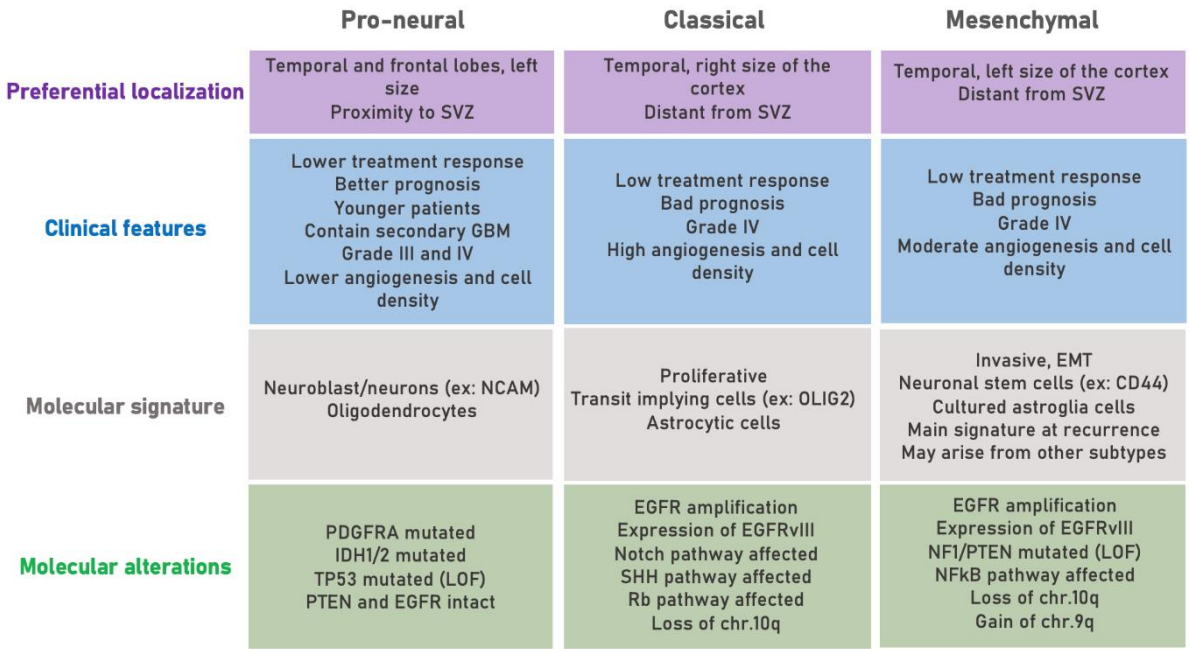


Figure 1.2 Molecular classification of high-grade glioma. Preferential localisation, clinical features, and molecular characteristics of pro-neural, classical, and mesenchymal subtypes. These subtypes, defined by their molecular signature, can each be analogised to a cell phenotype (neurons, proliferative, invasive cells), a neurogenesis state (neuronal, neuroblast, transit amplifying cells, neuronal stem cells) according to markers that they express, and a cellular type (oligodendrocytes, astrocytic cells, cultured astroglia cells). [Supraventricular zone (SVZ), Epithelial-to-mesenchymal transition (EMT), Chromosome (chr.)]. (Phillips *et al.* 2006, Rak *et al.* 2006, Verhaak *et al.* 2010, Friedmann-Morvinski *et al.* 2012, Steed *et al.* 2016, Rathore *et al.* 2018).

The classical subtype, which displays the most proliferative molecular signature, is characterised by amplification of EGFR, expression of EGFR variant III (EGFRvIII), and loss of chromosome 10q, which contains the *PTEN* gene (Verhaak *et al.* 2010). The pro-neural subtype, containing nearly all secondary GBMs, is generally identifiable by molecular alterations in PDGFRA, IDH1/2 (isocitrate dehydrogenase 1/2), and TP53 and is associated with better survival, younger age, and lower treatment response (Phillips *et al.* 2006). Finally, the mesenchymal subtype, which displays the most invasive signature, is characterised by amplification of EGFR, expression of angiogenic markers such as VEGF1, as well as NF1 and PTEN mutations (**Fig.1.2**) (Phillips *et al.* 2006). Overall, the critical oncogene and tumour-suppressor gene

involved in GBM development are respectively *EGFR* and *PTEN*, and mutations in these two genes have been shown to emerge during transition from anaplastic glioma towards GBM (Rak *et al.* 2006). EGFR and PTEN are intact in the pro-neural subtype, which principally encompasses anaplastic astrocytomas, whereas classical and mesenchymal subtypes contain mainly GBM and display loss of PTEN function (i.e., loss of the inhibition of the proliferative PI3K pathway) and amplification of EGFR.

Each subtype has a characteristic signature that can be compared to neurogenesis state and cellular type and assigned to a preferential anatomic localisation (**Fig.1.2**) (Steed *et al.* 2016, Rathore *et al.* 2018). 40% of GBM exhibit EGFR amplification inducing an elevated EGFR expression, which is often associated with additional expression of the constitutively activated variant form EGFRvIII. EGFRvIII, which is present in extracellular vesicles from the serum of GBM patients, has been shown to be a specific marker for glioma (Nishikawa *et al.* 1994). Other mutations found in GBM also include deletion of p16 and overexpression/amplification of MDM2 (the inhibitor of the tumour-suppressor p53), which are commonly found during late disease stages. Eventually, GBM cells carry many different mutations at recurrence, resulting from clonal and sub-clonal evolution, and therefore display high adaptation potential to chemotherapy (Kim *et al.* 2015).

Importantly, brain cancers have been reclassified during the course of this study (2021 WHO classification or WHO CNS5). The WHO classification system is no longer purely based on histological features alone but incorporates molecular and cytogenetic features. This classification update led to a change of guidelines by the European Association of Neuro-oncology (EANO) (Louis *et al.* 2021, Weller *et al.* 2021) summarised in **Figure 1.3**. Briefly, the new system states glioma is now classified in three basic groups based on the IDH genotype (*IDH1* or more rarely *IDH2*): mutant-IDH (m-IDH) astrocytoma, wild-type-IDH (wt-IDH) GBM, and m-IDH with 1p/19q co-deletion oligodendroglioma. The most common mutation for IDH is IDH1 R132H (more than 90% of cases) (Melhem *et al.* 2022). This means that m-IDH astrocytoma can no longer qualify as GBM, even when showing evidence of histological features of last grade glioma. Instead of diffuse, anaplastic, and GBM, these gliomas should be referred as grade 2, 3, or 4 m-IDH astrocytomas, using Arabic numbers (Louis *et al.* 2021). GBM is now defined by an astrocytic glioma with an absence of mutation in *IDH*

genes and histone *H3* genes in addition to presenting histological features such as necrosis and/or microvascular proliferation. Glioma that are wild type for *IDH* and *H3* genes but do not present histological features can still be classified as GBM if presenting specific molecular features: TERT promoter mutation, EGFR amplification, and/or loss of chromosome 10 with gain of chromosome 7 (+7/-10), since they are markers of poor prognosis. Consequently, almost all non-paediatric wt-IDH glioma now qualify as GBM, even in the absence of histologic hallmarks (**Fig.1.3**) (Weller *et al.* 2021). This represents an important change. However, as the histological classification was already very useful at predicting the IDH genotype, it is important to recognise that only a few GBM patients would likely be assigned to a different category following the new classification system.

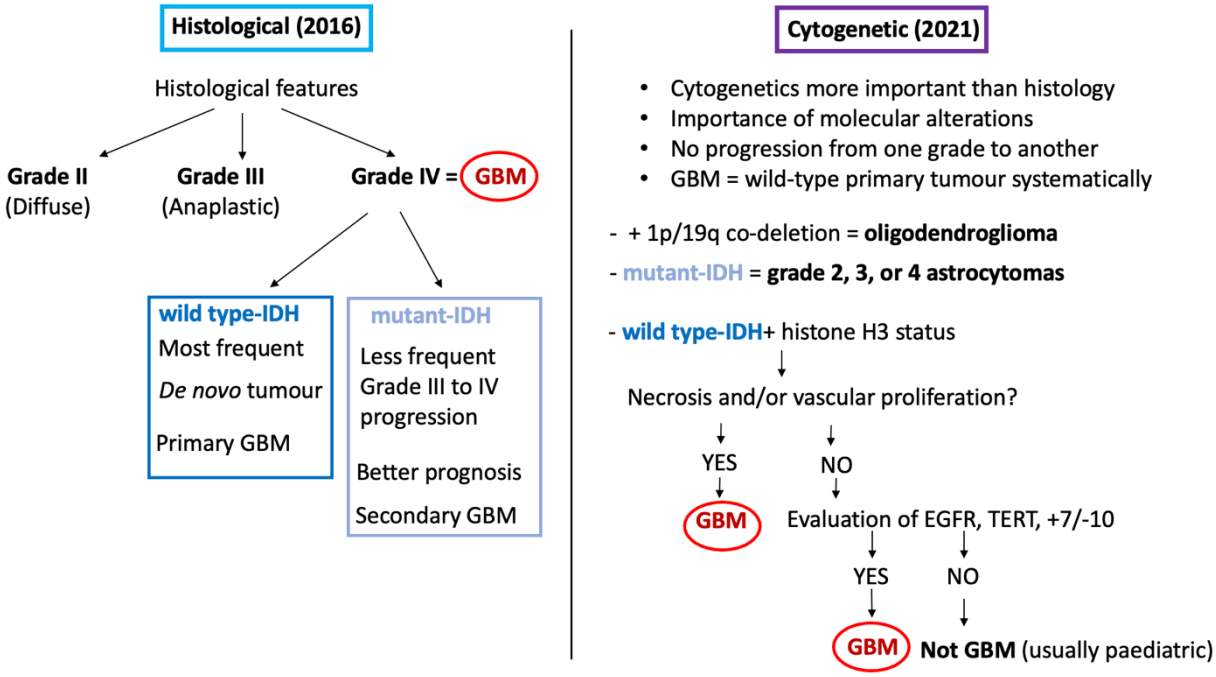


Figure 1.3 Comparison between WHO 2016 and 2021 classification of glioma. (Louis *et al.* 2021, Weller *et al.* 2021).

1.1.4 GBM treatments

The current conventional treatment for GBM consists of the combination between the common alkylating agent temozolomide (TMZ) and radiotherapy (or radiotherapy only), following surgical resection of the tumour (Davis 2016). Oligodendrogliomas and m-IDH diffuse astrocytomas are treated with procarbazine, lomustine, and vincristine

(PCV treatment) (Weller *et al.* 2021). The diffuse infiltration of isolated GBM cells into the surrounding tissue makes the complete surgical removal impossible, thus resection is rarely curative and recurrence at the resection site is almost systematically observed (Wick *et al.* 2011). The standard treatment (Stupp protocol), consisting of combined TMZ with radiotherapy after surgery, allows for prolonged progression-free survival (Stupp *et al.* 2005). Unfortunately, however, the treatment inexorably fails to stop GBM progression and maintenance of TMZ with radiotherapy do not prolong survival in wt-IDH astrocytomas (Weller *et al.* 2021). This treatment strategy presents several limitations since glioma cells that are left after surgery remain untreated for several weeks. These cells are already composed of different subclones when targeted by the treatment. In addition, surgery itself is believed to increase hypoxia. Therefore, it has been suggested that the Stupp protocol might contribute to early progression (Waqar *et al.* 2022). Maximal resection is still favoured as it extends survival and can be performed safely thanks to imaging techniques that are able to locate eloquent areas (Mukherjee *et al.* 2011, Taal *et al.* 2015). Another anti-angiogenic chemotherapeutic drug, the anti-VEGF bevacizumab, is also used in case of recurrence; however, this drug is not approved in Europe as it did not significantly improve overall survival in GBM patients (Hovey *et al.* 2017). Nitrosoureas treatment, such as lomustine, can also be given in case of recurrence (Taal *et al.* 2015). Both TMZ and lomustine function by adding an adduct to the O^6 position of guanine within the DNA, which triggers cell apoptosis (Taal *et al.* 2015).

In addition to TMZ and radiotherapy, a high number of targeted therapies have been investigated in GBM, although the molecular heterogeneity of GBM cells makes their efficacy very limited. For instance, therapies targeting proliferation pathways have been tested; those promoting superior improvement of overall survival during phase II clinical trials include the inhibitor of EGFR, erlotinib (Wick *et al.* 2011), the inhibitor of RAF, VEGFR, PDGFR, C-Kit, RAS, sorafenib (Hainsworth *et al.* 2010), and the inhibitor of mTOR, everolimus (Hainsworth *et al.* 2012). Other pathways have been targeted with limited efficacy in several studies. Targeting of integrin-modulated migration has also been tested with satisfactory results by using cilengitide (Stupp *et al.* 2010). Finally, studies using the drug rindopepimut, which targets EGFRvIII, led to limited improvement in overall survival (Babu *et al.* 2012). Although such treatments have shown promise, this has been undermined by limited efficacy and severe side

effects. There is therefore a need to develop more efficient and specific chemotherapies targeting new or currently investigated molecular targets, or alternative therapeutic strategies.

In addition to targeted therapy failure, GBM cells develop various resistance mechanisms against TMZ and radiotherapy. Chemotherapy resistance can be partly explained by the presence of GBM cell subpopulations exhibiting stem-like properties. These cells display proliferative and self-renewal properties and have been shown to develop high resistance after radiation (Haar *et al.* 2012). Additionally, cancer cells residing within the hypoxic centre of the tumour are believed to be protected from anti-proliferative chemotherapy as they are both quiescent and distant from the vasculature. Alternatively, cancer cells can resist death from chemotherapy via several molecular mechanisms, including expression of efflux proteins and DNA repair proteins (Haar *et al.* 2012). Indeed, TMZ cannot be given to patients with an active methylguanine-DNA methyltransferase (MGMT) enzyme (unmethylated MGMT promoter). MGMT is a DNA repair protein that specifically removes pro-mutagenic alkyl/methyl group from O^6 position of guanine, hence preventing mutations without causing DNA strand break. Resistance to radiotherapy has also been demonstrated within the context of GBM treatment. Radiotherapy was shown to promote transition from pro-neural to mesenchymal subtype (PMT), a process comparable to the epithelial-to-mesenchymal transition (EMT) observed in carcinoma (Nakano 2014). Following radiotherapy, an ensuing senescence-associated secretory phenotype (SASP) was also shown to induce the release of proteases and pro-inflammatory molecules, which in turn, led to increased cancer cell invasion (Jeon *et al.* 2016).

Finally, immunotherapeutic approaches have hold potential to treat GBM in recent years: 1) mRNA-pulsed dendritic cells, where dendritic cells are extracted from the peripheral blood of patients and re-injected with cancer-associated mRNAs for antigen presentation; 2) checkpoint inhibition, representing therapies targeting the inhibitory brakes of T-cells; 3) vaccine therapy, where dendritic cells are directly loaded with antigen and re-injected; and 4) modified autologous chimeric antigen receptor T-cells (CAR T-cells). However, these strategies present several limitations specific to brain tumours. Glioma and more specifically GBM, is known to have an immunosuppressive microenvironment coupled with a low immune surveillance in the brain and contains a

glioma stem cell niche with self-renewal capacity. For this reason, the most widely used checkpoint inhibition therapies anti-PD1 and anti-CTLA-4 have only showed modest results in glioma so far, although immunotherapeutic approaches using the whole tumour RNA have shown improvement of survival in glioma patients (Melnick et al. 2022; Vázquez Cervantes et al. 2021).

1.1.5 Biomarkers for GBM

The diagnosis of glioma is determinant for therapeutic decisions, yet, the grading highly depends on the observers, which can lead to subjectivity and ineffective therapeutic decision. Therefore, biomarkers that can help to guide diagnosis and choice of treatment in glioma are needed (Gravendeel *et al.* 2011). So far, several biomarkers with prognostic and diagnostic values have been identified in GMB. Those most investigated include loss of chromosome 10q heterozygosity, mutations in *IDH* genes, presence of the EGFR variant (EGFRvIII) (McNamara *et al.* 2013), and methylation status of MGMT and PTEN promoters (Riemenschneider *et al.* 2010).

Loss of heterozygosity of 10q chromosome is observed in 70% of primary GBM and is regarded as being crucial for both primary and secondary GBM formation (in the 2016 classification). Loci lost within the 10q chromosome frequently contains the *PTEN* gene (10q23 location), which is the phosphatase responsible for PI3K inactivation and one of the most potent and frequently mutated tumour-suppressor gene in cancer. The 10q loss is considered as a predictive biomarker for shorter survival in GBM patients (Ohgaki *et al.* 2004, McNamara *et al.* 2013).

IDH1/2 mutations are associated with the pro-neural subtype and secondary GBM as well as low-grade gliomas and are therefore an indicator of good outcome. They are rarely found in primary GBM; however, they also predict good prognosis when found in this type of tumour. GBM patients with mutated IDH1/2 have improved survival than those without (Riemenschneider *et al.* 2010, McNamara *et al.* 2013). The neomorphic mutation (leading to a new function) in IDH1 causes the conversion of the α-ketoglutarate cytosolic enzyme to the D-2-hydroxyglutarate enzyme. The neomorphic enzyme produced then leads to tumour initiation partly through an increase of histone and DNA methylation (hypermethylation), which is called ‘CpG islands hyper-methylator phenotype’ (CIMP) (Alzial *et al.* 2022, Rodríguez-Camacho *et al.* 2022).

Mutually exclusive mutation of the mitochondrial isozyme IDH2 is less common (Dang *et al.* 2009, Turcan *et al.* 2012, Cohen *et al.* 2013). The IDH genotype now plays a crucial role in the most recent WHO classification (2021) since it discriminates GBM (former wt-IDH GBM or primary GBM) from grade 2, 3, and 4 astrocytoma (former m-IDH glioma that could be classified in any grades depending on their histology with correspondence to secondary GBM in the case of grade IV) (Louis *et al.* 2021). In glioma patients, presence of the wild type version of IDH is correlated with increased infiltrating lymphocytes in the tumour and reduced anti-tumour T-cell response (Barthel *et al.* 2022).

Another type of biomarker in GBM is the presence of the constitutively activated EGFR variant EGFRvIII, which is observed in 20 to 30% of primary GBM. Identification of this variant in GBM patients is a predictor of poor prognosis (Riemenschneider *et al.* 2010, McNamara *et al.* 2013). EGFRvIII can be observed in microvesicles extracted from blood serum of some GBM patients and is thus a potential biomarker for liquid biopsies (Skog *et al.* 2008). In general, biomarkers that can be detected in liquid biopsies (blood serum, cerebrospinal fluid, urine) display several advantages compared to those detected in solid tissue. First, tissue biopsies may fail to be representative of the entire tumour. Besides, biomarkers from body fluids can be assessed several times as the method used to extract them is less invasive compared to solid biopsies; thus, they can serve in monitoring the progression of the disease. This type of biomarker has not been the most studied in GBM. Indeed, solid biopsies are almost always available in GBM, but furthermore in grade II and grade III gliomas, as maximal resection of the tumour is recommended as part of the standard treatment (Wager *et al.* 2011).

As mentioned above, methylation state of MGMT predicts the response of GBM to TMZ and is therefore used as a biomarker for treatment-response prediction (Riemenschneider *et al.* 2010). MGMT mRNA has been observed in extracellular vesicles of GBM patients and is believed to transfer resistance to TMZ between the cells (Simon *et al.* 2020). Therefore, monitoring extracellular vesicles in GBM patients, which can also be obtained from liquid biopsies, could be useful to evaluate and anticipate therapeutic response. Other genes whose promoter methylation state is found relevant in GBM include *PTEN*. Methylation of *PTEN* promoter is only observed in low-grade gliomas, in addition to secondary GBM that arise from them, and seems

to be mutually exclusive with mutations affecting the gene. Methylation of *PTEN* promoter, which leads to PTEN silencing, seems to be an alternative mechanism for aberrant activation of the proliferative PI3K signalling pathway in these tumours (Wiencke *et al.* 2007).

1.2 The invasive nature of GBM microenvironment

1.2.1 The extracellular matrix of GBM

In the brain, the normal extracellular matrix (ECM) is enriched in hyaluronic acid, tenascin, and lecticans (aggrecan, neurocan, versican, and brevican), whereas fibronectin and fibrillary collagen are relatively low and are exclusively found in vascular and perivascular areas (Bellail *et al.* 2004). The low level of fibrillary collagen and fibronectin confers a characteristic soft consistency to the brain. The blood-brain barrier is formed by endothelial cells surrounded by the basement membrane, with embedded pericytes. This capillary network is further surrounded by astrocytes that provide the cellular link to neuronal cells (Rempe *et al.* 2016). The ECM network in the brain is primarily produced by astrocytes and oligodendrocytes. Overall, the ECM density in GBM tends to be higher compared to the healthy brain, owing to both high GBM cell density and overexpression of ECM components (Mammoto *et al.* 2013). For instance, collagen is a rare component of the healthy brain microenvironment but is significantly increased during glioma development and progression (Dong *et al.* 2010). Interactions between GBM cells and the ECM components, such as collagen, play an important part in the invasion dynamics and GBM cell dissemination is facilitated in ECM-enriched regions (Seker-Polat *et al.* 2022). In brain tumours, collagen acts as a scaffold, providing adhesion sites for cell migration, and as a reservoir for ECM components, growth factors, and ligands for diverse signalling pathways. In particular, glioma cells produce their own collagen, particularly type IV collagen, instead of type I, which contributes to the high invasiveness observed in this tumour (Pointer *et al.* 2017).

During glioma progression, high dysregulation of the ECM composition is observed with glioma cells expressing their own pro-invasive matrix (Gritsenko *et al.* 2012). Alterations in the tumour microenvironment during GBM development and progression dramatically enhance the invasive properties of GBM cells (Akiyama *et al.* 2001,

Mahesparan *et al.* 2003, Sabari *et al.* 2011). Alterations in ECM composition is believed to support GBM cell migration, partly through modulation of their interaction between integrin receptors, present on the cell surface, and the ECM components. For instance, augmentation of $\alpha 2\beta 1$, which interacts with collagen, laminin, and tenascin, is observed in GBM, and seems to be involved in adhesion and migration abilities of GBM cells. Similarly, $\alpha v\beta 3$ is overexpressed with its ligand vitronectin at the invasive edge of the tumour as well as in new blood vessels and interaction between the two is believed to play a role in cancer invasion (Bellail *et al.* 2004).

1.2.2 GBM tumour microenvironment and mechanisms of progression

Histologically, the progression of anaplastic glioma towards GBM stems from a vascular occlusion occurring within the tumour, which induces severe hypoxia and necrosis. This process is sustained by concomitant downregulation of anticoagulants during tumour progression (Sawaya *et al.* 1995). Tumour cells migrate away from the hypoxic regions, subsequently creating a diffuse cellular wave from the central necrotic area towards the tumour edge (infiltrating/invasive edge). These cells, referred to as 'pseudopalisading' cells, have been observed and described only in GBM tumours. They are typically found around necrotic areas and they overexpress and secrete pro-angiogenic factors, such as VEGF, HIF-1, and IL-8 (Rong *et al.* 2006, Șovrea *et al.* 2022) (**Fig.1.4**). At the invasive edge, both pseudopalisading cells and tumour-associated macrophages (TAMs) co-exist, with the latter comprised of both resident microglia-derived cells and blood-borne macrophages, that secrete pro-invasive cytokines and growth factors. The newly formed vasculature is leaky and poorly developed with abnormal basement membrane composition, which produces an oedema (Cribaro *et al.* 2021). In GBM, both vasculogenesis (*de novo* vessel formation) and angiogenesis (resulting from the migration and proliferation of existing endothelial cells) occur as a consequence of hypoxia (Șovrea *et al.* 2022). They cause the multiplication of clusters of glomeruli-shaped vessels, a phenomenon called microvascular proliferation (MVP) or vascular hyperplasia, characteristic of GBM. The resulting microvasculature is often located near the infiltrative edge (**Fig.1.4**) (Șovrea *et al.* 2022)**.** Overall, the vessel network is greatly increased and disrupted in GBM, with open ends and virtually no basement membrane, but also aberrant receptor expression and vessel dilatation (Uyar 2022).

Tumour cells migrate in the white matter zone of the brain, through myelinated axon tracts and the perivascular space (Seker-Polat *et al.* 2022). Proteolytic enzymes secreted by tumour cells can cleave inhibitory molecules, present on the myelin surface, which may then serve as an adhesion surface for cancer cell migration (Gritsenko *et al.* 2012). During migration in the perivascular space, tumour cells are able to detach the astrocyte end-feet from vessels, contributing to the subsequent detachment of pericytes, and as such, the overall disruption of the blood-brain barrier (Watkins *et al.* 2014). In the GBM microenvironment, the communication of tumour cells with other cells or the ECM is also permitted by transfer of exosomes or extracellular vesicles. Extracellular vesicles carry a lot of different molecules and can thus be used as a reservoir of biomarkers, but they are also directly involved in GBM progression, principally via increasing angiogenesis. For instance, extracellular vesicles released from glioma stem cells transport miRNAs in addition to EGFRvIII to endothelial cells, where they regulate signalling pathways and increase production of VEGF. In addition, vesicles derived from hypoxic GBM cells contain VEGFA protein, which was shown to increase the blood-brain barrier permeability (Simon *et al.* 2020).

Among other molecules, various proteolytic enzymes including matrix metalloproteinases (MMPs), have been shown to play a critical role in GBM progression and tumour invasion (Quesnel *et al.* 2020). GBM and stromal cells interact within the tumour microenvironment via the expression and activity of a variety of proteolytic enzymes. These enzymes act from the first stage of cell invasion observed in progression from anaplastic glioma towards secondary GBM. During this process, cells anchor to the ECM via a panel of surface ligands including integrin receptors and the ECM is degraded by the proteolytic networks consequently supporting cell migration (**Fig.1.4**) (Friedl *et al.* 2008). Proteases, and especially MMPs, have also been shown to be either secreted directly or transported by extracellular vesicles within the tumour microenvironment. Alternatively, their expression can also be triggered indirectly by transcriptional factors transported through extracellular vesicles in cancer and stromal cells (Quesnel *et al.* 2022). The involvement of MMPs in GBM progression, and more particularly MMP9, will be further detailed in the next section.

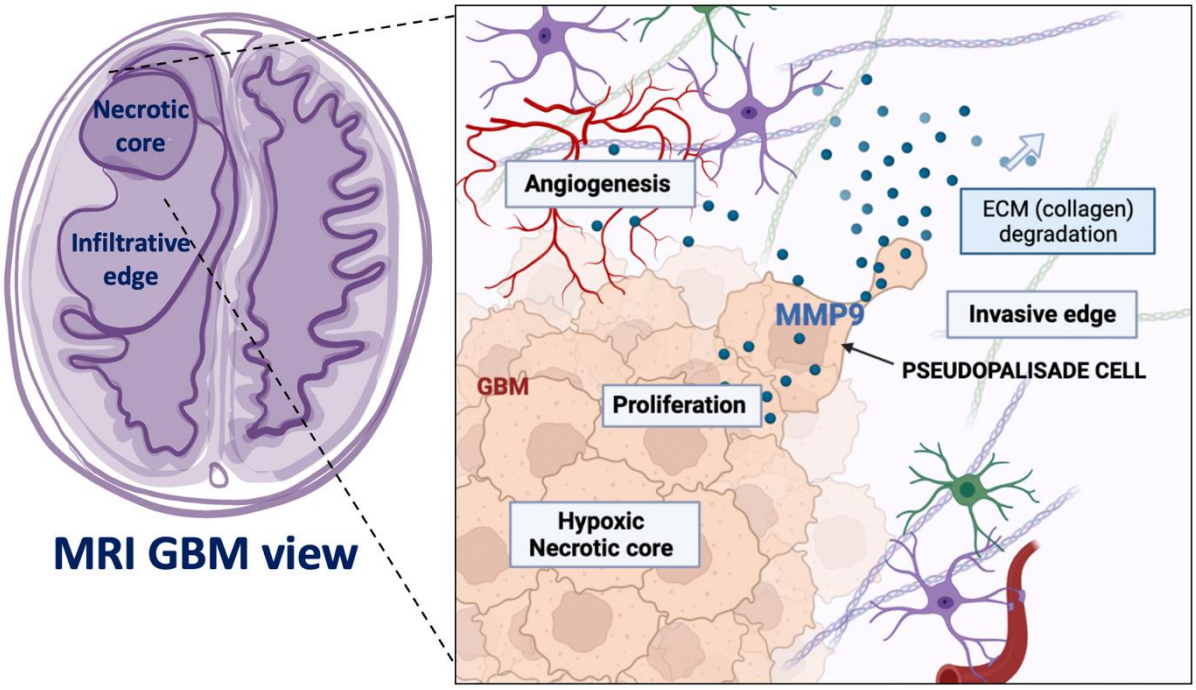


Figure 1.4 Mechanisms of invasion in GBM. GBM is composed of pseudopalisading cells constituting the infiltrative edge, surrounding the solid necrotic core. These cells have high invasive properties and detach from the core tumour to invade the surrounding tissue with the help of matrix metalloproteinases' (MMPs) secretion, and more particularly MMP9. The solid tumour core and the infiltrative edge can be discriminated with MRI images. [glioblastoma (GBM), extracellular matrix (ECM)]. *Figure created with Biorender.com*

1.3 Matrix metalloproteinases (MMPs) in GBM invasion and progression

1.3.1 The matrix metalloproteinase (MMP) family

Matrix metalloproteinases (MMPs) and adamalysins (ADAMs and ADAMTS) are protein subfamilies belonging to the broad metzincin family. The 26 human MMPs are highly homologous and can be categorised by their structural features and preferential substrates. They share three common domains: the pro-domain, the catalytic domain, and the hemopexin-like C-terminal (Hpx) domain linked to the catalytic site (**Fig.1.5**). MMPs can be either anchored to the cell surface or secreted. The pro-MMP (or zymogen) inactive protein can become active after the proteolytic removal of the pro-domain in the intracellular and extracellular space (Gialeli *et al.* 2011). Secreted MMPs

can be further classified according to their preferential ECM substrates. For instance, interstitial collagenases (MMP1, 8, and 13) mainly target fibrillar collagens (type I, II and III), whereas gelatinases (MMP2 and 9), also called type IV collagenases, mainly target basement membrane collagens, gelatine, and elastin. Stromelysins (MMP3, 10, and 11) are unable to degrade collagens, but instead target several ECM proteins, such as proteoglycans, fibronectins, and laminins. Finally, matrilysins (MMP7 and 26) target similar substrates as stromelysins, but lack the Hpx domain (Johansson *et al.* 2000) (**Fig.1.5**).

MMPs are also able to cleave inactivated forms of other MMPs (proMMPs), as well as pro-forms of growth factors, resulting in their subsequent activation. MMPs localised to the cell surface (MMP14, 15, 16, 17, 24, 25) are activated during their trafficking to the cell membrane (Johansson *et al.* 2000). In the extracellular space, MMPs are susceptible to inhibition by tissue inhibitors of metalloproteinases (TIMPs). TIMPs can bind to the active site of proteases to form a complex that will be recognised by macrophages (Visse *et al.* 2003). MMPs are further activated when secreted with only a low amount of their respective inhibitors; therefore, cell compartmentalisation is critical for the tight regulation of their activity. MMPs function in a variety of physiological processes that require a specific ECM (re)arrangement, as well as in multiple pathological processes, such as inflammation, autoimmunity, and cancer. In the brain, they significantly contribute to tissue formation, neuronal network remodelling, and blood-brain barrier integrity. Importantly, MMPs are involved in most brain diseases with a neuro-inflammatory component (e.g., Alzheimer's and Parkinson's diseases) and in brain cancers including glioma (Rempe *et al.* 2016). MMPs implicated in cancer progression are mainly expressed in the tumour cells, but also in the tumour-associated stromal cells (Tang *et al.* 2004).

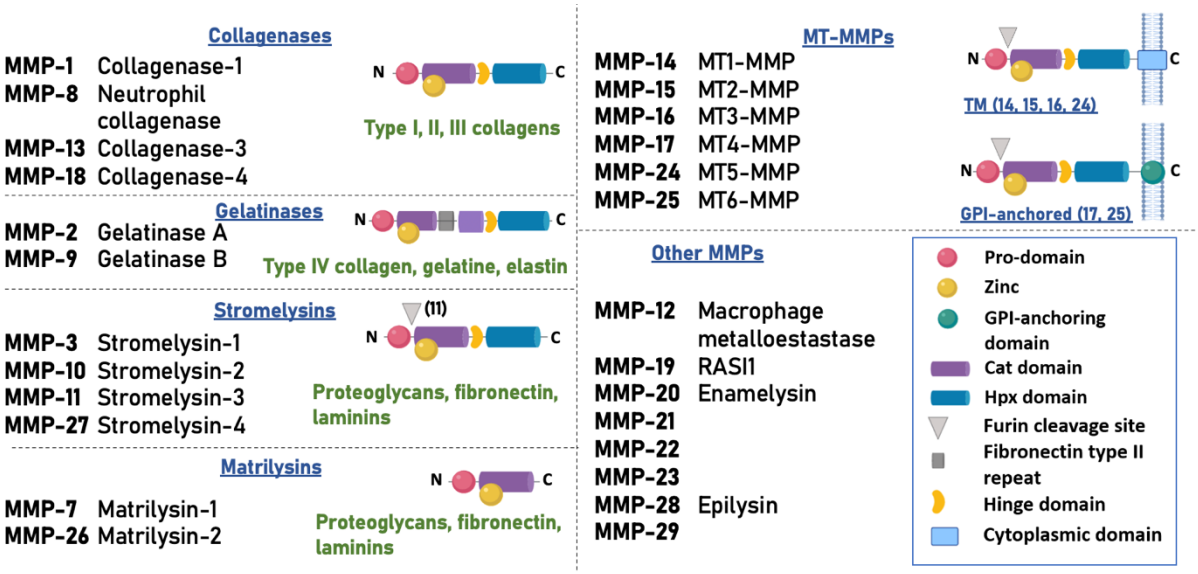


Figure 1.5 Classification, structure, and nomenclature of the MMP family. Preferential substrates are represented in green for each subtype [Matrix metalloproteinase (MMP), Membrane-type MMP (MT-MMP), Glycosylphosphatidylinositol (GPI), Catalytic (Cat), Hemopexin (Hpx), Transmembrane (TM)]. Adapted from (Quesnel *et al.* 2020).

1.3.2 Involvement of MMPs in GBM progression

Both ECM remodelling and degradation mediated by MMP-induced proteolytic cleavage play a prevailing role in cancer cell invasion. MMP14 disrupts fibrillar collagen, therefore inducing its rearrangement in the pericellular space and contributing to cancer cell invasion (Friedl *et al.* 2008). In general, there is evidence demonstrating that MMPs, along with ADAM proteins, contribute to cancer progression and are part of podosome-related proteases. In actin-rich podosomes of migrating cells, the scaffold protein Tks5/FISH binds to ADAMs where MMP2, 9 and 14 are present and promote invasion via direct ECM degradation (Sahin *et al.* 2004).

In glioma, the expression of multiple MMPs has been shown to be positively correlated with glioma histological grades (Hagemann *et al.* 2012), while 11 MMPs have been shown to be significantly overexpressed in GBM (Pullen, 2018, Quesnel *et al.* 2020) (**Table 1.1**). Apart from MMP15 and 19, the overexpression of these MMPs is correlated with poor survival. It is worth noting that the expression of furin, an enzyme that activates MMPs, is strongly associated with poor overall survival in GBM

(Shergalis *et al.* 2018), with the furin inhibition in astrocytoma cells leading to a reduction in cell migration (Mercapide *et al.* 2002). In general, more than half of human MMPs are correlated with GBM progression, with MMP2 and 9, being the most significant ones.

Proteases	Overexpression	Involvement in progression	Diagnostic value	Prognostic value
MMP1, 2, 7, 9, 11, 14	yes	yes	yes	yes
MMP3, 13, 16, 26	no	yes	no	yes
MMP8, 10, 23	yes	no	no	yes
MMP12	no	yes	yes	yes
MMP20, 21, 24	no	no	no	yes
MMP15, 19	yes	no	yes	no

Table 1.1 Overexpression, involvement in progression, and diagnostic/prognostic value of MMPs in GBM. Involvement in progression: the protein has been demonstrated to have a role in disease progression (invasion, proliferation, angiogenesis) from experimental studies. Diagnostic value: the protein has been shown to have the potential of discriminating between the different grades of glioma and to be differentially expressed in GBM compared to lower grades. Prognostic value: the expression of the protein has been demonstrated to be associated with poor survival. Adapted from (Quesnel *et al.* 2020).

1.3.3 Involvement of MMP9 in GBM progression

MMP9, among all MMPs, has been shown to be considerably involved in GBM progression (Quesnel *et al.* 2020). IHC analysis showed that its active form was not found in healthy brain, but in half of GBM tissue samples (Choe *et al.* 2002). The expression of MMP9 also correlates significantly with high-grade glioma and is considered as a potential significant prognostic factor (Xue *et al.* 2017). Low MMP9 expression is associated with favourable outcome and good response to TMZ treatment (Li *et al.* 2016). Overexpression of MMP9 triggers proliferation in GBM cell lines, while its blockade decreases the volume, weight, and microvascular density of

the tumour in mice (Sun *et al.* 2013). Importantly, MMP9 expression seems to be involved in resistance to treatment as its expression increases after radiation in recurrent GBM (Zhou *et al.* 2019).

Several mechanisms of regulation have been suggested for MMP9 expression and activation in GBM proliferative foci (**Fig.1.6A**), including the uPA/uPAR-mediated pathway. Urokinase plasminogen activator (uPA) is a serine protease, upregulated in high-grade gliomas (Sawaya *et al.* 1995) that converts plasminogen into plasmin, with better efficacy when anchored to its receptor (uPAR). Plasmin, in turn, is responsible for both MMPs and uPA activation (Shergalis *et al.* 2018). uPA is also capable of directly activating MMP9, which in turn can degrade fibronectin and has therefore been suggested to participate in GBM progression (**Fig.1.6B**) (Yamamoto *et al.* 1994, Zhao *et al.* 2008). MMP9 has been shown to be part of several intracellular signalling pathways. In GBM, MMP9 expression positively correlates with EGF and EGFR expression. MMP9 transcription is both stimulated by EGFR and EGFRvIII in GBM, with signalling pathways involved such as PI3K/AKT (PKB), STAT3/5, NFk-B, ERK, and Sonic hedgehog (SHH) (**Fig.1.6B**) (Choe *et al.* 2002, Chang *et al.* 2015, Chen *et al.* 2017).

MMP9 promotes cell migration via various mechanisms in the invasive edge (**Fig.1.6A**). First, via its ability to degrade several ECM components, preferentially type IV collagen, gelatine, and elastin; but also, via cleavage of cell-cell adhesion molecules and release of growth factors bound to the ECM (**Fig.1.6C**) (Huang 2018). Additionally, MMP9 can act on cell motility via direct cleavage of CD44. CD44 is a surface glycoprotein involved in cell-ECM interaction via binding to several ECM ligands, mainly hyaluronic acid. The direct shedding of CD44 by MMP9 releases its extracellular portion in the extracellular space and contributes to cell migration and invasion (**Fig.1.6C**) (Chetty *et al.* 2012). In addition, an *in vitro* model of GBM carrying PTEN mutation showed that deficiency of PTEN phosphatase activity leads to hyaluronic-acid-induced MMP9 expression (Park *et al.* 2002).

Finally, MMP9 has been shown to function as a mediator of angiogenesis. In GBM, MMP9 is expressed in proliferative endothelial cells (Rao *et al.* 1996) and induces

basement membrane degradation, enabling the migration of endothelial cells for neo-vasculature formation (**Fig.1.6D**). This is further supported by the fact that MMP2, 9, and 14 expression correlates with VEGF expression in GBM, which is the most crucial regulator of angiogenesis. Proliferation and migration of endothelial cells induced via VEGF in GBM is dependent on the binding of VEGF to VEGFR-2 (Munaut *et al.* 2003). MMP9 is induced by the hypoxic factor HIF-1 and increases VEGF action by cleaving its bound to the ECM and therefore acting on its bioavailability within the extracellular space (Du *et al.* 2008). MMP9 is also expressed in cells that correspond to vascular smooth muscle cells and pericytes and contributes to their proliferation (**Fig.1.6D**) (Forsyth *et al.* 1999).

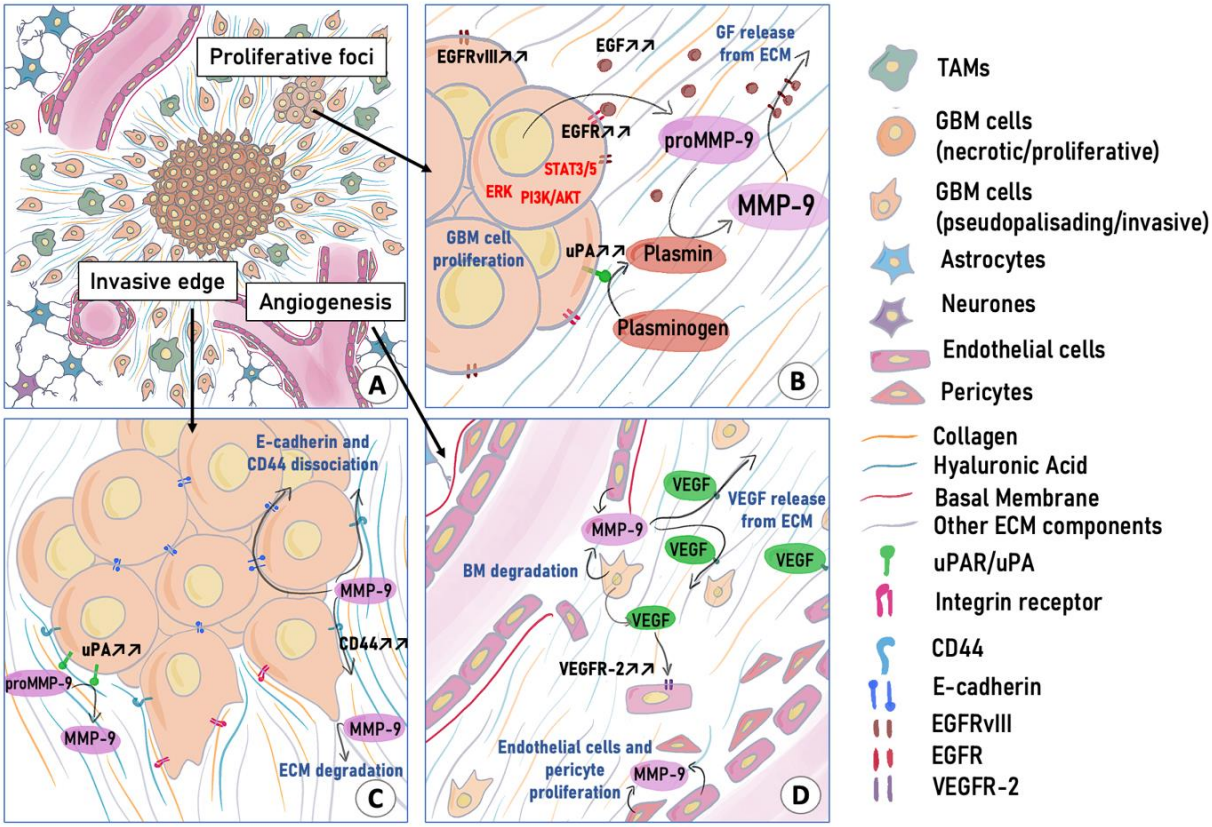


Figure 1.6 MMP9-mediated mechanisms of GBM progression. (A) GBM and its microenvironment, with the central necrotic tumour and the invasive edge with cells migrating away from the hypoxic centre (pseudopalisading cells); foci of proliferative GBM cells; and angiogenesis areas with vascular proliferation and new vessel formation. **(B)** Proliferative foci of GBM cells, pro-MMP9 is converted into active MMP9 by plasmin. Plasmin is converted from plasminogen by uPA/uPAR, whose expression increases in GBM. EGFR, EGFRvIII, and EGF are increased in GBM and lead to MMP9 expression via downstream proliferative pathways. MMP9 controls bioavailability of growth factors (GF) bound to the ECM. **(C)** Infiltrative edge of GBM, MMP9 is directly converted into its active form by uPA and leads to CD44 and E-cadherin dissociation as well as ECM degradation, promoting cell migration away from the edge. **(D)** Perivascular area of the tumour, MMP9 is expressed in endothelial cells and invasive cells and leads to VEGF release from the ECM; and basement membrane (BM) degradation, which leads to endothelial cell migration. VEGFR-2 expression is upregulated in GBM and promotes angiogenesis via VEGF binding. MMP9 is expressed in endothelial cells and pericytes and promotes their proliferation. Adapted from (Quesnel *et al.* 2020).

1.3.4 Therapies targeting MMPs in GBM

Small molecule MMP inhibitors have been previously developed for impeding proteolysis-driven cancer progression. The mechanisms of their action are divided into different categories. The peptidomimetic inhibitors (i.e., batimastat and marimastat) mimic the structure of collagen at the MMP cleavage site and, as such, act as competitive inhibitors. On the other hand, non-peptidomimetic inhibitors (i.e., tanomastat and prinomastat) mimic the tertiary conformation of the MMP active site (Gialeli *et al.* 2011). In addition, chemically modified tetracyclines have been used to inhibit MMP activity via zinc binding. Other specific molecules developed to inhibit specific MMPs (MMP2 and 9) have also been suggested for therapies, such as CGS-27023A and SB-3CT (Gialeli *et al.* 2011).

Two drugs that function as MMPs inhibitors, marimastat and prinomastat have both revealed promise in treating GBM. Marimastat is a broad spectrum metzincin inhibitor, which ultimately was shown to have limited impact on progression-free and overall survival. In a recent study, the use of marimastat in combination with TMZ and radiation resulted in downregulation of MMP14 via the parallel downregulation of microRNA374 and induction of cell cycle arrest. Of note, marimastat specifically inhibits the growth of cancer cells but not that of normal astrocytes (Ulasov *et al.* 2013). However, in Phase II trial, the use of marimastat with TMZ in anaplastic glioma not only showed very limited improvement of chemotherapy efficacy compared to standard treatment, but also generated additional joint toxicity (Groves *et al.* 2006). Prinomastat, on the other hand, targets MMP2, 3, 9, 13, and 14, which all together highly contribute to angiogenesis and invasion observed in GBM as mentioned above. Prinomastat-treated gliomas in mice are smaller in tumour size, have lower rates of proliferation, and are less invasive compared to untreated gliomas (Price *et al.* 1999). This drug has been associated with high toxicity and lack of efficacy during Phase III clinical trials (Bissett *et al.* 2005); hence, its potential applicability is currently on hold.

To sum up, small molecule MMP inhibitors have been previously developed, and some of them have been tested in GBM patients. Nonetheless, these treatments so far have shown limited efficacy in clinical trials for improving the patient outcome. This could be partly due to the tumour suppressive roles that MMPs have been shown to display during specific time-windows of tumour progression. For instance, MMP9 is

responsible for a rapid increase of angiogenesis at early stage of tumour progression during which it degrades basement membrane components. However, the resulting products of basement membrane cleavage by MMP9 can have an anti-angiogenesis effect. This has been demonstrated for tumstatin, a cleavage product of MMP9, which hampers tumour development during the later stage of progression (Hamano *et al.* 2003). However, it is unclear if these MMPs also have tumour-suppressive effects in GBM, but lessons should be learned from their application in other types of cancer, and careful considerations should be made for avoiding similar drawbacks in the GBM therapeutics field (Quesnel *et al.* 2020).

1.4 EGFR family of receptors

1.4.1 EGFR family and cancer

The epidermal growth factor receptor (EGFR) family is composed of four different receptors: HER1 to HER4, with HER1 being simply known as EGFR. These receptors interact with a variety of growth factor ligands present in the extracellular space and transduce signals inside the cells that provoke a wide array of cellular mechanisms including migration, growth, adhesion, and differentiation. Given the triggered cellular mechanisms, these receptors have been shown to be commonly dysregulated during cancer initiation, progression, and chemoresistance (von Achenbach *et al.* 2018). Except for HER3, EGF receptors contain an intrinsic tyrosine kinase activity in their cytoplasmic tail. However, HER3 and its neuregulin ligands can still be responsible for signal transduction within the cell, when coupled with HER2 (HER2 has no affinity for any ligand). The ligands able to act on EGF receptors include tumour growth factor α (TGFα), epithelial growth factor (EGF), and amphiregulin (AREG), which have affinity for EGFR (HER1); epiregulin (EREG), B-cellulin (BTC), and heparin binding (HB)-EGF, which have affinity for heterodimers composed of EGFR and HER4; and the neuregulins (NRG1 to 4), which bind to either HER3 or HER4 (von Achenbach *et al.* 2018).

1.4.2 EGFR alterations in GBM

EGFR is the most commonly amplified gene in GBM and is overexpressed in some 10% of secondary GBM and 60% of primary GBM. This amplification is associated with worsen prognosis and overall survival in GBM patients (Karpel-Massler *et al.* 2009).

Besides, EGFR give rise to different aberrant transcript variants, the most common one being the constitutively activated form EGFRvIII, which lacks the extracellular ligand-binding domain (von Achenbach *et al.* 2018). This variant is found in half of GBM carrying EGFR amplification and the mutation leading to this variant represents 60% of all mutations found in EGFR gene in GBM (Karpel-Massler *et al.* 2009). Other EGFR variants are less common and less characterised. Like EGFR, HER2 displays a very heterogenous expression in GBM samples but it has been estimated that this receptor is mutated in 8% of GBM (Wager *et al.* 2011). HER3 and HER4 have an extremely low expression but HER3 might be specifically expressed in GBM stem cells (Duhem-Tonnelle *et al.* 2010). Besides, EGFR and HER3, as well as their respective phosphorylated forms, are detected in GBM LN-18 cells whereas, high EGFR and HERs expression are detected in GBM A-172 cells. HER3 is expressed by both cell lines at mRNA level but to a lesser extent than EGFR and HER2, and HER4 is only expressed at low level in A-172 cells (von Achenbach *et al.* 2018).

1.4.3 Therapies targeting EGFR in GBM

EGFR can be targeted by the tyrosine kinase inhibitors gefitinib and erlotinib, with the latter inhibiting also the EGFRvIII variant. These treatments compete with ATP binding within the intracellular tyrosine kinase domain. However, these chemotherapies have failed in improving survival compared with standard treatments and thus are not traditionally used to treat GBM. Considering the implications of EGFR in GBM progression, it is surprising that targeting this pathway is not efficient and the underlying reasons are not yet well understood. However, it seems that the existence of alternative pathways as well as subpopulations of cells within the tumour that display different combinations of EGFR variants may account for rapid adaptation and resistance to tyrosine kinase inhibitors (Westphal *et al.* 2017). Overexpression of EGFR is not sufficient to predict response to tyrosine kinase inhibitors and other molecular alterations must be taken into considerations. For instance, loss of PTEN seems to be involved in resistance mechanisms to these treatments (Karpel-Massler *et al.* 2009).

1.5 Mucins

1.5.1 Family of mucins

Mucins are heavily O-glycosylated proteins expressed at the cell surface of healthy epithelial cells. Mucins can be either secreted or membrane-anchored (transmembrane mucins). Secreted mucins form a mucous gel in the extracellular space and serve as a protective barrier for epithelial cells. They are essentially secreted by goblet cells within the healthy epithelial layer (Kufe 2009). Transmembrane mucins include MUC1, MUC3, MUC4, MUC13, MUC14, MUC15, MUC16, MUC17, MUC20, MUC21, and MUC22 (Dhanisha *et al.* 2018). MUC1 is the best-characterised mucin and is present at the apical membrane of secretory epithelial cells in normal physiological conditions. Mucins comprise a domain rich in proline, threonine, and serine residues, called the 'PTS domain', which constitutes tandem repeats for O-glycosylation (Kufe 2009). The main physiological function of transmembrane mucins is to lubricate the epithelial layer for protection against pathogen entrance (Chaturvedi *et al.* 2008). Besides, the large, glycosylated PTS domain is thought to protect the proteins against proteolytic cleavage by pathogens and host proteases (van Putten *et al.* 2017).

Transmembrane mucins contain a sea urchin sperm protein (SEA) domain, unique to the mucin family. Transmembrane mucins are typically composed of the PTS domain, two EGF-like sequences, the SEA domain, the transmembrane domain, and the cytoplasmic tail. After that the addition of O-glycans occurred, it accounts for 80% of the total weight of the mature mucin (Jonckheere *et al.* 2013). The most commonly O-linked sugars are N-acetyl-glucosamine (GlcNAc) and N-acetyl-galactosamine (GalNAc). O-glycosylation is initiated by GalNAc transferase (GalNTs) on serine and threonine residues to constitute the Tn-epitope during trafficking through the Golgi apparatus. Subsequently, serial addition of galactose and GlcNAc occurs and forms the primary glycan structure. The O-glycan chains can then be further extended by other transferases with glucose, fucose, and sialic acid (N-acetyl-neuraminic acid) and can reach over 100 nm (Hanson *et al.* 2016, Reily *et al.* 2019), while addition of mannose and sulphate is more occasional. Glycosylation varies between cell type since transferases are not expressed at the same level in all the cells. The highly glycosylated extracellular part of transmembrane mucins can be released in the extracellular space and be detected in several fluids. The cytoplasmic domain, and sometimes the entire protein, have been observed in the nucleus. The extracellular

part of transmembrane mucins can be released either via shedding at the SEA domain or via cleavage by MMP14. After cleavage of the extracellular part, the cytoplasmic domain can be cleaved by a γ-secretase within the cell. However, it is not known yet if phosphorylation of the cytoplasmic tail and signal transduction occur after release of the extracellular part (van Putten *et al.* 2017, Dhanisha *et al.* 2018).

1.5.2 Mucin 4 (MUC4)

Mucin 4 (MUC4) has a unique structure compared to the other transmembrane mucins. The extracellular region of MUC4 embraces a Nidogen-like (NIDO) domain, an adhesion-associated (AMOP) domain, a Von Willebrand factor (VWF) domain and, between this latter domain and the transmembrane domain, three EGF-like sequences (**Fig.1.7**)**.** The VWF, AMOP, and NIDO domains replace together the SEA domain of the other transmembrane mucins (Kufe 2009). The EGF-like sequences are highly conserved, which suggests important function. This sequence is believed to act as a ligand for the EGFR family members. Surrounding the EGF-like sequences*,* N-glycan chains are added in the endoplasmic reticulum on asparagine residues (Jonckheere *et al.* 2013, Dhanisha *et al.* 2018). The different domains of MUC4 and their functions have been investigated and showed that: the NIDO domain could play a role in invasion and extravasation in carcinomas (Senapati *et al.* 2012); the AMOP domain is present in other cell adhesion proteins and as such, could have adhesion function; the function of the VWF is lost in MUC4, and finally; the cytoplasmic tail may have a function in signal transduction (Jonckheere *et al.* 2013). MUC4 is translated into a single polypeptide divided at the GDPH site (Gly-Asp-Pro-His) into two subunits: the α subunit covering the entire extracellular part, and the β subunit covering the transmembrane and cytoplasmic domains (**Fig.1.7**) (Dhanisha *et al.* 2018). It is believed that MUC4 can be cleaved at the GDPH site and be secreted in order to strengthen the extracellular mucous gel in some conditions (Xia *et al.* 2017).

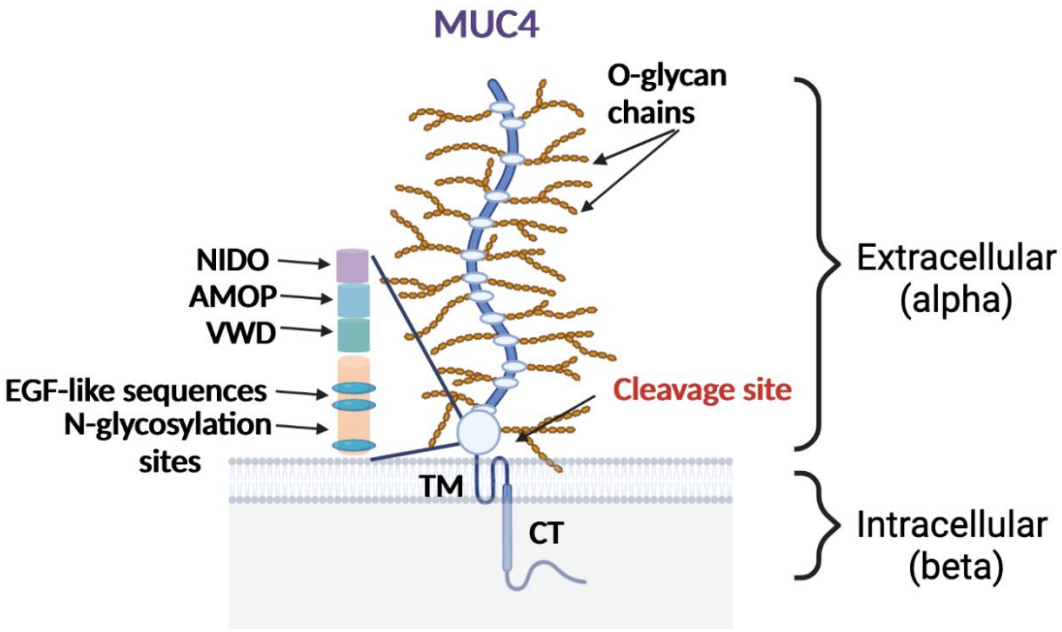


Figure 1.7 Structure of MUC4 protein. MUC4 is a membrane-anchored protein composed of a transmembrane (TM) domain, a cytoplasmic tail (CT), the Von Willebrand factor (VWF) domain, the adhesion-associated (AMOP) domain, the Nidogen-like (NIDO) domain, tandem repeats of PTS domains with O-glycosylation chains attached to it, and the EGF-like domain with three EGF-like sequences interspersed with N-glycosylation sites. *Figure created with Biorender.com.*

1.5.3 Mucins in cancer

Transmembrane mucins are overexpressed in various carcinomas including pancreatic (MUC4, MUC1, MUC16), breast (MUC1, MUC3, MUC4), ovarian (MUC1, MUC4, MUC13, MUC16), colorectal (MUC1, MUC3, MUC13), gastric (MUC3), and lung and prostate carcinomas (MUC1, MUC4) (Hanson *et al.* 2016, Dhanisha *et al.* 2018). It has been shown that shedding of the extracellular part of transmembrane mucins is upregulated during cancer metastasis. Overexpression of these mucins in carcinoma cells *in vitro* increases invasiveness as interaction between collagen and laminin with the cells are hampered by the bulk of the PTS barrier. In addition to decrease adhesion and increase invasion, the dense sugar barrier can also protect cancer cells against destruction by the immune system and by anti-cancer drugs. Besides, transmembrane mucins can inhibit anoikis (i.e., cell death induced by loss of adhesion) by masking the Fas ligand at the surface of the cells with their PTS layer.

Moreover, MUC1 can form a complex with β-catenin via the cytoplasmic domain and regulate genes involved in proliferation and differentiation. MUC16 can also directly bind to the chromatin and interact with transcription factors (van Putten *et al.* 2017).

Aberrant glycosylation of mucins is observed in carcinoma where they constitute tumour-associated carbohydrate antigens (TAAs) that can be used as glycosylation-based biomarkers. O-glycan chains observed in cancer typically display more sialyl motifs and less sulphate motifs compared to normal O-glycan chains. They can be truncated because of incomplete synthesis, as well as aberrantly extended. Mucins are thus both aberrantly expressed and glycosylated in cancer. Differential glycosylation can be explained by differential expression of transferases between healthy and tumour cells (Jonckheere *et al.* 2013, van Putten *et al.* 2017). Mucins are present in the serum of patients, for instance MUC1, MUC13, MUC16, and their aberrant glycosylated forms have been shown to have a prognostic value in cancer (Filippou *et al.* 2018, Reily *et al.* 2019). These truncated carbohydrate antigens (Tn, T, sialyl-Tn) can additionally serve as potential targets for immunotherapies (Hanson *et al.* 2016). Differential glycosylation of proteins between healthy and tumour cells are believed to play a role in cancer progression. Increased level of TAAs can be associated with malignancy and poor prognosis in breast, pancreatic, and colorectal cancer patients (MUC1) and in ovarian cancer patients (MUC16). It is considered that MUC4 and MUC1 expose the most potent TAAs (van Putten *et al.* 2017).

1.5.4 MUC4 in cancer

MUC4 overexpression is observed at the surface of almost all carcinoma cells, notably in pancreatic, gall bladder, ovarian, breast, prostate, oesophagus, and lung carcinomas. In pancreatic ductal adenocarcinoma, overexpression of MUC4 is a biomarker for poor prognosis. MUC4 is the most differentially expressed gene in this type of cancer and its expression correlates with the disease grades (Kufe 2009). Moreover, MUC4 is a biomarker for poor prognosis in other carcinomas including lung (Tsutsumida *et al.* 2007), gall bladder (Lee *et al.* 2012), oral squamous cell (Hamada *et al.* 2012) and, ovarian carcinomas (Subhash C Chauhan *et al.* 2006). In a rat model, MUC4 was shown to act as a pro-oncogene as its overexpression induced a large array of oncogenic mechanisms including disruption of cell-cell and cell-ECM

interactions, promotion of cell survival, and augmentation of growth and invasion. Of interest, the phosphorylation of the cytoplasmic domain of MUC4 and its involvement in intracellular signal transduction has not yet been demonstrated (Xia *et al.* 2017).

MUC4 is believed to act by blocking the surface of the cells with its large PTS layer when overexpressed, impeding the fixation of antibodies like trastuzumab in the case of breast cancer, thus rendering the cells resistant to monoclonal antibody-targeted therapy (Jonckheere *et al.* 2013, Chugh *et al.* 2015, Mercogliano *et al.* 2017) (**Fig.1.8A**). MUC4 silencing counteracted the resistance to trastuzumab in gastric cancer cells (Jonckheere *et al.* 2014). Via a similar mechanism, MUC4 could prevent ECM components to reach their receptors at the surface of the cells, like integrins and CD44, during migration (Chaturvedi *et al.* 2007) (**Fig.1.8A**). In addition, in cancer cells, MUC4 was shown to abrogate different normal cell mechanisms including apoptosis (Komatsu *et al.* 2001), contact-inhibition (Pino *et al.* 2006) (**Fig.1.8B**), and recognition by immune cells (**Fig.1.8A**). Besides these mechanisms, it is possible that aberrant glycosylation of MUC4 in cancer can be, at least in part, responsible for its oncogenic role. The PTS layer seems to promote cancer progression through several mechanisms, but paradoxically, its absence was shown to have similar effects. For example, a hypo-glycosylated form of MUC4 could expose its AMOP domain, which was shown to play a role in metastasis of pancreatic cancer cells (Hanson *et al.* 2016). Moreover, a splice variant of MUC4 lacking the PTS domain is upregulated in poorly differentiated pancreatic tumours and induces proliferation, invasion, and adhesion to the ECM (Jahan *et al.* 2018).

MUC4 also supports cancer progression and metastasis through different signalling mechanisms. Galectin-3 (Gal-3) can bind to O-glycan chains of MUC4 (as it is observed for MUC1) and induces the interaction between cancer cells and endothelial cells (Senapati *et al.* 2011). MUC4 interacts with Gal-3 via the T antigen on its surface, which sparks off the clustering of MUC4 at the cell surface. Adhesion molecules, such as integrins, are thus exposed and facilitate the docking of cancer cells to endothelial cells for extravasation initiation (**Fig.1.8C**). In pancreatic cancer, MUC4 induces trafficking of β-catenin to the nucleus via lysosomal degradation of E-cadherin. Accumulation of β-catenin in the nucleus activates its targeted genes including VEGF

and CD44 and thereby, promotes proliferation, metastasis, and angiogenesis (Zhi *et al.* 2014). It was shown that MUC4 overexpression could trigger an EMT process associated with increased MMP9 expression in ovarian cancer cells (Ponnusamy *et al.* 2010).

MUC4 expression can be regulated by diverse factors including cytokines. IFNγ induces MUC4 expression in pancreatic cell lines. TGFβ, TNFα, INFγ, IL-6, and retinoic acid positively regulate MUC4 via transcription factor activation, mainly the STAT pathway. In turn, MUC4 can regulate several factors involved in tumourigenicity. Importantly, in pancreatic cancer, the AMOP domain of MUC4 leads to increased expression of VEGFA, MMP9, and ANG-2 via activation of the NOTCH3 signalling pathway (Tang *et al.* 2016).

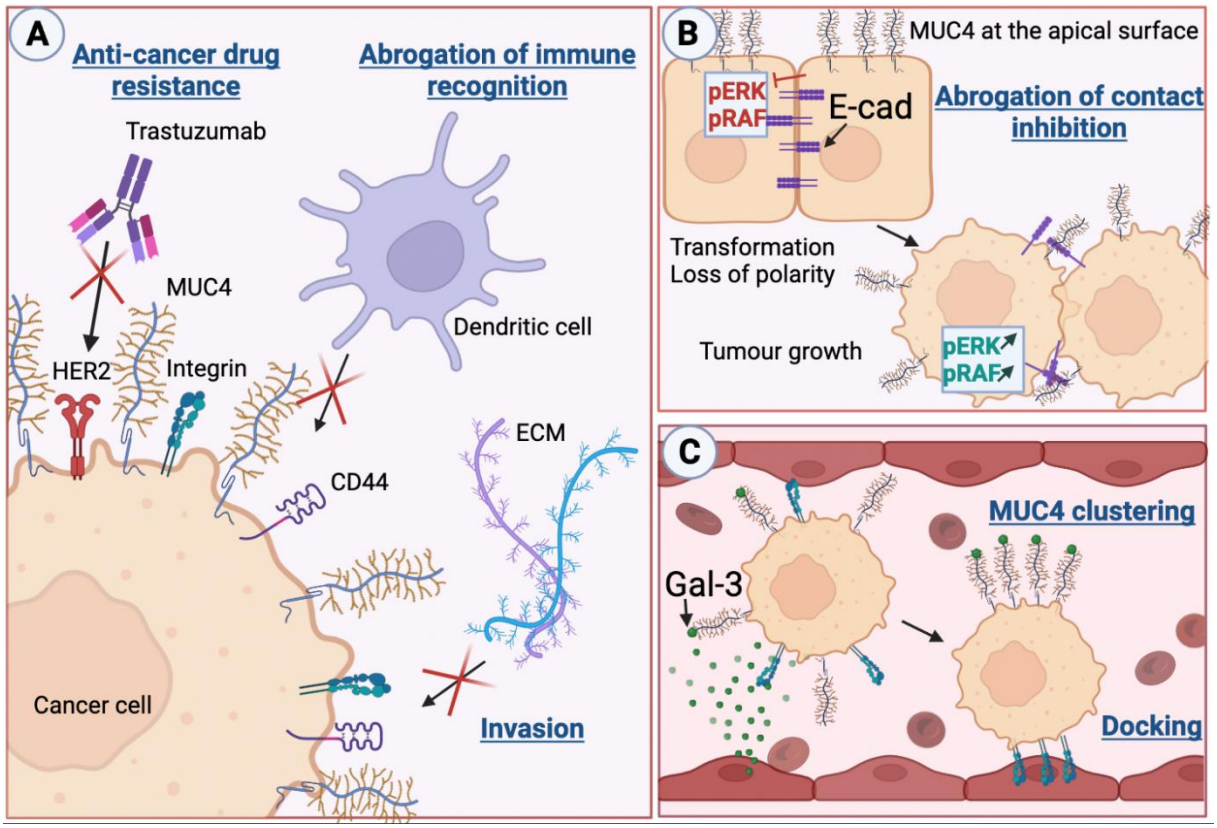


Figure 1.8 MUC4 implication in carcinogenesis. (A) Overexpression of MUC4 on cancer cells abrogates anti-HER antibodies' binding, which leads to drug resistance; interactions between cancer cells and the extracellular matrix (ECM), which promotes invasion; and recognition of cancer cells by dendritic cells, which contributes to carcinogenesis. **(B)** In non-carcinogenic epithelial cells, cell-cell interactions mediated by cadherins maintain low activation of the RAF/ERK pathway in the cell. After transformation, the cell polarity is lost and MUC4 is not restricted to the apical surface anymore but expressed circumferentially, which prevents cadherin molecules to form heterodimers at the lateral side of the cell. This abrogates normal contact-inhibition and leads to increased level of activation of the RAF/ERK pathway, leading to tumour growth. **(C)** Galectin-3 (Gal-3) is secreted by endothelial cells and binds to MUC4 expressed at the surface of cancer cells that penetrated the circulation, leading to MUC4 clustering at the surface of the cell and exposition of integrins. Integrins permit the docking of cancer cells to endothelial cells, which initiates extravasation, the first step of metastatic dissemination in carcinoma.(Pino *et al.* 2006, Chaturvedi *et al.* 2007, Senapati *et al.* 2011, Jonckheere *et al.* 2013, Chugh *et al.* 2015, Mercogliano *et al.* 2017). *Figure created with Biorender.com.*

1.5.5 MUC4-based targeted therapies

MUC4 is not targeted yet in cancer by classical targeted chemotherapy, however, it was suggested that its sequence may be directly targeted in the future by miRNA-based therapy, such as miR200c, miR21913p, and miR150 in pancreatic cancer and melanoma (Kufe 2009, Dhanisha *et al.* 2018). In addition, cancer vaccines based on TAA presentation could be used for the production of tumour-specific antibodies. This approach relies on the activation of T-helpers via antigen presentation, which in turn can activate cytotoxic cells and B cells. Synthetic MUC4 glycopeptides carrying TAAs could induce a specific immune response. In that case, the glycosylation pattern used dictates the specificity and efficacy of the immune response and therefore, would need to be carefully optimised beforehand (Torres *et al.* 2012).

Interestingly, some plant-derived compounds with antioxidant properties have proven to downregulate MUC4 expression in several studies. Thymoquinone (TMQ), extracted from Nigella sativa (black seed), was shown to decrease MUC4 protein level in pancreatic cells potentially via removing sialic acid (N-acetyl-neuraminic acid) from the PTS domain, which is involved in selectin recognition (Macha *et al.* 2015). Importantly, TMQ is non-toxic to normal cells. However, the exact molecular targets are yet to be characterised even though some pathways were identified. Importantly, TMQ was shown to induce cell cycle arrest, apoptosis, and inhibition of proliferation. It was also shown to inhibit angiogenesis, and to overcome chemoresistance. For these reasons, TMQ could potentiate chemotherapy in GBM when administered in association. TMQ has been shown useful in combination to various anti-cancer drugs in different cancer cell lines and importantly, has been shown to reduce MMP2 and MMP9 expression in breast cancer, lung, and pancreatic cell lines (Imran *et al.* 2018). Finally, TMQ has been shown to reduce the MUC4 protein level in pancreatic cancer cells (Torres *et al.* 2010), and to reduce secretion of MMP9 and MMP2 protein level in GBM cells (Kolli-Bouhafs *et al.* 2012).

1.5.6 Interaction between EGFR and mucins in cancer

MUC1 plays a role in carcinogenesis in part through activation of EGFR. MUC1 is overexpressed at the surface of carcinoma cells and was shown to interact with several growth factor receptors including c-MET (HGF receptor), EGFR, and PDGFRb (Hanson *et al.* 2016). The truncated form of MUC1 interacts with EGFR (HER1) and

supports its activation when their respective ligands, Gal-3 and EGF, are available in the extracellular space. MUC1 can activate EGFR when it is linked to its ligand EGF without Gal-3. However, the binding of Gal-3, whose expression is altered in carcinogenesis, on the truncated form of MUC1 potentiates the activation of EGFR and prolongs the downstream activation of the ERK1/2 pathway (Piyush *et al.* 2017). The use of an inhibitory peptide blocking the interaction between MUC1 and EGFR led to reduced malignancy of breast cancer cells in an *in vitro* assay and inhibited growth and recurrence *in vivo* (Bitler *et al.* 2009).

MUC4 has also been shown to interact with EGFR family members at the surface of carcinoma cells. MUC4 forms a complex with HER2, which is located at the lateral side of non-carcinogenic epithelial cells. This interaction induces their translocation at the apical surface of the cell and prevents the pro-oncogenic interaction between HER2 and its partner HER3, which, in the presence of neuregulin, would induce proliferation, invasion, and survival. Besides, the HER2/MUC4 complex activates the p38 pathway, which maintains cell differentiation. Thus, interaction of HER2 and MUC4 in physiological conditions maintains normal cell polarity and differentiation and prevents HER2 from exerting its oncogenic effects. During neoplastic transformation and loss of cell polarity, the normal interaction between HER2 and MUC4 is lost over the formation of an aberrant HER2/HER3 complex stabilised by MUC4 and neuregulin. This new interaction leads to phosphorylation of the cytoplasmic domain of HER2 and activation of the ERK and AKT pathways (Carraway *et al.* 2009) (**Fig.1.9A**).

MUC4 is also thought to act as a transmembrane ligand that induces the dimerisation and transactivation of HER2, which does not have any extracellular ligand (Dhanisha *et al.* 2018). Activation of HER2 has been shown to disrupt the PAR-2/PAR-6 complex that controls and maintains cell polarity, leading to the disruption of tight junctions. MUC4 may also be responsible for β-catenin sequestration via again interaction with HER2 at the apical membrane, which would cause disruption of adherens junctions (Kufe 2009). In addition, some MUC4 transcriptional variants lack the transmembrane domain and thereby could be secreted. In this case the EGF-like sequences of MUC4 could directly act as a ligand on EGF receptors (Chaturvedi *et al.* 2008). In another study, a model was proposed in which MUC4 forms a complex with HER2 at the

surface of ovarian cancer cells, leading to upregulation of N-cadherin, which would induce the EMT process with an increase of MMP9 through the N-cadherin/Akt/ERK pathway (Ponnusamy *et al.* 2010). A similar model was recently proposed in pancreatic cancer, where MUC4 and HER2 physically interact via the EGF-like sequences of MUC4 (Stoup *et al.* 2021) and induce invasion through activation of several signalling pathways such as mTOR, Akt, and β-catenin. In parallel, it was also suggested that MUC4 could induce invasion indirectly through MMP9 secretion and action in pancreatic cells (Tang *et al.* 2016, Stoup *et al.* 2021). These observations support that targeting of MUC4 in cancer with different strategies, as mentioned earlier (miRNA-based therapy, anti-cancer vaccines, natural compounds) could be beneficial for cancer patients through downregulation of MUC4-EGFR interaction and MMP9 expression (**Fig.1.9B**).

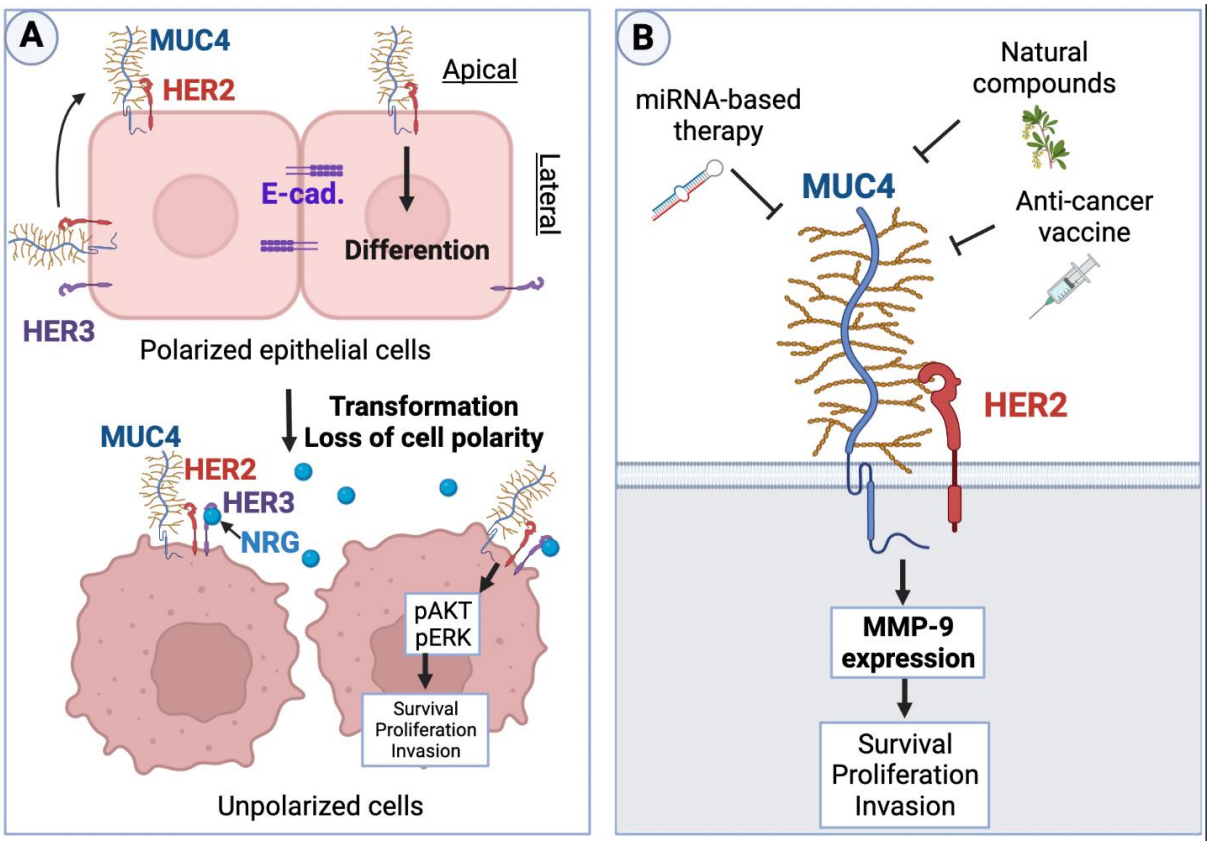


Figure 1.9 Model of MUC4-EGFR interaction in cancer. (A) In normal epithelial cells, HER3 is restricted to the lateral side whereas interaction between MUC4 and HER2 leads to their translocation to the apical side where they maintain normal differentiation of the cells. After oncogenic transformation, polarity and HER3 lateral restriction are lost, MUC4 forms a complex with HER2 and HER3. When neuregulin (NRG) binds to HER3, the complex is activated, HER2 tyrosine kinase activity transduces the Akt and ERK signalling pathway, which promotes downstream oncogenic responses. **(B)** Interaction between MUC4 and HER2 through EGF-like sequences of MUC4 at the surface of carcinoma cells promotes survival, proliferation, and invasion via downstream activation of oncogenic pathways that are believed to trigger MMP9 expression. Potential strategies for MUC4 inhibition to stop cancer progression are represented. (Carraway *et al.* 2009, Ponnusamy *et al.* 2010, Tang *et al.* 2016, Stoup *et al.* 2021). *Figure created with Biorender.com.*

1.5.7 Mucins in GBM

Taken together, these above-mentioned findings suggest that mucins play crucial roles in carcinogenesis that are still under ongoing investigations. However, very few studies on their expression and function in non-epithelial cancers have been conducted. Research studies investigating the involvement of mucins in glioma progression are limited, probably because mucins are not expressed in the healthy brain. A few studies showed that mucins are being increasingly correlated with GBM progression. First, the expression of MUC15 has been found upregulated in glioma samples compared with healthy brain tissue and the expression of MUC15 correlated with both tumour progression and shorter survival (Yang *et al.* 2014). Moreover, MUC13 was also found overexpressed in GBM stem cells and its overexpression enhanced the cell migrative behaviour (Li *et al.* 2017). MUC1 C-terminal subunit expression was found to be higher in grade III glioma and GBM compared to the lower grades. Its overexpression increased colony formation ability of glioma cells *in vitro*. Expression of MUC1 C-terminal subunit in glioma was shown to be regulated by miR-194 and the lncRNA LINC00909, whose expression is dysregulated in glioma (Liu *et al.* 2019).

MUC4 was shown to be overexpressed in GBM samples and GBM cell lines. Ectopic expression of MUC4 in GBM cell lines induced increased proliferation and invasion abilities with similar effects to those observed after EGFR overexpression in these cells. Besides, MUC4 inhibition in these GBM cell lines resulted in downregulation of EGFR expression both at mRNA and protein levels, which indicates that EGFR and MUC4 proteins may be part of a common pathway. Both proteins are found overexpressed together in GBM samples (Li *et al.* 2014). To sum up, MUC1, MUC15, MUC13, and MUC4 have been found overexpressed and involved in GBM and more importantly this seems to be linked to the EGFR pathway in the case of MUC4.

1.6 Hypothesis and aims

1.6.1 Hypothesis

In the light of the data mentioned above, one can posit that in GBM, aberrantly glycosylated/regulated variants of MUC4 may interact with EGFR. Furthermore, this

interaction could be, at least in part, responsible for high MMP9 expression observed in GBM and therefore, might considerably account for GBM aggressiveness.

The components of this pathway, taken together, could constitute a panel of potential biomarkers detectable from tissue and liquid biopsies, as well as therapeutic targets. Of note, MUC4 could constitute a particularly interesting biomarker since it is heavily glycosylated and therefore potentially easier to identify. In addition, its absence in healthy glial cells makes MUC4 potentially a highly specific biomarker in brain cancer. Anti-EGFR therapy in GBM has represented a promising treatment strategy but fails at being efficient when given alone. Besides, MUC4 inhibition is not yet contemplated for GBM treatment because of a knowledge gap on its involvement in such tumours. Although MMP9 has been shown to be a promising target in GBM, current therapy targeting directly this protein is not considered safe and specific enough. Moreover, the exact biology of MMP9 and MUC4 during GBM progression must be better clarified before these two proteins can be used in combination as therapeutic targets. This would allow among others, to avoid the toxic effects that their inhibition may induce and to determine the correct time windows for efficient delivery.

Innovative biomarkers are rarely deemed efficient for clinical decision-making. For this reason, many recently discovered biomarkers are not used in clinics even if they have been shown to be highly implicated in the disease progression, as is the case for MMP9. However, biomarkers that have diagnosis and/or prognosis value can be used to predict poor response to standard treatment and thus, can help to guide patients towards more innovative treatments (e.g., enrolment in clinical trials). Moreover, biomarkers that can be available from liquid biopsies are particularly helpful: they allow disease progression and treatment response monitoring as they are less invasive; they can be assessed during early stage of the disease (before surgical resection), and finally, might be more representative of the cancer than tissue biomarkers detected from a restricted portion of the tumour.

1.6.2 Aims

The aims of the research studies carried during this PhD thesis are to:

1. **Evaluate the diagnostic and prognostic potential of EGFR, MMP9, and MUC4 in glioma tissue and serum samples.** The first chapter of this thesis aims at characterizing the expression of MUC4, alone and in association with MMP9 and EGFR, in glioma and determine its prognostic (relation with survival) and diagnostic (relation with histological grades and IDH genotypes) potential as a biomarker. The localisation and putative co-localisation of the proteins within the tissue will be assessed. To address this, immunohistochemistry (IHC) and immunofluorescence (IF) labelling will be performed in formalin-fixed, paraffin-embedded (FFPE) biopsy slides from grade II, grade III, and grade IV glioma patients. The association between protein expression and the survival will be determined with the Kaplan-Meier method. Finally, to assess the value of the proteins as non-invasive liquid biopsy biomarkers, the levels of MMP9 and MUC4 in the serum of the patients will also be studied by using ELISA immunoassays.

2. **Delineate the putative MUC4/EGFR/MMP9 pathway in GBM cells with *in vitro* assays.** The second chapter aims at deciphering the effect of MUC4 in glioma cells, as well as the localisation and potential interaction of MUC4 and EGFR, and to investigate the regulation of this pathway at the gene expression level. To do this, GBM cell lines will be used. MUC4 will be transfected in glioma cells to observe the effect of the protein's expression on the cell behaviour in spheroids. Secondly, immunocytochemistry (ICC) will be performed to study the proteins' localisation. RT-qPCR will be used to study the gene expression after the addition of different ligands and drugs, which will shed light on the regulation of this pathway.

3. **Determine to which extent biomolecular and glycosylation changes between different glioma cells can be captured with Raman spectroscopy (RS).** The third chapter aims at assessing the efficacy of Raman spectroscopy at discriminating between glioma grades and more importantly between grade III and grade IV based on the level of glycosylation patterns present in live cells, notably the one introduced by MUC4 overexpression. Changes in the bio-composition of the samples in the different glioma grades will be explored using an in-house glycosylation database and a more general database from the literature.

4. **Assess the potential of Raman spectroscopy (RS) for glioma diagnosis and glycosylation monitoring in clinical samples.** The fourth chapter aims at determining whether the findings of the preceding chapter can be applied in clinical FFPE samples for objective glioma grade discrimination. In addition, the biomolecular changes will be assigned by comparison to the database and parallel monitoring of biological changes between the grades will be performed with a particular focus on glycosylation. Finally, the same approach will be used on serum sample (liquid biopsy), which would constitute a better type of sample since collected in a less invasive way for the patients.

Chapter 2: The EGFR/MUC4/MMP9 diagnostic and prognostic value in glioma tissue and serum samples

ABSTRACT

Background: There is currently an urgent need for novel biomarkers and therapeutic targets for GBM patients. In this chapter, the combined role of mucin 4 (MUC4), matrix metalloproteinase-9 (MMP9), and the epidermal growth factor receptor (EGFR) in the progression and clinical outcome of glioma patients was investigated.

Method: Immunohistochemistry (IHC) and immunofluorescence (IF) in formalin-fixed, paraffin-embedded (FFPE) tissue samples were used to evaluate the expression and localisation of EGFR, MMP9, and MUC4 in glioma. Kaplan-Meier survival analysis was also performed to test the prognostic significance of the proteins for glioma patients. Finally, the protein levels were assessed with enzyme-linked immunosorbent assay (ELISA) in blood serum of glioma patients, to further investigate their potential as non-invasive liquid biomarkers.

Results: Both MUC4 and MMP9 expression was significantly correlated with the histological glioma grades and IDH1 genotypes, with the highest expression observed in GBM/wt-IDH1 patients. Moreover, MUC4 and MMP9 were co-expressed in the proliferative microvasculature of GBM, suggesting a new combined role for MUC4/MMP9 in angiogenesis and microvascular proliferation (MVP). The combined high expression of MUC4/MMP9, and MUC4/MMP9/EGFR was significantly associated with poor overall survival (OS). Finally, MMP9 levels were significantly higher in the serum of GBM patients compared to grade III patients, whereas MUC4 level was low in all cases.

Conclusion: These results strengthen the hypothesis that MUC4, along with MMP9, might account for GBM progression, representing a potential novel therapeutic target. Moreover, this suggests the 'MUC4/MMP9/EGFR axis' may be used in GBM diagnostics and prognostics.

2.1 Introduction

Glioma diagnosis is usually made by assessing histological characteristics and IHC staining of biomarkers in solid tissue samples and is thus susceptible to pathologist-dependent bias. Naturally, accurate diagnosis in glioma patients is crucial for treatment decisions. Usually, using a single protein as a biomarker is insufficient for an accurate and specific diagnosis; thus, there is the tendency to combine several biomarkers in multi-analyte panels to improve diagnostic accuracy (Yurkovetsky *et al.* 2006, Borgia *et al.* 2009). So far, several biomarkers have been identified in GBM. IDH1/2 genotyping is the prevailing current marker for distinction between GBM and m-IDH grade 4 astrocytoma in the 2021 WHO classification (Chen *et al.* 2017). However, as the prognosis of GBM remains poor, a need for novel biomarkers and potential therapeutic targets is warranted.

MUC4 is overexpressed in various carcinomas and has been demonstrated to interact with various EGFR family members (Carraway *et al.* 2009, Dhanisha *et al.* 2018). However, only a few studies (Li *et al.* 2014, Yang *et al.* 2014, Cheng *et al.* 2020) have investigated the implication of mucins in glioma. According to the Human Protein Atlas database, MUC4 is not expressed in the healthy brain. Nonetheless, MUC4 overexpression is observed in GBM patient samples and cell lines. In addition, *in vitro* assays have shown the implication of MUC4 in proliferation and invasion properties of GBM cells (Li *et al.* 2014). Of note, MUC4 was one of the three most frequently mutated genes in grade III patients progressing towards histological grade IV astrocytoma, suggesting its involvement in gliomagenesis (Seifert *et al.* 2020). In GBM, EGFR is the most commonly amplified gene (Karpel-Massler *et al.* 2009), while MMP9 has been previously suggested as a prognostic and diagnostic factor in glioma (Xue *et al.* 2017). MMP9 has been shown to be transcriptionally regulated by EGFR signalling and involved in the regulation of key signalling pathways in glioma (Choe *et al.* 2002, Chang *et al.* 2015, Chen *et al.* 2017). In general, EGFR interacts with members of the mucin family in cancer (Carraway *et al.* 2009, Dhanisha *et al.* 2018), and as such, MUC4 may physically interact with EGFR in a ligand-dependent manner. This interaction may lead to downstream signalling and MMP9-mediated glioma initiation and progression.

In this chapter, EGFR, MMP9, and MUC4 were hypothesised to be part of a common signalling axis that could constitute a panel of novel biomarkers for diagnosis and

prognosis of glioma patients. The goal of this chapter was to evaluate the diagnostic and prognostic potential of EGFR, MMP9, and MUC4 in glioma tissue and serum samples. First, the protein expression and cellular (co)-localisation were investigated in tissue samples from patients diagnosed with grade II, grade III, and grade IV gliomas, as well as their association with clinicopathological features. The prognostic significance was next evaluated by analysing the association between protein expression levels and clinical outcome of glioma patients using the Kaplan-Meier survival analysis method. Finally, the potential of MUC4 and MMP9 as liquid biomarkers was explored by measuring the protein levels in serum samples of high-grade glioma patients (grade III, IV) and non-glioma (control) individuals.

2.2 Material and methods

2.2.1 Patients and clinical samples

To analyse the significance of the three potential biomarkers (EGFR, MMP9, and MUC4) in glioma patients, tissue sections from grade II (diffuse astrocytoma and oligodendroglioma), grade III (anaplastic astrocytoma and anaplastic oligodendroglioma), and grade IV (GBM) gliomas were assessed. Since grade I glioma (pilocytic astrocytoma) is molecularly and histologically distinct from the other glioma grades, this grade was considered to be non-relevant to this study (Arcella *et al.* 2020).

20 μm-sections of 60 formalin-fixed, paraffin-embedded (FFPE) tumour sections, obtained from tumour debulking surgery or biopsy, were investigated in total. 10 grade II, 10 grade III, 10 grade IV samples were obtained from the Manchester Cancer Research Centre (MCRC) Biobank (Manchester, UK) and corresponding numbers from the NovoPath Biobank Newcastle (Newcastle-upon-Tyne, UK). Non-glioma brain tissue samples used as controls (2 patients) were obtained from Newcastle Brain Tissue Resource (NBTR) (Newcastle-upon-Tyne, UK). All diagnoses were performed by a local consultant neuropathologist, following the 2016 WHO classification of the central nervous system tumours (the samples have been collected before the 2021 WHO classification update).

30 blood serum samples (10 non-glioma benign tumours (controls), 10 grade III, and 10 grade IV) were obtained from the MCRC biobank (Manchester, UK). The serum samples were stored at −80 °C prior to analysis.

2.2.2 Ethical approval

The Research Ethics Boards approved sample collection of the respective Biobanks (Manchester Cancer Research Centre (MCRC) (REC ref 18/NW/0092) and NovoPath Biobank Newcastle (REC Ref 17/NE/0070), UK) and the Teesside University Research Ethics Committee upon receipt of the ethical approval. Brain tissue was obtained from the NBTR, a UK Human Tissue Authority-approved research tissue depository, and ethical approval granted by the Newcastle University ethics board and the Joint Ethics Committee of Newcastle and North Tyneside Health Authority (ref: 08/H0906/136). All procedures followed The Declaration of Helsinki.

2.2.3 Immunohistochemistry (IHC)

IHC is a well-established method that allows to visualise the localisation of a protein of interest within a tissue section, or other types of organic samples, using antibodies (primary and secondary). The primary antibody is designed to specifically bind to the protein of interest, while the secondary antibody is designed to bind to the primary antibody. The latter is linked to an enzyme (peroxidase) producing a brown-coloured reagent (chromogen) after the addition of a substrate, usually diaminobenzidine tetrahydrochloride (DAB) (Erskine *et al.* 2019). The tissue slide can then be mounted and visualised easily with microscopy. The localisation pattern and the intensity of the staining, which is representative of the protein expression, can then be analysed visually from microscope images. Statistical analyses can also be performed on categorised variables. For instance, samples can be assigned to different intensity IHC-scores or different scores corresponding to different percentages of expression on which later, statistical tests can be performed. Qualitative (localisation of the protein) and quantitative (level of expression categorised) approaches were both used in this study for EGFR, MUC4, and MMP9 expression.

Primary antibodies were first tested and concentrations optimised on a few glioma samples (several slides were available for each patient). Negative controls in one sample of each grade were also used, following the same procedure with the only

difference that the addition of the primary antibodies was omitted, to verify that the non-specific background staining was low. FFPE tissue sections were deparaffinised by serial incubation for 10 minutes in xylene (Honeywell, #534056-4L) and decreasing concentrations of ethanol (VWR chemicals, #20821.330) mixed with distilled water (100%, 95%, 70%, and 50% ethanol) before being soaked in pure distilled water. Heat-induced antigen retrieval was then performed to unmask the target protein by boiling the slides in citrate buffer pH 6 or Tris-EDTA buffer pH 9 (**Table 2.1**), depending on the manufacturer's instruction for each antibody (**Table 2.2**). Tissue slides were soaked in the antigen retrieval buffer and boiled using a microwave for 15 minutes. A hydrophobic pen (Merck, #Z377821) was used to draw a liquid-repellent barrier around the samples and the following incubation was performed by directly placing the appropriate solution onto the sample (200 to 500 µl depending on the sample size). Endogenous peroxidases were blocked to avoid unspecific staining by incubation in a H_2O_2 3% solution (Sigma-Aldrich, #7722-84-1) at room temperature for 20 minutes. Samples were also blocked to avoid unspecific binding with host immunoglobulins by incubation with a blocking solution (**Table 2.1**) for one hour at room temperature. Samples were incubated with the primary antibodies (**Table 2.2**) diluted in the blocking solution overnight at 4°C.

The next day, universal probe horseradish peroxidase (HRP) and DAB (A. Menarini diagnostic kit, #MP-860-K100) were used to reveal the primary antibodies by using the same incubation time for all the samples stained with the same antibody. The slides were then mounted with Eukitt quick-hardening mounting medium (Sigma-Aldrich, #03989) and let to dry. Images were taken with a Leica microscope DM75 (Leica microsystem, UK) at total magnifications of x200 and x400. Haematoxylin (Sigma-Aldrich, #517-28-2) staining was performed on a parallel slide for each case to confirm histological grading. To avoid the inclusion of artefacts, slides were scored by the investigator, who was blind to clinicopathological data, rather than by automated techniques using software tools. Both the intensity and the proportion of expressing cells were considered during the quantitative analysis. Blinded scoring was performed from 10 views at a total magnification of x200 for each slide and each marker. 1800 images were analysed in total (60 patients x 3 antibodies x 10 images).

Since the IDH genotype is crucial in the 2021 classification of glioma, IHC staining for IDH1 mutant R132H was performed on the samples that had not been genotyped for IDH in the hospital. The same protocol described was followed by using citrate buffer as an antigen retrieval. The antibody used was IDH1 (mutant R132H) rabbit monoclonal antibody (Thermofisher scientist, #Z2010RT) used at a 1:200 dilution.

Buffers	Composition
Citrate buffer pH 6	0.05 % Tween-20
	3 mM sodium citrate
Tris EDTA pH 9	1 mM EDTA
	10 mM Tris
Tris-buffered saline (TBS) (1X)	150 mM NaCl
	20 mM Tris
Blocking solution	2% bovine serum albumin (BSA)
	10% normal horse serum
	TBS (1X)

Table 2.1 Composition of buffers used for IHC staining.

Protein	Dilution	Reference	Antigen retrieval buffer	Species
MUC4	1:250	Abcam, ab60720	Citrate, pH 6	Mouse
MMP9	1:750	Abcam, ab76003	Citrate, pH 6	Rabbit
EGFR	1:100	Abcam, ab52894	Tris-EDTA, pH 9	Rabbit

Table 2.2 Primary antibodies used for IHC staining.

2.2.4 Immunofluorescence (IF)

IF is a technique very similar to IHC with the difference that the secondary antibody is linked to a fluorophore, which can be visualised directly with a fluorescent microscope (Joshi *et al.* 2017). This technique allows to label more than one protein of interest on the same sample, since different fluorophores can be used and discriminated by the microscope thanks to their different excitation and emission wavelengths. Therefore, this technique can be used to reveal co-expression of pairs of biomarkers (double IF). Different technical aspects need to be carefully considered before performing double IF. First, the primary antibodies need to be produced from different species (in this study, mouse, and rabbit) whenever possible, to avoid cross-binding, and should not be produced from the host species (human) to avoid cross-reaction with the host immunoglobulins. The secondary antibodies (linked to different fluorophores) need to be produced from a fourth species (goat), and serum from that species (goat serum) should be used as a blocking agent during the blocking step. Finally, each secondary antibody needs to be targeted against the species of that the first antibodies it should recognise and label (Im *et al.* 2019).

FFPE slides were chemically dewaxed by serial incubation in xylene and decreasing concentrations of ethanol as described earlier. Antigen retrieval was performed by boiling the slides in citrate buffer pH 6 as described before. The slides were incubated in blocking solution (TBS, 0.05% Tween (TBS-T), 7% goat serum, 5% BSA) for one hour at room temperature and next with each primary antibody solution containing the primary antibody at the appropriate dilution (**Table 2.3**) diluted in antibody diluent solution (TBS-T, 2% goat serum, 5% BSA) at 4°C overnight. Fluorescently labelled secondary antibodies used were goat anti-rabbit IgG H&L (Alexa Fluor® 647 pre-adsorbed; Abcam, ab150083) and goat anti-mouse IgG H&L (Alexa Fluor® 555; Abcam, ab150114) diluted in the same antibody diluent as described before (1:500 dilution for both). The slides were incubated in the secondary antibody solution for one hour at room temperature and mounted with fluorescent mounting medium, which contained DAPI (Abcam, ab104139) to stain the nuclei. The mounting medium was left on the samples for one minute before mounting the slides. The slides were then left to dry overnight at room temperature in the dark and then stored at 4°C in the dark. Images were captured with a Leica SP8 fluorescent microscope (Leica microsystem, UK).

Protein	Dilution	Reference	Species
Anti-MMP9	1:500	Abcam, ab58803	Mouse
Anti-MMP9	1:1000	Abcam, ab38898	Rabbit
Anti-MUC4	1:250	Abcam, ab60720	Mouse
Anti-EGFR	1:400	Abcam, ab52894	Rabbit
Anti-CD31	1:200	Sigma, SAB5700639	Rabbit

Table 2.3 Primary antibodies used for IF staining.

2.2.5 ELISA immunoassays

ELISA is a well-established analytical technique that uses the same principle as IHC to detect the presence of a protein of interest in a liquid sample. It is typically used to assess the precise concentration of a specific protein in blood serum, cell line supernatant, or cell/tissue extract with high sensitivity and specificity. In this technique, micro-wells contained on a plate are coated with a capture antibody specific for an antigen present on the protein of interest. The bound antibody is incubated with the protein solution whose concentration needs to be determined. The detection antibody is next added and binds to another antigen present on the protein, so that the protein of interest is in the middle of the capture antibody and the detection antibody (sandwich-type technique). The latter is biotinylated and therefore binds to the enzyme streptavidin-peroxidase after its addition, or alternatively, the enzyme can be pre-linked to the detection antibody. The addition of a substrate generates a blue chromogen (product formation) when in contact with the enzyme. Sulfuric acid is added to stop the reaction before optical density measurement. This technique allows a precise determination of the concentration as a standard curve is performed with the protein of interest.

ELISA immunoassays were used to assess the concentration of MMP9 and MUC4 proteins in the serum samples. Measurement of the two forms of MMP9 (92 KDa pro-

form and 82KDa activated form) was performed using the Human MMP9 Quantikine ELISA kit (R&D systems, #DMP900). Human MMP9:TIMP-1 Complex DuoSet ELISA kit (R&D systems, # DY1449) was used to measure MMP9:TIMP-1 complex. Detection of MUC4 was performed using the Human MUC4 ELISA kit (Fine Bioteck co, # EH1160). Each kit was used according to the manufacturer's instructions.

For the MMP9 Quantikine and MMP9:TIMP1 ELISA assays, samples and standards were incubated for two hours in a plate pre-coated with a monoclonal antibody specific for human MMP9 at room temperature. For detection of MMP9, an HRP-conjugated anti-MMP9 antibody was directly added before addition of the substrate solution. For detection of MMP9:TIMP1, a biotinylated anti-TIMP1 antibody was added to each well before a two-hours incubation at room temperature followed by a 20-minutes incubation with Streptavidin-HRP. The substrate solution, tetramethylbenzidine chromogen (TMB) was added and left for 30 minutes at room temperature before a stop solution of sulfuric acid was added and optical density measured in each well. Three washes were performed between each step.

For the human MUC4 ELISA immunoassay, samples and standards were incubated for 90 minutes at 37°C in a plate pre-coated with a captured antibody specific for human MUC4. A biotinylated antibody was added to each well and left for one hour at 37°C. Then, an HRP-streptavidin conjugate was added for 30 minutes at 37°C and finally, the TMB substrate was added for 20 minutes in the dark at 37°C before the stop solution was added and optical density was measured. Three washes were performed between each step.

Optical density was measured with an Epoch 2 microplate spectrophotometer (BioTeK, VT, USA) at 450 nm with a wavelength correction set at 540 nm, made in duplicate for each sample. The mean blank value was subtracted. A standard curve for each protein (using different concentrations of the protein standard) was made for each assay from averaged duplicates and used to determine the protein concentration in each sample. The concentration of total protein content in each sample was also determined using a Quick start Bradford assay (Bio-Rad, CA, USA) in which a standard curve of BSA was used. The protein concentration was calculated using the relevant standard curve

for each ELISA (ng/ml) and normalised based on the total protein concentration (mg/ml) determined by the Bradford assay.

The MUC4 ELISA kit used was able to detect a minimal standard concentration of 0.31ng/ml, with a limit of detection of 0.188 ng/ml. For this assay, some values, calculated from the standard curve (for four patients) were inferior to this concentration. These values were arbitrarily defined as 0. All other MUC4 concentrations for each patient were multiplied by the dilution factor, then adjusted to the total concentration of protein in the serum of each patient (measured beforehand using the standard Bradford method).

2.2.6 Statistical analysis

Chi^2 test, Tukey's multiple comparison test, and log-rank test were performed using GraphPad Prism 9 software (GraphPad software, CA, USA). The Chi^2 test was used to evaluate the dependence between more than two categorical variables in glioma tissue samples. Tukey's multiple comparison test is used for pair-wise comparison of means of two populations. One-way ordinary ANOVA, followed by Tukey's multiple comparison test, was used to test whether the mean concentrations of proteins in the serum between pairs of grades were significantly different. All p-values given were 2-sided and a p-value ≤ 0.05 at a 95% confidence interval was considered statistically significant. p-values significance are defined as followed: *** indicates a p-value ≤0.001; ** indicates a p-value ≤0.01; * indicates a p-value ≤0.05.

The number of months between surgery and death (if any) was used to define the overall survival (OS) and the number of months between the surgery and relapse (if any) was used to define the progression-free survival (PFS). The Kaplan-Meier method and log-rank test were used to test the null hypothesis of no difference between two survival curves. The EGFR-high group comprised score-2 and the MMP9-high and MUC4-high groups comprised score-1 and score-2. This decision was made to compare groups of similar size since EGFR was expressed in most patients.

2.3 Results

2.3.1 Patient clinicopathological characteristics of tissue samples

The histological 2016 WHO classification was followed in this study as the samples were collected in 2019 before the 2021 WHO classification update, which now mostly relies on cytogenetic characteristics (Arcella *et al.* 2020). However, IDH-genotype, which is the most important feature in the 2021 WHO classification, was also available for almost all the patients. Therefore, for patients not clinically IDH-genotyped (6 grade IV patients), IHC staining for IDH1 mutant R132H, which is the most common IDH mutation (more than 90%) (Parsons *et al.* 2008), was performed on parallel glass slides. GBM cases with a known IDH1 genotype were used as controls (one mutated (m) and one wild-type (wt)). All six GBM patients with an unspecified genotype were wt-IDH1 (negative staining) from the IHC. Eventually, IDH1 genotypes for all the glioma patients in the study were known.

Patient characteristics are listed in **Table 2.4** with 20 patients per grades (60 in total).

Case	Grade	Sex	Age (yrs) at surgery	IDH genotype	OS (months since surgery)	Relapse	Treatment pre-surgery	Case	Grade	Sex	Age (yrs) at surgery	IDH genotype	OS (months since surgery)	Relapse	Treatment pre-surgery
1	II	F	40	m				**31**	III	M	68	m			
2	II	F	60	m	78	Yes		**32**	III	M	57	WT	3		
3	II	F	54	m				**33**	III	F	64	m			
4	II	F	28	m				**34**	III	M	76	WT	14		
5	II	M	49	m				**35**	III	M	56	WT	16		
6	II	F	24	m		Yes		**36**	III	M	33	m			
7	II	F	72	m				**37**	III	M	59	WT	20		
8	II	M	37	m		Yes		**38**	III	M	36	m	89	Yes	XRT
9	II	F	67	m	20			**39**	III	M	36	m			
10	II	F	58	m		Yes		**40**	III	M	54	m			
11	II	M	63	m				**41**	IV	F	64	WT	18	Yes	
12	II	M	67	m				**42**	IV	M	58	WT		Yes	
13	II	F	41	m	10			**43**	IV	M	56	WT	10	Yes	
14	II	F	79	WT	4	Yes		**44**	IV	M	64	WT	0		
15	II	M	52	m				**45**	IV	F	52	WT	3	Yes	
16	II	F	57	m				**46**	IV	F	73	WT	7		
17	II	M	23	m				**47**	IV	F	74	WT			
18	II	F	38	m	63	Yes		**48**	IV	M	69	WT	18	Yes	
19	II	M	26	m		Yes		**49**	IV	F	41	WT	17	Yes	
20	II	M	25	m				**50**	IV	F	70	WT	6		
21	III	M	53	m	30	Yes	XRT, PCV, TMZ	**51**	IV	F	67	WT	3	Yes	
22	III	M	53	m	93			**52**	IV	M	84	WT			
23	III	F	36	m		Yes		**53**	IV	M	40	WT	6	Yes	
24	III	F	44	m				**54**	IV	M	44	WT	8		
25	III	F	68	m	67		XRT	**55**	IV	M	73	WT	2	Yes	
26	III	M	41	m	37	Yes	XRT	**56**	IV	F	57	WT	7	Yes	
27	III	F	42	WT	28	Yes	PCV	**57**	IV	F	44	m	60	Yes	XRT
28	III	F	56	m	33	Yes	PCV and TMZ	**58**	IV	M	61	WT	14		
29	III	F	25	m				**59**	IV	M	68	WT			
30	III	F	42	m				**60**	IV	M	61	WT	12		

Table 2.4 Patient characteristics in fixed glioma tissues. [Years (yrs), Overall survival (OS), Radiotherapy (XRT), procarbazine, lomustine, vincristine (PCV), temozolomide (TMZ), mutated (m), wild-type (WT)].

The number of wt-IDH patients significantly increased with the glioma grade ($p<0.0001$, Chi^2). This was expected as m-IDH IV gliomas are generally scarce and are likely to have arisen from a lower-grade glioma (Wesseling 2018, Louis *et al.* 2021) (**Fig.2.1A**). There was no significant difference between the number of women and men in each grade ($p=0.42$, Chi^2), nor in the number of relapses between the grades ($p=0.23$, Chi^2). The mean age was 48 years old for grade II patients, 50 for grade III, and 61 for grade IV, and the number of patients younger than 65 years old was not significantly different between the grades (pairwise comparison, Fisher's exact) (**Fig.2.1A**).

Overall survival (OS) is defined by the number of months between surgery (tissue collection) and death. The number of patients surviving less than 12 months was significantly higher in grade IV than in grade III ($p=0.006$, Fisher's exact) (**Fig.2.1A**). There were 15 patients still alive in the grade II group, 9 in the grade III group, and 4 in the grade IV group at the time of data collection. The mean OS was 35, 39, and 12 months for grade II, III, and IV, respectively. The mean OS for grade IV when excluding the only m-IDH patient was reduced to 8.7 months. Finally, the difference between the survival curves of m-IDH patients and wt-IDH patients showed high significance ($p<0.0001$, Log-rank) by using the Kaplan-Meier method (**Fig.2.1B**).

A

	N(%)	Grade II	Grade III	Grade IV	P-value (Chi-2)
IDH genotype	IDH-m	19 (31.6)	15 (25)	1 (1.6)	<0.0001 (****)
	IDH-WT	1 (1.6)	5 (8.3)	19 (31.6)	
Sex	Female	12 (60)	8 (40)	9 (45)	0.42 (ns)
	Male	8 (40)	12 (60)	11(55)	
Relapse	Yes	7 (35)	6 (30)	11 (55)	0.23 (ns)
	No	13 (65)	14 (70)	9 (45)	
	N (%)	Grade II	Grade III	P-value (Fisher's exact)	
Age	<65	16 (80)	17 (85)	>0.99 (ns)	
	≥65	4 (20)	3 (15)		
OS	<12	2 (40)	1 (9)	0.2 (ns)	
	≥12	3 (60)	10 (91)		
		Grade III	Grade IV	P-value (Fisher's exact)	
Age	<65	17 (85)	12 (60)	0.15 (ns)	
	≥65	3 (15)	8 (40)		
OS	<12	1 (9)	10 (62)	0.006 (**)	
	≥12	10 (91)	6 (38)		
		Grade II	Grade IV	P-value (Fisher's exact)	
Age	<65	16 (80)	12 (60)	0.3 (ns)	
	≥65	4 (20)	8 (40)		
OS	<12	2 (40)	10 (62)	0.6 (ns)	
	≥12	3 (60)	6 (38)		

B

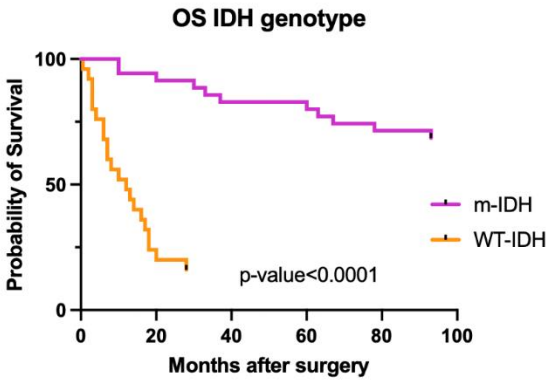


Figure 2.1 Association between patient characteristics. (A) Association between IDH genotype, sex, and relapse, with glioma histological grades was assessed by chi^2 test; and association between age and overall survival (OS) with glioma grades was assessed by Fisher's exact test. The association between wt-IDH and higher glioma grades was strongly significant (****) and OS was significantly lower in grade IV patients compared with grade III patients (**). **(B)** Overall survival (OS) curves of mutated-IDH (m-IDH) and wild-type-IDH (wt-IDH) patients were generated with the Kaplan-Meier method. The OS of wt-IDH patients was significantly lower than the one of m-IDH patients as assessed by the Log-rank test (p<0.0001).

2.3.2 Protein expression in non-glioma tissue samples

Staining of glioma tissue samples with EGFR antibodies is very well characterised (Jeuken *et al.* 2009, Burel-Vandenbos *et al.* 2013, Sartori *et al.* 2017), which is considerably less the case with MMP9 and MUC4. Therefore, the expression of MMP9 and MUC4 was first tested in non-glioma brain tissues (n=2 per condition). The samples used were from normal brains (brain of patients who deceased from a cause unrelated to brain pathology) and brains from Alzheimer's disease patients, so that two different types of non-cancer brain tissues were tested. This decision was made to confirm the absence of MUC4 and MMP9 expression in normal brain and non-cancer brain, but also to ensure that the antibodies would give a negative staining in tissue with expected negative expression (negative control). In normal conditions, MMP9 is highly expressed in lymphocytes, bone marrow, and testis, while MUC4 is highly

expressed in the genital and gastrointestinal tracts, and respiratory system, and both show a very low to negative expression in the brain (*data from Human Protein Atlas*). The antibodies showed negative to very low expression for MMP9 and MUC4 in both normal and Alzheimer's Disease cells and surrounding matrix (**Fig.2.2**).

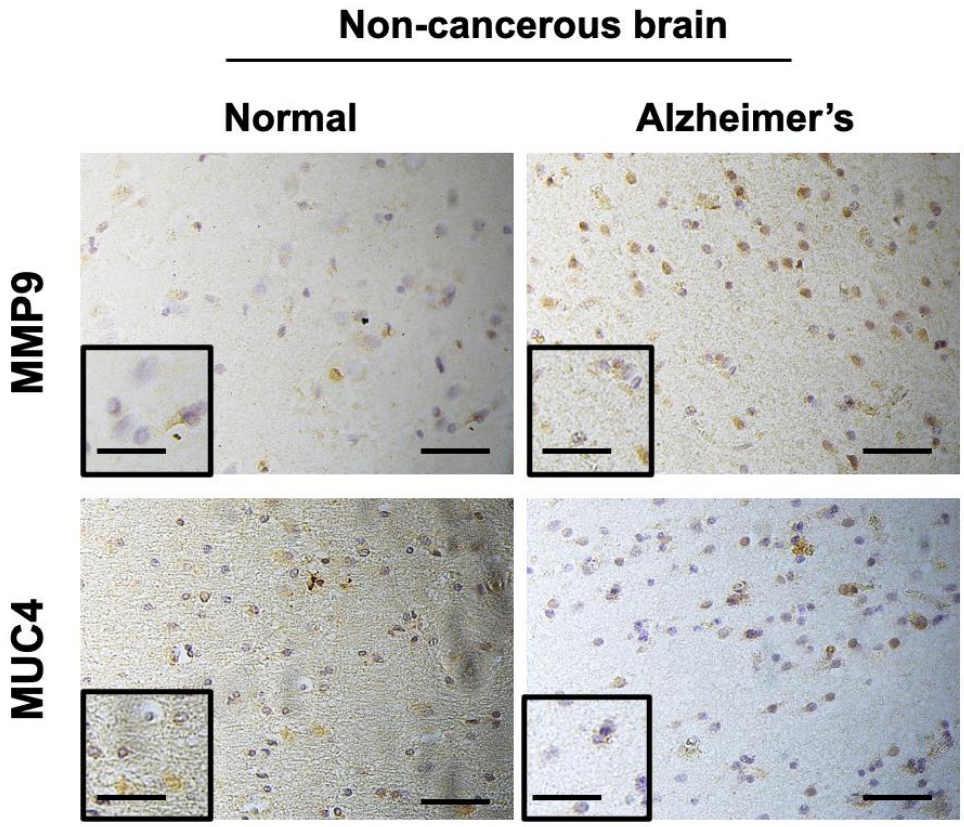


Figure 2.2 MMP9 and MUC4 expression in control non-glioma tissue samples. MMP9 and MUC4 protein expression was assessed by IHC in control tissue samples (normal and Alzheimer's disease brain). Both proteins showed very low to no expression in non-cancerous brain tissue. Total magnification, x400, scale bar = 50 μm, and 25 μm in zoomed panels (bottom left).

2.3.3 EGFR/MUC4/MMP9 tissue expression in glioma grades

EGFR, MMP9, and MUC4 expression levels were assessed in FFPE tissue biopsies from the 60 glioma patients (n=20 per grade). 1800 images were analysed in total. The tumours were categorised into three IHC scores (0, 1, and 2) for each protein, depending on both the proportion of expressing cells (0-30%, 30-70%, or 70-100%)

and the expression intensity (low, medium, or high) (**Fig.2.3A**). Examples of scored images are shown for each protein and for each score (**Fig.2.3B**).

A

		Intensity		
	IHC scores	**Low**	**Medium**	**High**
% of expressing cells	0-30%	0	1	1
	30-70%	1	1	2
	70-100%	1	2	2

B

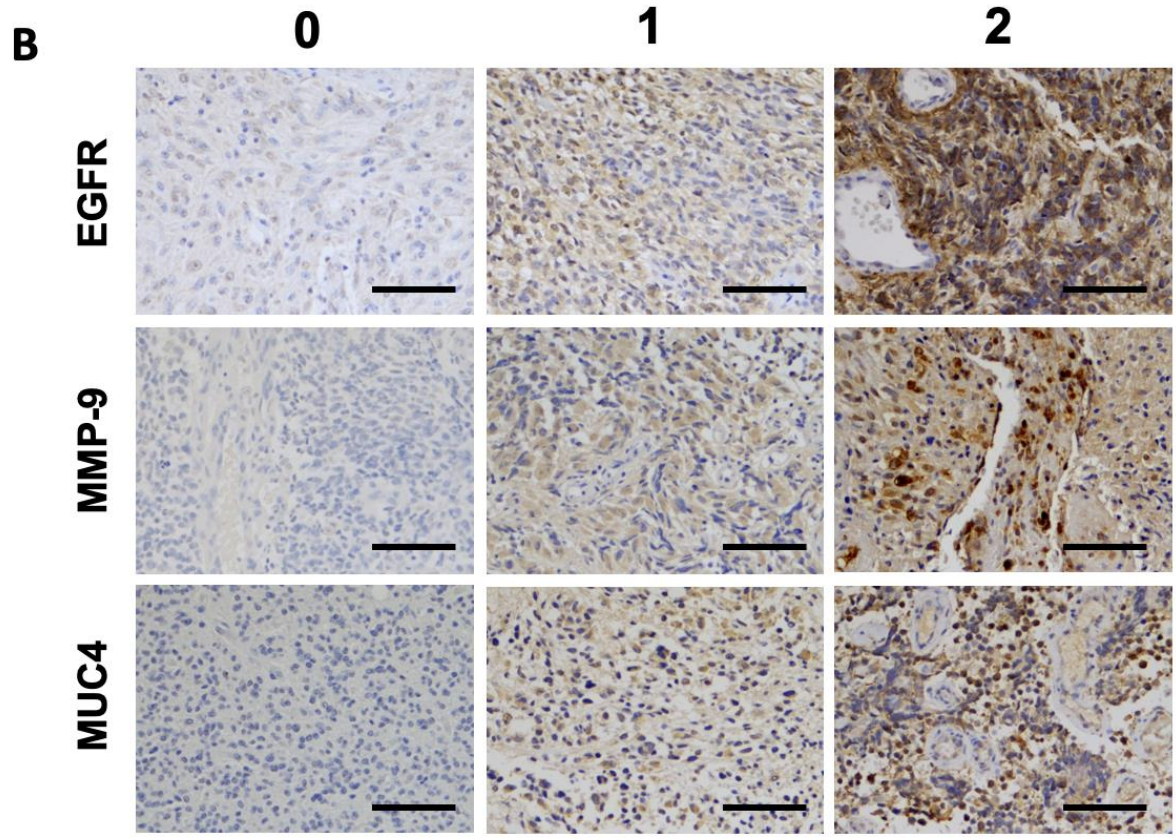


Figure 2.3 Tissue expression scoring system. IHC was performed on glioma tissue sections with EGFR, MMP9, and MUC4 antibodies. **(A)** The slides were categorised into three scores according to the percentage of area and intensity of staining. Score 2 represents the highest expression. **(B)** Examples of expression of EGFR, MUC4, MMP9 for each score are shown. Total magnification, x200, scale bar = 100 μm.

MMP9 and MUC4 tissue protein expression were significantly correlated with glioma histological grades. Specifically, MUC4 and MMP9 expression was higher in grade IV compared to grade III, and in grade III compared to grade II, and the association between the expression and the grades was significant for both proteins, as assessed by the Chi^2 test (p=0.0001 for MUC4 and p=0.0008 for MMP9) (**Fig.2.4A-B**). When comparing 'negative expression' (score 0) versus 'positive expression' (score 1 and 2 combined) to augment the number of samples per group, the results were similar, with the positive association between the grades and the protein expression significant for both MUC4 and MMP9 (p=0.0003 and p=0.0001, respectively) (**Table 2.5**). High expression of MUC4 (score 2) was observed in some grade II tumours. This was not the case for MMP9, which was totally absent from low grade cases (grade II). Among all grades, the number of patients with high expression was higher for MUC4 than for MMP9 (26.6% vs 11.6% in total). A high number of patients expressed EGFR (91.7%), and although the number of patients showing high expression was higher in the malignant grades (III and IV) than in the lower grade (II), no statistically significant differences were observed between any of the groups for this marker. 62% of high-grade glioma patients (grade III and IV) overexpressed EGFR (score 2).

A

	Number (%)			χ2 p-value
EGFR score	**0**	**1**	**2**	
II	3 (5)	10 (16.6)	7 (11.7)	0.323 (ns)
III	1 (1.6)	6 (10)	13 (21.6)	
IV	1 (1.6)	7 (11.7)	12 (20)	
MMP9 score	**0**	**1**	**2**	
II	15 (25)	5 (8.3)	0	0.0008 (***)
III	11 (18.3)	7 (11.7)	2 (3.3)	
IV	2 (3.3)	13 (21.7)	5 (8.3)	
MUC4 score	**0**	**1**	**2**	
II	14 (23.3)	4 (6.7)	2 (3.3)	0.0001 (***)
III	6 (10)	11 (18.3)	3 (5)	
IV	2 (3.3)	7 (11.7)	11 (18.3)	

B

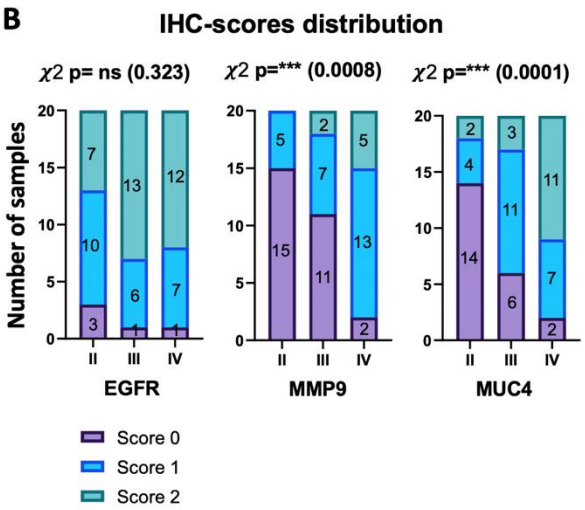


Figure 2.4 Association between IHC scores (tissue expression) and glioma histological grades. (A) Statistical differences were assessed with the Chi^2 test (95% confidence interval). For EGFR, there was no statistical significance (ns= non-significant). For MMP9 and MUC4, the rate of high expression significantly increased during glioma progression (higher in GBM compared to III and II). **(B)** For visualisation purpose, stacked bar graphs represent the distribution (contingency) of scores in grade II, grade III, and grade IV glioma for EGFR, MMP9, and MUC4 (N=60 cases; 20 per grade).

MUC4-MMP9 N(%)	0	1	2	$\chi 2$ p-value
II	16 (80)	2 (10)	2 (10)	0.0019 (**)
III	10 (50)	6 (30)	4 (20)	
IV	4 (20)	5 (25)	11 (55)	
MUC4-MMP9	**0 (no expression)**		**1-2 (expression)**	$\chi 2$ **p-value**
II	16 (80)		4 (20)	0.0007 (***)
III	10 (50)		10 (50)	
IV	4 (20)		16 (80)	
MUC4	**0 (no expression)**		**1-2 (expression)**	$\chi 2$ **p-value**
II	14 (70)		6 (30)	0.0003 (***)
III	6 (30)		14 (70)	
IV	2 (10)		18 (90)	
MMP9	**0 (no expression)**		**1-2 (expression)**	$\chi 2$ **p-value**
II	15 (75)		5 (25)	0.0001 (***)
III	11 (55)		9 (45)	
IV	2 (10)		18 (90)	

Table 2.5 Association between combined MMP9-MUC4 IHC-scores and glioma grades. Statistical differences were assessed with the Chi^2 test (95% confidence interval). The combined high expression of MMP9 and MUC4 was significantly and positively associated with the grades (p=0.0019; and p=0.0007 when comparing 'no expression' with 'expression'). MUC4 high expression and MMP9 high expression were significantly and positively associated with the grades when comparing 'no expression' with 'expression' status (p=0.0003 and 0.0001, respectively).

2.3.4 Correlation of protein expression with clinicopathological parameters

The clinicopathological features of the patient cohort were compared with the tissue expression for each protein separately (**Table 2.6**). Overall, there was no significant correlation between age or sex and the expression for any of the three proteins.

Since IDH1 genotype is crucial in the current cytogenetic classification, protein tissue expressions were also correlated with IDH1 genotype (mutant or wild type). For MMP9 and MUC4, but not EGFR, there was a significant association between high expression

and presence of wild type IDH1 and the significance was higher in the case of MUC4 (p=0.002 and p=0.0007, for MMP9 and MUC4 respectively). This observation was expected since wt-IDH, a marker predictive of worse survival in glioma, is strongly associated with higher histological grades (90% of wt-IDH are grade IV glioma in the 2016 WHO classification) (**Table 2.6**) (Cohen *et al.* 2013). When comparing 'no expression' (score 0) with 'expression' (score 1 and 2) to increase the number of patients per group and therefore the statistical power, the results were similar (p=0.001 for MUC4 and p=0.0002 for MMP9, Fisher's exact, data not shown). This result further confirms that MUC4 and MMP9 proteins are positively correlated with GBM (both histologically and cytogenetically) among glioma samples, and therefore with glioma progression and aggressiveness.

MMP-9	N (%)	IHC-scores			χ^2 p-value
		0	1	2	
Gender	**Female**	16 (26.6)	16 (26.6)	4 (6.6)	0.59 (ns)
	Male	12 (20)	9 (15)	3 (5)	
Age	**≤65**	22 (36.7)	30 (18)	5 (8.3)	0.84 (ns)
	>65	6 (10)	7 (11.7)	2 (3.3)	
IDH gene	**Mutated**	23 (38.3)	10 (16.6)	2 (3.3)	0.002 (**)
	Wild type	5 (8.3)	15 (25)	5 (8.3)	
MUC4	**N (%)**	**IHC-scores**			**χ^2 p-value**
		0	**1**	**2**	
Gender	**Female**	14 (23.3)	9 (15)	6 (10)	0.19 (ns)
	Male	8 (13.3)	13 (21.7)	10 (16.7)	
Age	**≤65**	17 (28.3)	17 (28.3)	11 (18.3)	0.8 (ns)
	>65	5 (8.3)	5 (8.3)	5 (8.3)	
IDH gene	**Mutated**	19 (31.6)	12 (20)	4 (6.6)	0.0007 (***)
	Wild type	3 (5)	10 (16.6)	12 (20)	
EGFR	**N (%)**	**IHC-scores**			**χ^2 p-value**
		0	**1**	**2**	
Gender	**Female**	5 (8.3)	11 (18.3)	13 (21.7)	0.05 (ns)
	Male	0	12 (20)	19 (31.7)	
Age	**≤65**	4 (6.7)	20 (33.3)	21 (35)	0.2 (ns)
	>65	1 (1.7)	3 (5)	11 (18.3)	
IDH gene	**Mutated**	4 (6.6)	14 (23.3)	17 (28.3)	0.5 (ns)
	Wild type	1 (1.6)	9 (15)	15 (25)	

Table 2.6 Association between tissue protein expression and patient characteristics. Statistical differences were assessed with the Chi^2 test (95% confidence interval). MMP9 and MUC4 lower expression was significantly associated with m-IDH patients (p=0.002 and 0.0007, respectively).

2.3.5 Localisation of EGFR, MUC4, and MMP9 proteins in GBM tissue

Next, the localisation of the protein expression was assessed in the tissue. In GBM patients, EGFR was variably expressed in the cytoplasm of neoplastic cells, ranging from moderate to strong expression, across the entire section (**Fig.2.5**, four examples shown). The expression was homogeneous among all neoplastic cells, but EGFR did not appear in endothelial cells of some visible microvascular proliferation (MVP) structures in all the samples (**Fig.2.6A**, four representative examples are shown). These structures can be identified easily as they contain a lumen that can be filled with red blood cells and/or small rounded structures that can be surrounded by several

layers of cells with a characteristic glomeruli-like shape. However, when analysed more thoroughly, the inner endothelial wall of some MVP structures was positive for EGFR with similar intensity as observed in the surrounding neoplastic tissue (**Fig.2.6B**, four representative examples are shown).

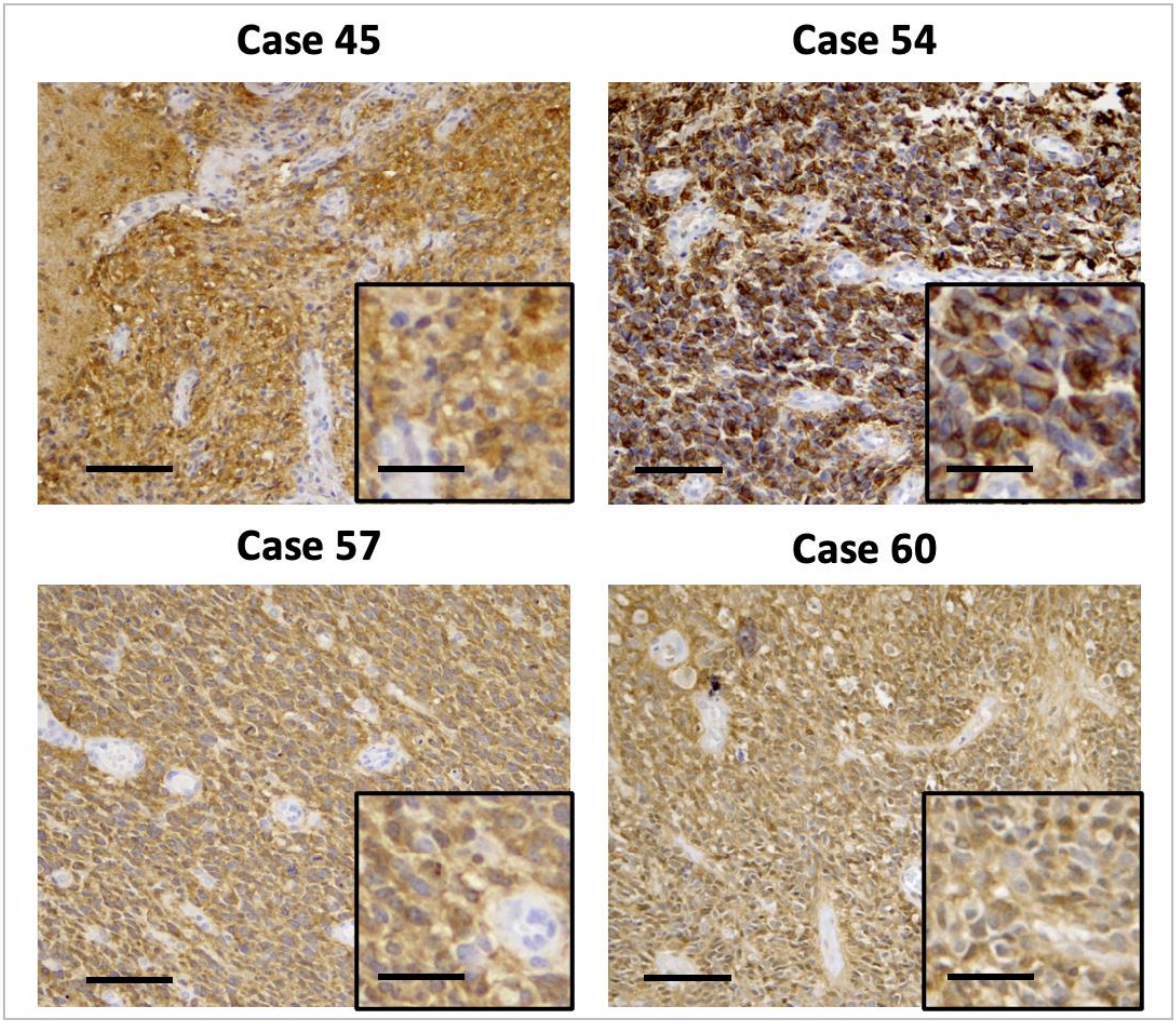


Figure 2.5 Expression of EGFR in GBM. EGFR expression was assessed in GBM tissue with IHC. Total magnification, x400, scale bar = 50 μm. Images are zoomed at the bottom right of each image, scale bar = 10 μm.

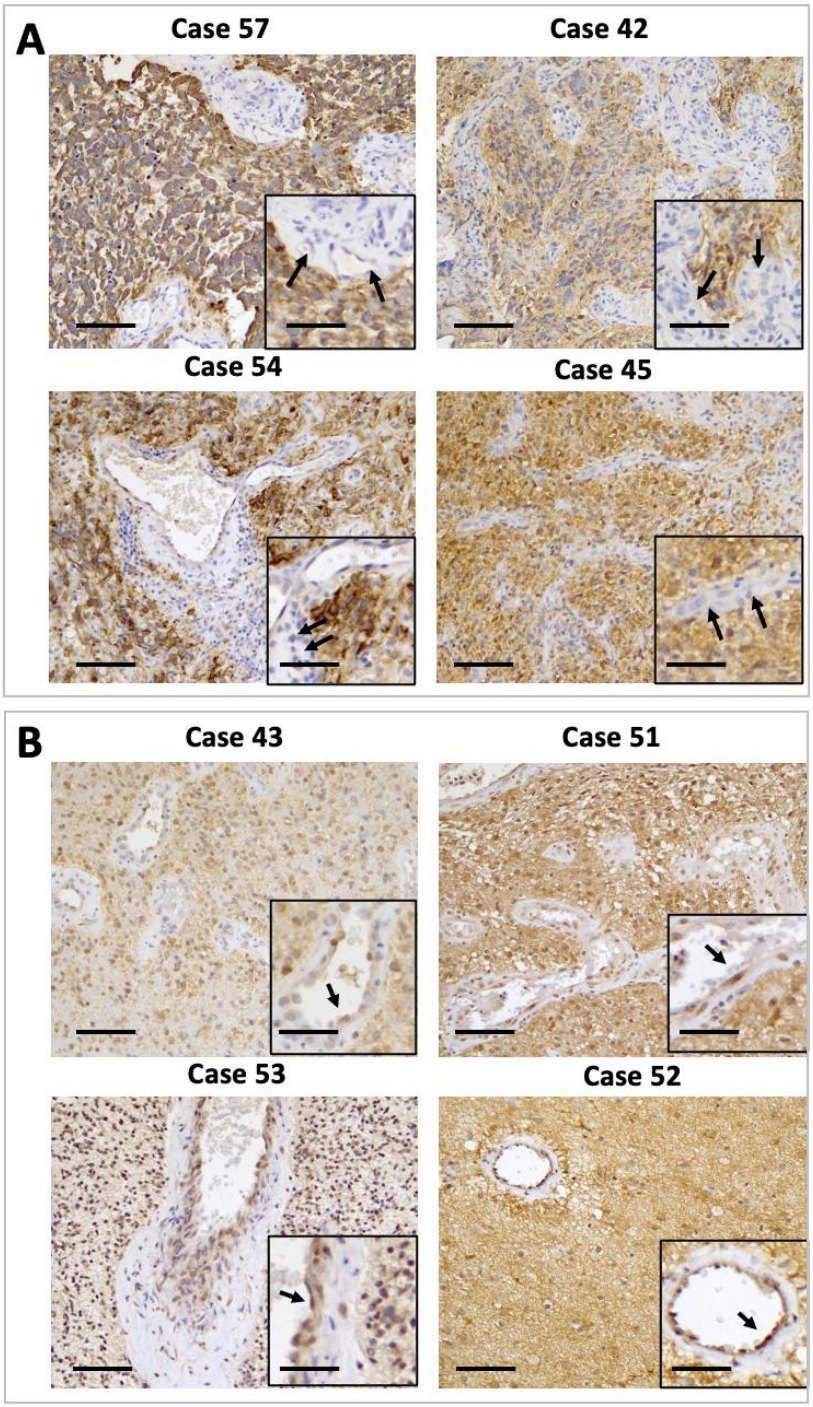


Figure 2.6 Expression of EGFR in GBM microvascular proliferation (MVP) structures. EGFR expression was assessed in GBM tissue with IHC. Total magnification, x200, scale bar = 100 μm. Images are zoomed at the bottom right of each image, scale bar = 20 μm. **(A)** EGFR was absent from visible microvascular proliferation (MVP) structures. **(B)** EGFR was expressed in endothelial cells in some patients (representative GBM samples shown). Black arrows indicate endothelial cells.

The expression of MMP9 in GBM patients was less dispersed and more heterogeneous than for EGFR, with cytoplasmic and nuclear expression showing in pleomorphic cells (cells highly heterogeneous in shapes and sizes and characteristic of GBM) (**Fig.2.7A**, four representative examples shown). The same observation was made for MUC4 (**Fig.2.7B**, four representative examples shown).

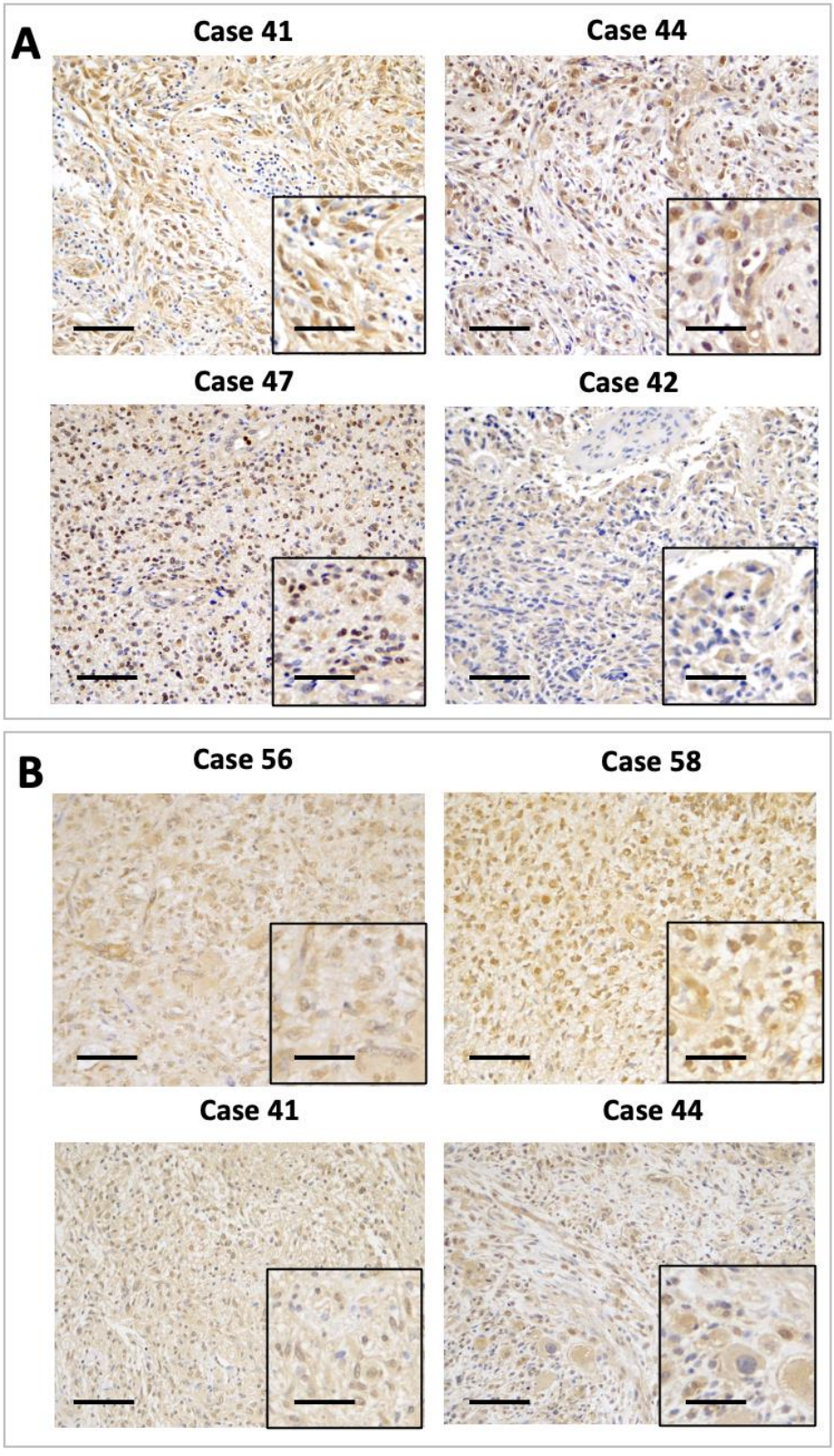


Figure 2.7 Expression of MMP9 and MUC4 in GBM. (A) MMP9 and **(B)** MUC4 were expressed in GBM tissue. Total magnification, x200, scale bar = 100 μm. Images are zoomed at the bottom right of each image, scale bar = 20 μm .

In addition, MMP9 was strongly and specifically expressed in nucleated cells located inside or near blood vessels, corresponding most probably to neutrophils or monocytes/lymphocytes, which is a normal localisation of MMP9 expression in healthy conditions (data from the *Human Protein Atlas*). The number of these MMP9-expressing cells increased in higher grades (**Fig.2.8**, two examples shown per grade). MMP9 was also strongly expressed in a small portion of nascent MVP structures in GBM (characterised by their glomeruli-like shape), which is in accordance with MMP9 implication in angiogenesis (**Fig.2.9**, four representative examples shown).

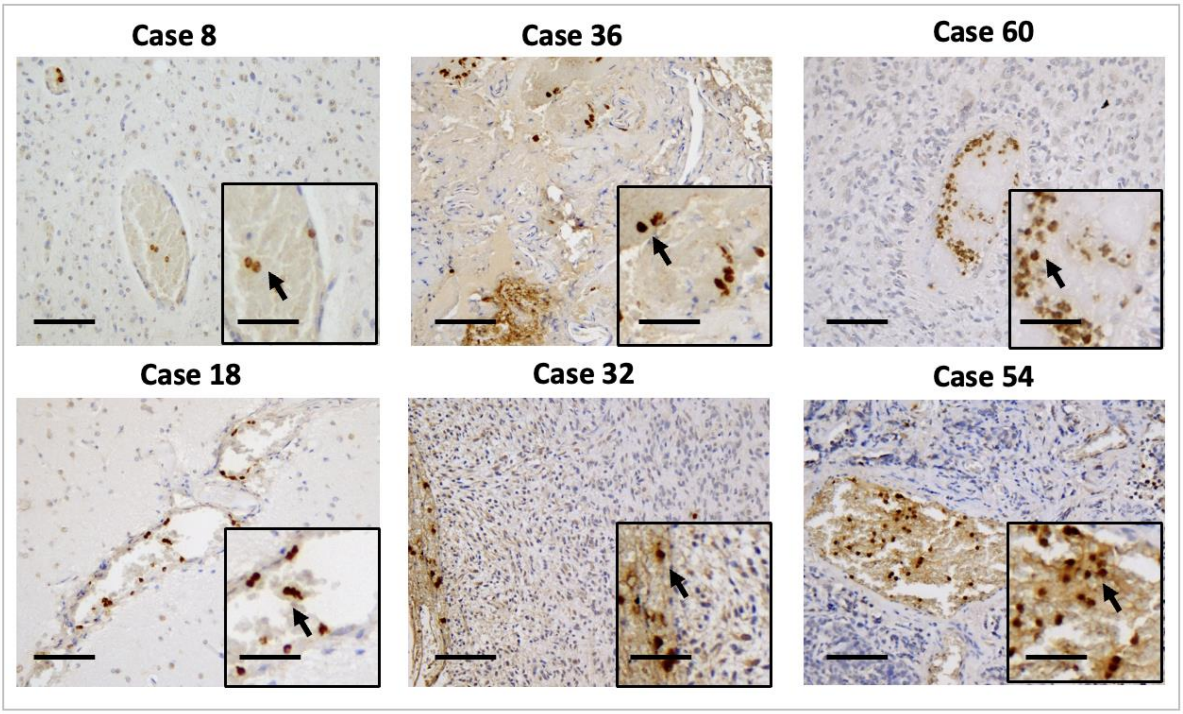


Figure 2.8 MMP9-expressing haematological cells in glioma. MMP9 was expressed in haematological cells inside vessels in glioma (indicated by black arrows) as assessed by IHC. The number of these cells increase in higher grades. Examples shown in grade II (case 8 and 18), grade III (case 36 and 32), and GBM (case 60 and 54). Total magnification, x200, scale bar = 100 μm . Images are zoomed at the bottom right of each image, scale bar = 20 μm

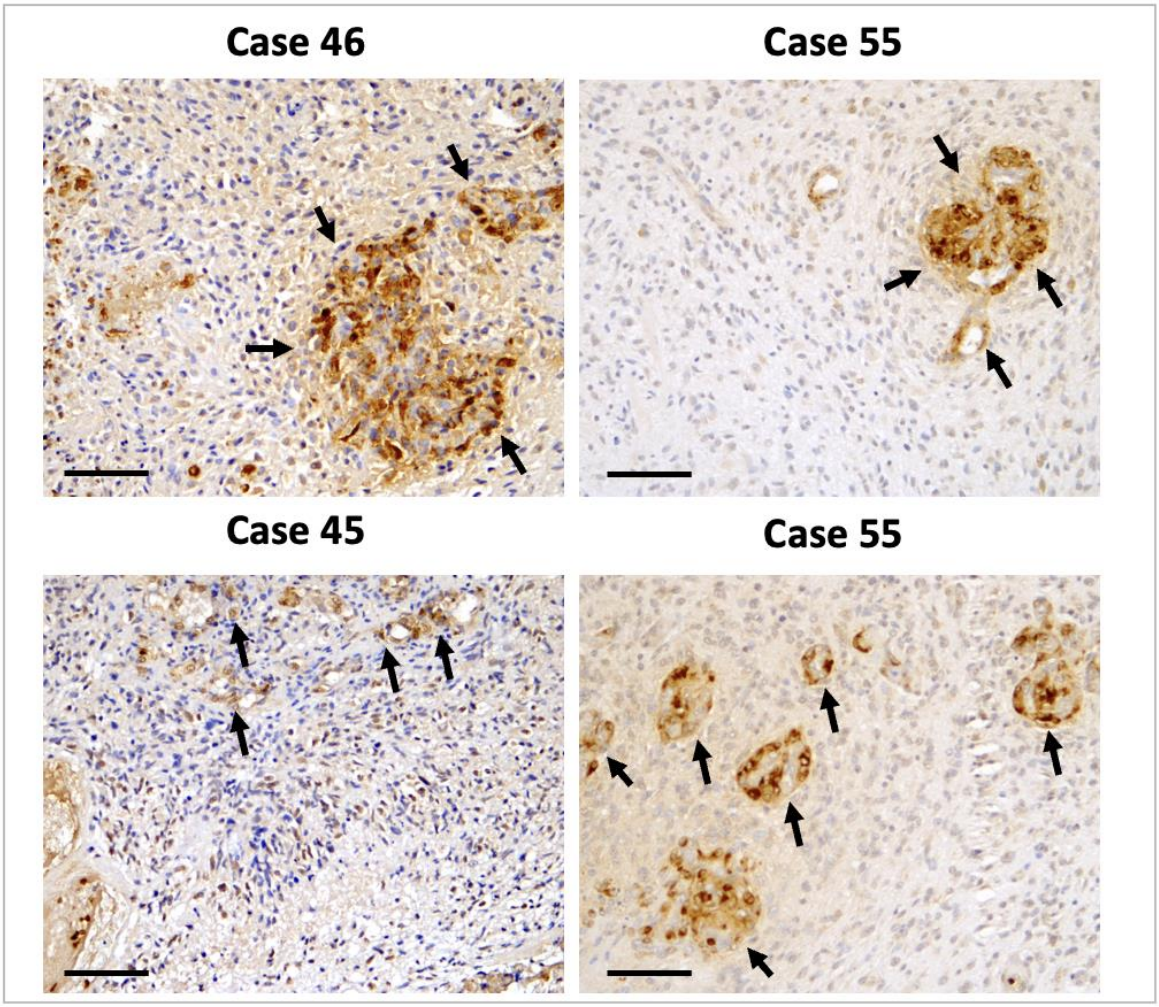


Figure 2.9 MMP9 expression in MVP structures in GBM. MMP9 was expressed in small microvascular proliferation (MVP) structures (indicated by black arrows) as assessed by IHC. Total magnification, x200, scale bar = 100 μm.

In GBM cases displaying the highest MUC4 expression, MUC4 was prominently expressed in cancer cells (**Fig.2.7B**). In GBM patients expressing MUC4, the protein was found in the cytoplasm of cancer cells, like EGFR. Strikingly, MUC4 was also undoubtedly expressed in the endothelial cells of the GBM-associated MVP structures (**Fig.2.10**, six examples shown).

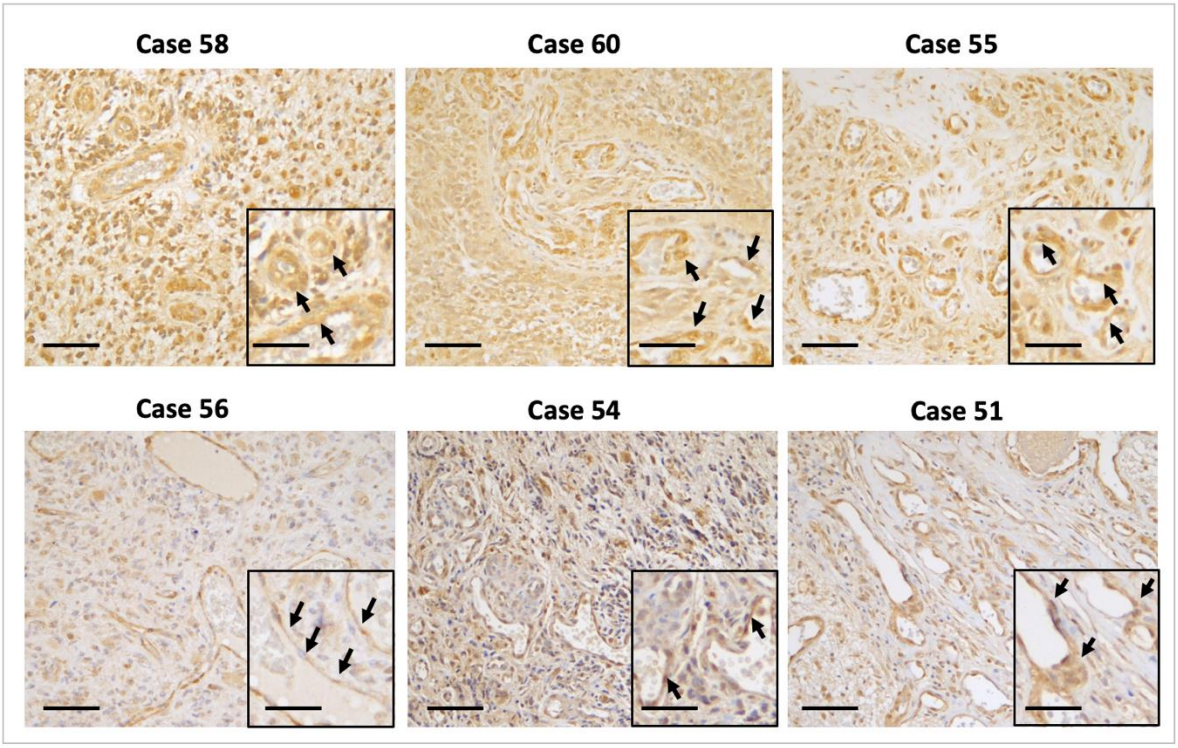


Figure 2.10 MUC4 expression in MVP areas in GBM. MUC4 was expressed in MVP structures (indicated by black arrows) as assessed by IHC. Total magnification, x200, scale bar = 100 μm . Images are zoomed at the bottom right of each image, scale bar = 20 μm.

2.3.6 MMP9, EGFR, and MUC4 co-localisation in MVP structures of GBM tissue

Firstly, to confirm that MMP9 and MUC4 were individually expressed in the MVP structures of GBM, double IF staining with the common pan-endothelial marker CD31, used to mark MVP in brain cancer (Peddinti *et al.* 2007), was performed in selected GBM patients. Patients showing obvious MMP9 or MUC4 expression in MVP structures from the IHC staining results were selected. The staining showed that CD31 was expressed in cellular structures that appeared to be MMP9 and MUC4-positive in these patients (**Fig.2.11A and B**, respectively). CD31 appeared to co-localise in merged images with MMP9 and MUC4, separately. This further confirmed that structures expressing either MMP9 or MUC4 indeed corresponded to CD31-positive regions (MVP structures).

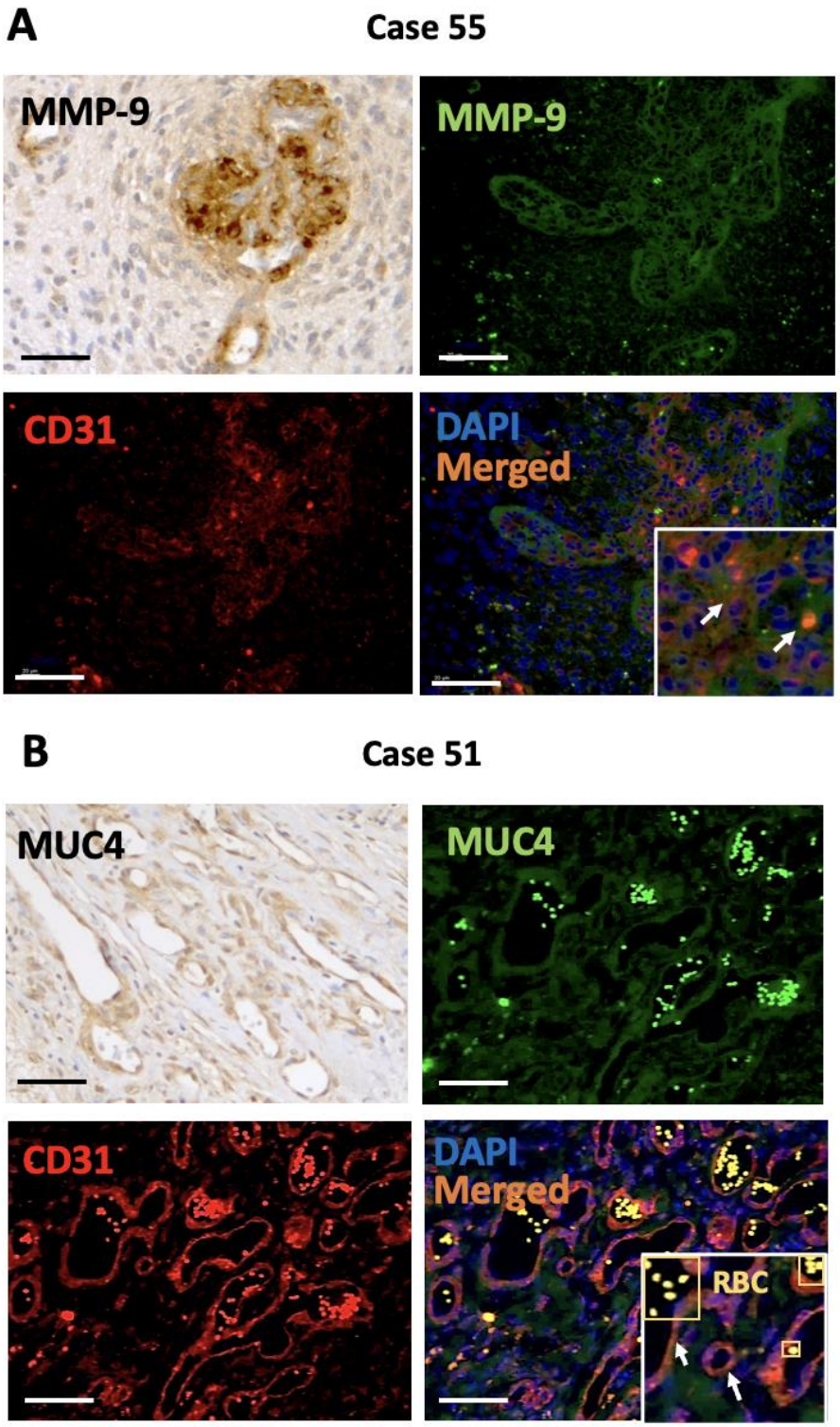


Figure 2.11 MMP9 and MUC4 expression in MVP structures in GBM. (A) MMP9 and **(A)** MUC4 **(B)** were expressed in vascular cells using IHC and co-localised with the MVP/angiogenesis marker CD31 in double IF, respectively. Red blood cells (RBC) emit natural strong fluorescence. White arrows indicate co-localisation. Total magnification, x200 for IF images and x400 for IHC images, scale bar = 100 μm. IHC and IF images are represented in parallel for reference only.

14 out of 20 GBM patients expressed MMP9 in the MVP structures, and 12 out of 20 expressed MUC4 in the MVP structures, while 8 out of 20 expressed both. To test whether MMP9 and MUC4 could be expressed together within the same MVP structures, some of these patients (showing both MMP9 and MUC4 expression in MVP structures from the IHC staining) were selected and analysed. Double IF staining revealed in these patients that MUC4 and MMP9 proteins were expressed together in the same MVP structures (**Fig.2.12A**).

MUC4 has also been shown to interact with EGFR family members in several cancer types (Carraway *et al.* 2009, Dhanisha *et al.* 2018). In the GBM samples, some cases also displayed a similar pattern of MUC4 and EGFR expression, with both proteins being expressed in MVP structures. Again, double IF staining revealed that MUC4 and EGFR proteins were expressed together in the same MVP structures in the selected patients (**Fig.2.12B**).

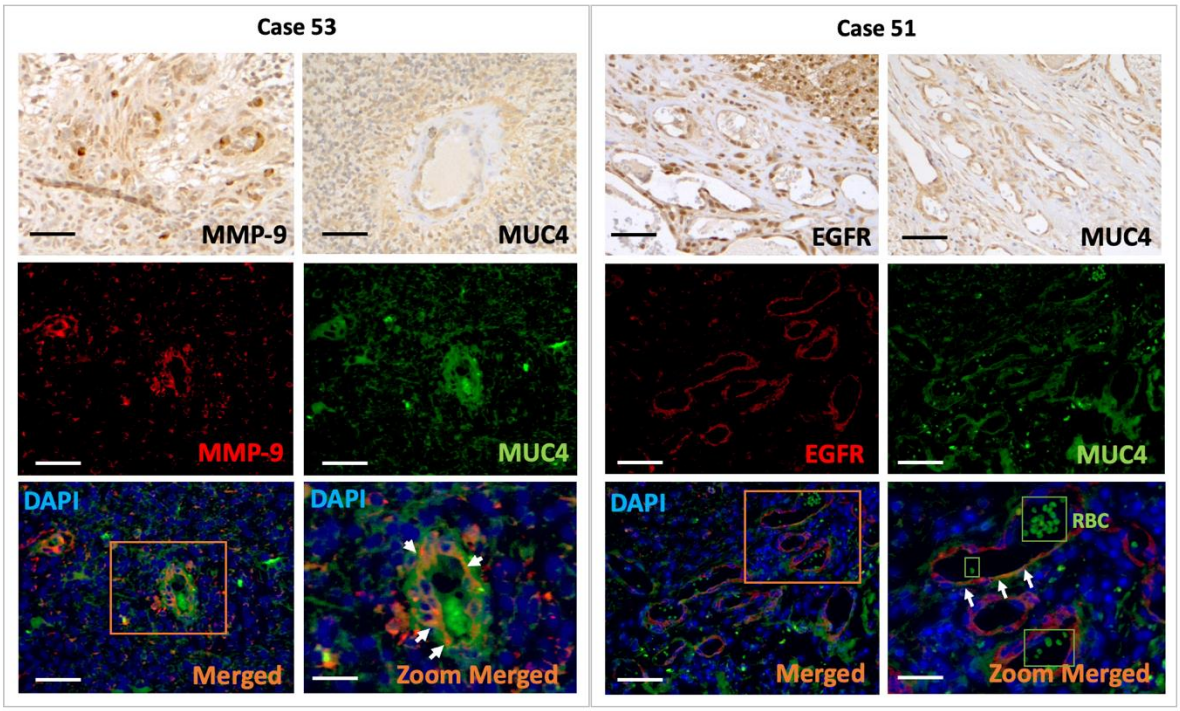


Figure 2.12 MUC4 co-localisation with MMP9 and EGFR in MVP structures in GBM. (A) MMP9 and MUC4 concomitant vascular expression in representative glioblastoma (case 53). Co-localisation of both proteins is observed in the vascular areas with double IF. **(B)** EGFR and MUC4 concomitant expression in vascular areas in representative case (case 51) of GBM observed by IHC and IF. Co-localisation of both proteins is observed with double IF. White arrows indicate co-localisation. Total magnification, x200 for IF images and x400 for IHC images, scale bar = 100 μm and 50 μm in zoom merged images.

2.3.7 Association between tissue expression and patient survival

Next, the clinical outcome was compared between patients with a low level of protein tissue expression and patients with a high level of protein tissue expression, using the Kaplan-Meier survival analysis method. This aim of that method is to assess the impact of the protein tissue expression on the patients' survival. This type of analysis can help determining whether these proteins can be used as prognostic biomarkers. The clinical outcome was evaluated by looking at two different variables: the overall survival (OS), defined by the number of months between surgery and death (if any at the time of analysis), and, the progression-free survival (PFS), defined by the number of months between surgery and relapse of the disease (if any). The three IHC-scores were first categorised into two groups: High and Low expression, and the clinical outcome was compared for each protein.

When investigated individually, MMP9-High and MUC4-High expression groups tended to indicate worse prognosis than MMP9-Low and MUC4-Low groups, respectively, for both OS and PFS. However, the difference was not significant (OS: p=0.105 for MMP9; p=0.136 for MUC4, PFS: p=0.35 for MMP9; p=0.36 for MUC4, Log-rank) (**Fig.2.13A-D**).

For EGFR, no difference was observed between the two groups for both OS and PFS and the two survival curves overlapped in each case (OS: p=0.63, PFS: p=0.6, Log-rank) (**Fig.2.13E-F**).

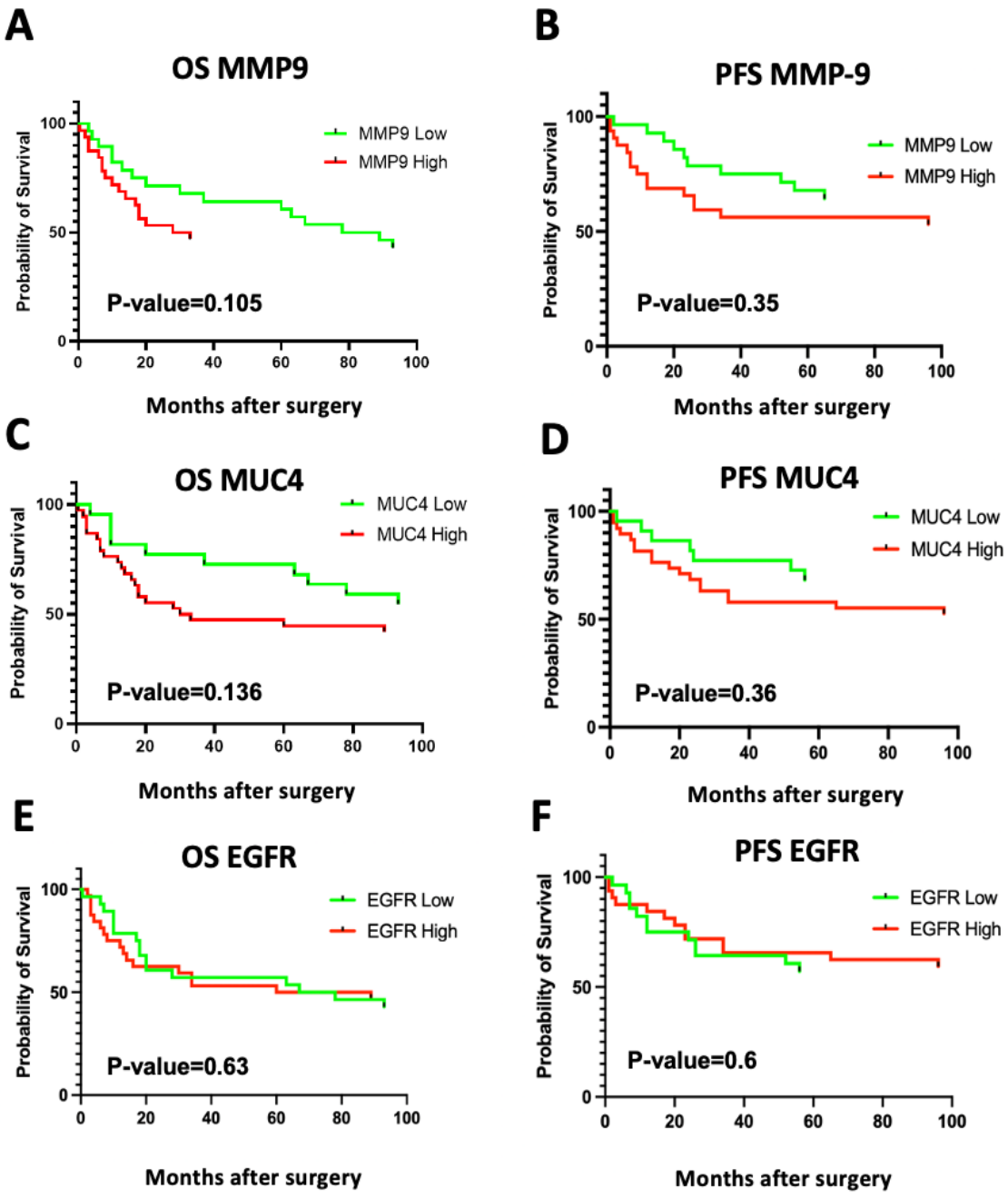


Figure 2.13 Association between patient survival and protein tissue expression. Comparison of overall survival (OS) and progression-free survival (PFS) in all glioma patients using the Kaplan-Meier method, between two tissue expression groups (low and high) for **(A-B)** MMP9, **(C-D)** MUC4, and **(E-F)** EGFR. **(A-D)** High expression of MMP9 or MUC4 predicts worse survival (OS and PFS), but the difference was not significant. **(E-F)** There was no visible difference for EGFR. p-values are indicated on the graphs.

The protein expression was next combined and correlation with clinical outcomes was made. First, clinical outcomes were investigated by combining MMP9 and MUC4 proteins without EGFR (**Fig.2.14A-D**). For MMP9 and MUC4 combined, two different approaches were tested with one by generating three groups (**Fig.2.14A-B**) of combined expressions and one generating four groups of combined expressions (**Fig.2.14C-D**). First, three groups were generated: one containing patients with both MMP9 and MUC4 proteins with low tissue expression (L-L), one with both proteins at high expression (H-H), and one with one of the proteins only at high expression and the other one at low expression (L-H and H-L). With this approach, the L-L group appeared to predict better outcome for both OS and PFS than the two other groups. The difference between the L-L curve and the H-H curve was more important than in the survival analysis looking at MMP9 and MUC4 individually. For the OS, the difference was significant (p=0.023, Log-rank) between the L-L groups and the H-H group, but not for PFS (p=0.27, Log-rank) (**Fig.2.14A-B**).

In the second approach, four groups were generated for the pair of proteins containing the different combinations of expression: Low MMP9 expression-Low MUC4 expression (L-L), High-High (H-H), Low-High (L-H), and High-Low (H-L). Overall, this approach did not change the results since patients displaying MUC4 low expression with a high MMP9 expression were very scarce [n=2 (3.3%)]. Overall, combining MMP9 with MUC4 was more useful to predict the outcome than looking at the protein individually, since the difference between the L-L group and the H-H group was significant. The intermediate combinations, corresponding to the patients that express the two proteins at different levels, displayed an intermediate outcome (**Fig.2.14C-D**).

Next, all three proteins were combined, and four groups were generated: ‘triple low’ contained patients with low expression for the three proteins; ‘triple high’ patients with high expression for the three proteins; ‘double low’ patients with low expression for two proteins out of three; and ‘double high’ patients with high expression for two proteins (**Fig.2.14E-F**). The MUC4/MMP9/EGFR combined low expression significantly predicted better survival in glioma patients in comparison with the ‘double high’ group and the difference was again more significant than the previous

analysis that combined MMP9 and MUC4 only (without EGFR) (OS: p=0.004, PFS: p=0.087, Log-rank). Therefore, adding EGFR was useful to improve the accuracy of outcome prediction even though there was no prognostic significance for this protein when used alone. The biggest difference was observed between ‘triple low’ and ‘double high’ and not between ‘triple low’ and ‘triple high’ (p=0.15 for OS). However, this was likely due to the smaller size of ‘triple high’ (n=8) in comparison with the ‘double high’ group (n=15) (**Fig.2.14E**).

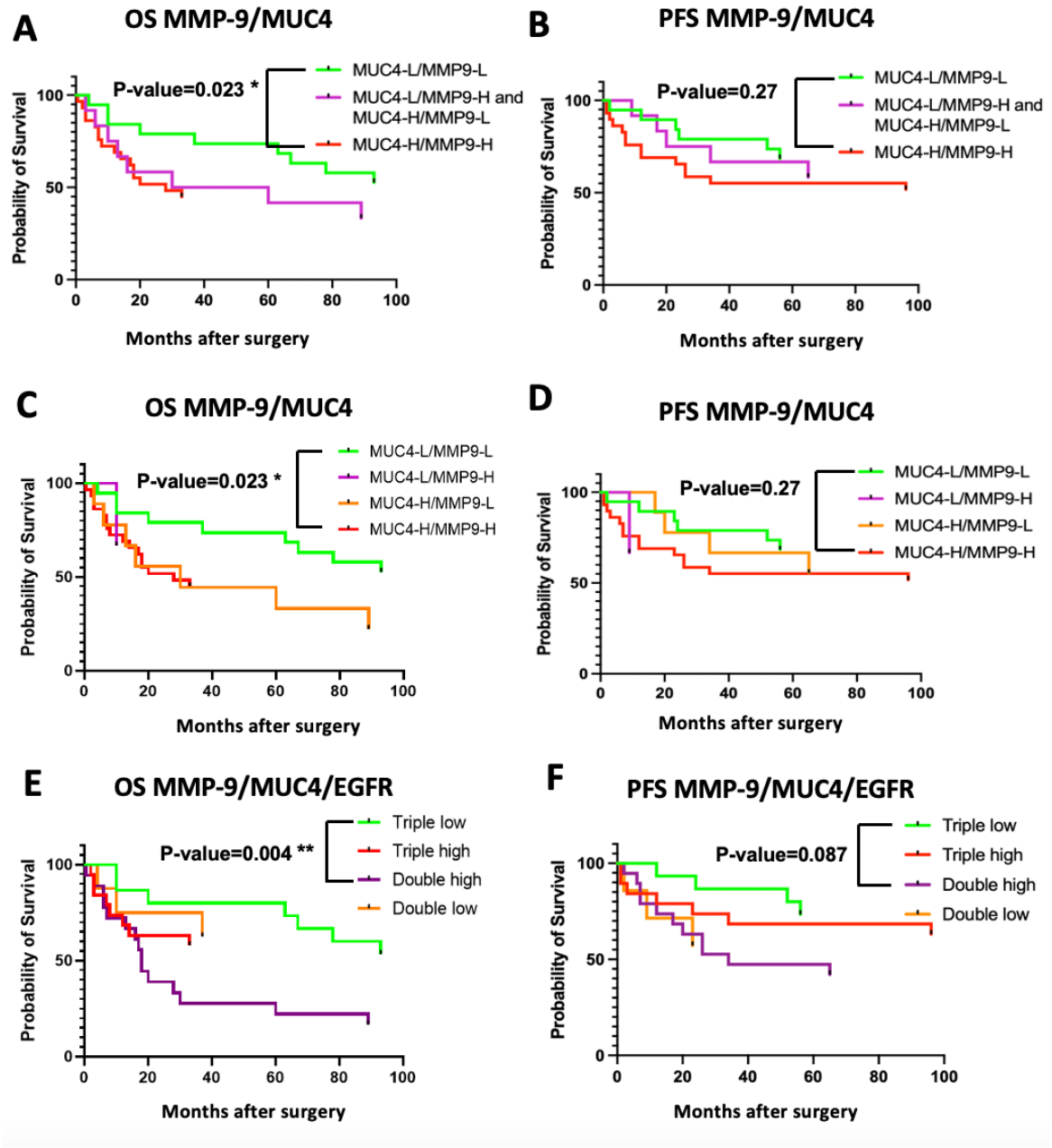


Figure 2.14 Association between patient survival and MUC4/MMP9 or MUC4/MMP9/EGFR combined tissue protein expression. Comparison of OS and PFS in all glioma patients using the Kaplan-Meier method, between **(A-B)** three tissue expression groups combining MMP9 and MUC4 expression, **(C-D)** four expression groups combining MMP9 and MUC4 expression, and **(E-F)** four expression groups combining MMP9, MUC4, and EGFR. **(A-B)** MMP9-MUC4 combined high expression predicted worse prognosis than MMP9-MUC4 combined low expression, which was significant for the OS as assessed by the Log-rank test (p=0.023). **(C-D)** The same results were observed when using four expression groups. **(E-F)** The combined high expression of MMP9, MUC4, and EGFR predicted

worse prognosis than the group containing high expression for two proteins only (double high), which was significant for OS as (p=0.004). The combination of the three proteins gave more significance than combination of MMP9 and MUC4 only (p=0.004 vs p=0.023 respectively).

2.3.8 Patient characteristics of blood serum samples

Since blood samples are collected more easily and less invasively than solid biopsies, liquid biomarkers are helpful for disease monitoring at different time points. Therefore, the protein levels of MMP9 and MUC4, were next assessed in blood serum samples from glioma patients (grade III and IV, n=10 each) and from non-glioma control individuals (n=10) to test their potential use as liquid biomarkers. Patient characteristics are listed in **Table 2.7**. Patients 21 to 29 (grade III) and 41 to 50 (all grade IV serum patients) corresponded to the same patients whose tissue biopsies were used previously, while patients 61 to 63 (grade III) and non-glioma control individuals were added to the study.

Case	Grade	Tumour type	Sex	Age (yrs) at surgery	IDH genotype	OS (months)	PFS (months)	Treatment pre-surgery
21	III	Tumour Progression to an Anaplastic Oligodendroglioma	M	53	m	30	14	Radiotherapy, PCV, and TMZ
22	III	Anaplastic Oligodendroglioma	M	53	m	93		
23	III	Anaplastic Oligodendroglioma	F	36	m		73	
26	III	Anaplastic Oligodendroglioma	M	41	m	37	23	
27	III	Anaplastic Oligodendroglioma	F	42	WT	28	25	
28	III	Anaplastic Oligodendroglioma	F	56	m	33	7	PCV and TMZ
29	III	Mutant Anaplastic Oligodendroglioma	F	25	m			
61	III	Recurrent Glioma (Anaplastic Astrocytoma)	M	32	m	46	24	
62	III	Anaplastic Astrocytoma	M	28	WT			
63	III	Anaplastic Astrocytoma	F	62	WT		33	
41	IV	Glioblastoma	F	64	WT	18	7	
42	IV	Glioblastoma	M	58	WT		34	
43	IV	Glioblastoma	M	56	WT	10	9	
44	IV	Glioblastoma	M	64	WT	0		
45	IV	Glioblastoma	F	52	WT	3	3	
46	IV	Glioblastoma	F	73	WT	7		
47	IV	Glioblastoma	F	74	WT			
48	IV	Glioblastoma	M	69	WT	18	12	
49	IV	Glioblastoma	F	41	WT	17	7	
50	IV	Glioblastoma	F	70	WT	6		
64	CTRL	Uterus: Sections of the fibroids from the myometrium show features of benign leiomyomas	F	50				
65	CTRL	Benign serous cyst or hyrosalpinx	F	38				
66	CTRL	Ovary: Paratubal/Para-ovarian Benign Cyst, Uterus: Benign Leiomyoma	F	44				
67	CTRL	Benign Mucinous Cystadenoma	F	79				
68	CTRL	Benign Leiomyoma	F	73		Non-glioma CTRL		
69	CTRL	Benign Leiomyoma	F	80				
70	CTRL	Benign Hyperplasia, no evidence of malignancy	M	76				
71	CTRL	Benign Serous Cystadenofibroma	F	83				
72	CTRL	Benign Leydig Cell Tumour	M	45				
73	CTRL	Right Lateral Wall of Orbit: Benign Meningioma	F	52				

Table 2.7 Glioma and control serum samples, patient characteristics. [Years (yrs), Overall survival (OS), Radiotherapy (XRT), procarbazine, lomustine, vincristine (PCV), temozolomide (TMZ), mutated (m), wild-type (WT), controls (CTRL)].

The mean age was 43 years-old in the grade III group and 62 years-old, both in the grade IV and the control group. In total, the grade III group contained four wt-IDH1 patients and six m-IDH1 patients and the grade IV group contained wt-IDH1 patients only. Among the patients who had deceased, the mean OS was 44.5 months in the grade III group and 10 months in grade IV group. Among the patients who had relapsed, the mean PFS was 28.4 months and 12 months in grade III and IV, respectively.

2.3.9 Protein levels of MMP9 and MUC4 in glioma serum

Next, the potential use of the MMP9 and MUC4 proteins as serum biomarkers for discrimination between grade III and IV glioma was investigated. Since it was previously found in this study that these proteins were upregulated in glioma, and especially in GBM, their level in the serum of 30 patients (10 benign non-glioma

tumours (controls), 10 grade III, and 10 GBM) were further investigated with ELISA immunoassays.

First, the total amount of protein was measured in each serum sample using the Bradford assay method. The mean total protein concentration was 57.25 mg/ml in the control group, 59.1 mg/ml in the grade III group, and 65.63 mg/ml in the grade IV group. Then, the concentrations of MMP9 free form and TIMP1-bound form, as well as MUC4, were measured in each sample with ELISA immunoassays. Since the concentration of total protein tended to be higher in the glioma grades, even though the difference was not significant after unpaired multiple t-tests, the concentration for the proteins was normalised to 1 mg of total protein for each sample to avoid any bias.

The ELISA immunoassay for MMP9 could detect two forms of MMP9 that can be found in the blood (the 92 KDa pro-form and the 82 KDa activated form). The mean MMP9 concentration in the control group was 4.92 ng/mg of total protein. The mean MMP9 concentration in the grade III and grade IV group was 3.34 and 6.1 ng/mg, respectively. The structures of the two MMP9 forms and the scatter graph of the normalised values with the mean are represented in **Figure 2.15A**. Overall, MMP9 protein levels were more elevated in GBM compared to both grade III and control samples. The increase of mean MMP9 protein level in GBM compared to grade III was significant ($p=0.022$, Tukey's multiple comparison), but the difference with the control was not significant ($p=0.25$ for grade III and 0.46 for grade IV). Moreover, the inter-patient heterogeneity was high in the control group, which was not the case in the grade III group. This could be explained by the fact that the control group was composed of patients affected by different benign non-glioma tumours rather than healthy individuals.

Since MMP9 may also be bound to its inhibitor TIMP1, and this complex can also have a potential diagnostic value (Marcus *et al.* 2010), the level of the MMP9:TIMP1 complex in the same samples was also measured. The MMP9:TIMP1 complex concentration was respectively 0.28, 0.12, and 0.52 ng/mg of total protein in the control group, grade III group, and grade IV group. The structures of the two forms and the scatter graph of the normalised values with the mean are represented in

Figure 2.15B. The trend for MMP9:TIMP1 complex was similar to the free MMP9 form, with a significant increase of MMP9:TIMP1 complex mean protein level in GBM compared to grade III (p=0.018, Tukey's multiple comparison), but no significant difference with the control (p=0.47 for grade III and p=0.2 for grade IV), which again displayed higher heterogeneity compared with the grade III group. The percentage of MMP9 protein bound to its inhibitor was low in all cases compared to the MMP9 free forms, with ~5% of MMP9 bound to TIMP-1 in GBM on average, twice than what was observed in controls and grade III (2.7% and 2.4%, respectively).

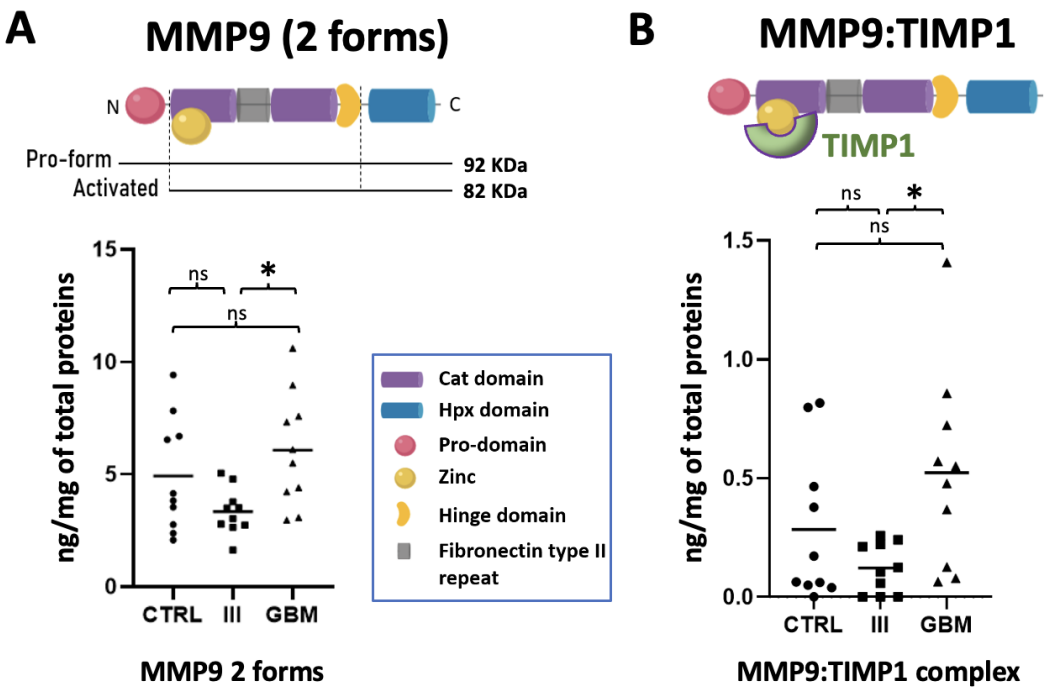


Figure 2.15 MMP9 protein levels in serum. Scatter graphs showing the protein levels of **(A)** two forms of MMP9 (MMP9 pro and active forms) and **(B)** MMP9 bound to TIMP1 in the serum of non-glioma (CTRL: controls), glioma grade (III), and glioblastoma (GBM) patients. N = 10 for each group (30 patients in total). The structures of the different MMP9 forms measured are represented. Protein levels were normalised based on the total protein content; results are given as ng per mg of total protein. MMP9 two forms and MMP9:TIMP1 complex mean protein level was significantly higher in the GBM group than in grade III group (Tukey's multiple comparison test, p=0.022 and 0.018, respectively). The bars represent the means.

Finally, mean MUC4 serum protein levels tended to be low overall, and higher in the malignant glioma grades (0.014, 0.068, and 0.030 ng/mg). The structure of the MUC4 and the scatter graph of the normalised values with the mean are represented in **Figure 2.16**. The mean protein level of MUC4 was very low in all cases and was only slightly more elevated in glioma compared with the controls and the difference had no statistical significance (p= 0.42 and 0.92 for controls versus grade III and IV, respectively, and p=0.64 for grade III versus grade IV, Tukey's multiple comparison).

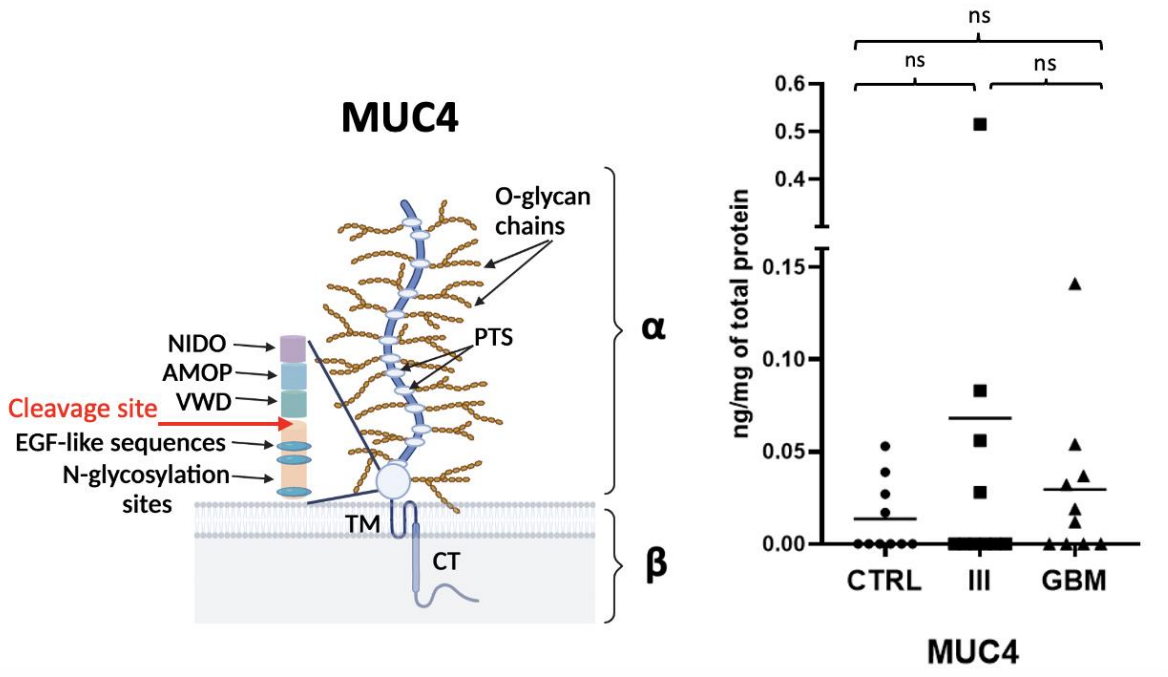


Figure 2.16 MUC4 protein levels in serum. MUC4 structure and scatter graphs showing MUC4 protein levels. The values are represented on a stretched Y axis. Protein levels were normalised based on the total protein content; results are given as ng per mg of total protein. Higher concentrations of MUC4 are found in malignant glioma grades compared with controls but the difference does not reach significance (Tukey's multiple comparison test). [Nidogen-like domain (NIDO), Adhesion-associated domain (AMOP), and Von Willebrand factor Domain (VWD)].

2.4 Discussion

2.4.1 MMP9 and MUC4 overexpression in glioma

MUC4 is involved in carcinogenesis and suggested as a potential biomarker for several cancers with epithelial origin including pancreatic (Kufe 2009), lung (Tsutsumida *et al.* 2007), gallbladder (Lee *et al.* 2012), and oral squamous cell carcinoma (Hamada *et al.* 2012). It was also demonstrated that MUC4 is overexpressed in GBM solid biopsies and cell lines (Li *et al.* 2014). The interaction between MUC4 and EGFR at the surface of the cells is believed to contribute to carcinogenesis (Carraway *et al.* 2009, Dhanisha *et al.* 2018) and expression of MMP9, whose involvement in glioma progression is well characterised (Choe *et al.* 2002, Chang *et al.* 2015, Chen *et al.* 2017, Quesnel *et al.* 2020). For these reasons, it was hypothesised that these three proteins together should be assessed to further elucidate their individual diagnostic value and to investigate their combined effect that may improve glioma diagnosis and prognosis. With that aim, the expression of MUC4, EGFR, and MMP9 in glioma biopsies of different histological grades was investigated. This study showed that MMP9 and MUC4, but not EGFR, tissue expression was correlated with increased glioma grades and wt-IDH1 genotype, which is used in the most recent glioma classification to define GBM. Overall, MMP9 has been shown to be a significant potential prognostic and diagnostic factor in glioma by several previous studies (Choe *et al.* 2002, Zhao *et al.* 2012, Liu *et al.* 2015, Xue *et al.* 2017). In the present study, MUC4 was at least as helpful as MMP9 for grade discrimination. Although further investigation is needed, this entails that it could serve as a potential biomarker in glioma diagnosis.

2.4.2 EGFR expression in GBM

EGFR has been shown to be overexpressed, amplified and/or mutated in GBM, and to a lesser extent, in all the other glioma grades. Its prognostic and diagnostic value as well as its therapeutic targeting have led to different and contradictory results (Saadeh *et al.* 2018). Even though its involvement in glioma development is undeniable, its relevance as a biomarker for glioma remains contestable. Here, it was found that EGFR was expressed in most of the samples (91,7%) and overexpressed in ~60% of the GBM patients, which certainly explains why it fails at discriminating between glioma grades. EGFR is the most commonly amplified gene

in GBM and is overexpressed in around 60% of primary GBM according to other studies, which is in total agreement with the findings of the current study (Karpel-Massler *et al.* 2009). These results imply that EGFR tissue expression, when examined individually, is not a powerful biomarker in glioma diagnosis. However, looking at EGFR amplification or the expression of specific EGFR variants rather than its overall expression could yield different results.

In the present study, it was found that EGFR was highly expressed in the cytoplasm of the neoplastic cells while being clearly absent from a high number of visible MVP structures in IHC staining. It has been shown previously *in vitro* and *in vivo* that GBM cells with stem cell abilities can transdifferentiate into endothelial-like cells that could contribute to the tumour microvasculature. These cells of neoplastic origin are not endothelial cells as they are negative for CD34 (Soda *et al.* 2011). Although they cannot generate an endothelium, they can form structures resembling vessels, a process which has been designated as 'vasculogenic mimicry'. It has been suggested that this process could exist in GBM patients and may be responsible for resistance to anti-VEGF treatments (Soda *et al.* 2011), although, this has not been clearly demonstrated in clinics so far. In the proliferative microvasculature of GBM tumours, cells carrying molecular alteration characteristics of GBM, like an EGFR overexpression, are not found by using IHC (Rodriguez *et al.* 2012). This agrees with the present study since even in GBM patients showing high EGFR expression, it appeared that there was no EGFR expression within the visible MVP structures. However, the IF revealed EGFR expression in some of these structures and when looking back at the IHC staining, some vessels appeared to be indeed positive for EGFR. This suggests that the majority of MVP structures could emerge from existing endothelial cells (EGFR-negative) while others could be the results of vascular mimicry from neoplastic cells (EGFR-positive), the latter being hard to identify in IHC when the omnipresent and strong expression of EGFR masks the vasculature shape. Recently, a study revealed that vascular mimicry is more frequent in samples from GBM patients with radiotherapy history. This study does not consider these parameters and most patients had not been treated with radiotherapy as standard treatment for GBM is typically performed after the surgical resection. This could explain why the phenomenon is not observed in abundance in the tissue biopsies (De Pascalis *et al.* 2018).

2.4.3 MMP9 expression in GBM

In this study, it was found that MMP9 displayed a specific expression and was not systematically expressed across the tumour. For example, MMP9 was strongly expressed specifically in cells that have a haematological origin and were located inside vessels. Given their shape and their high number, these cells most probably correspond to immune cells such as neutrophils, or monocytes/lymphocytes, which is a healthy location for MMP9 expression according to the Human Protein Atlas. MMP9 is involved in neutrophil migration across the basement membrane (Delclaux *et al.* 1996) and is expressed by neutrophils contributing to the inflammatory response (Bradley *et al.* 2012, Medeiros *et al.* 2017). Neutrophil-derived MMP9 has also been shown to play a role in vessel remodelling/angiogenesis during tissue regeneration (Heissig *et al.* 2010) and cancer progression (Acuff *et al.* 2006, Nozawa *et al.* 2006, Bekes *et al.* 2011). Specifically, it has been demonstrated that MMP9 released from neutrophils is in its free form (not bound to its inhibitor TIMP-1) and is therefore an active and very potent angiogenic mediator (Ardi *et al.* 2007, Bekes *et al.* 2011). Given the importance of angiogenesis in glioma progression and the involvement of MMP9 in this process, it is reasonable to think that what is observed in the investigated samples might correspond to MMP9-expressing neutrophils increasing during glioma progression. This is in accordance with findings in other types of cancer (Acuff *et al.* 2006, Nozawa *et al.* 2006, Bekes *et al.* 2011). Interestingly, a recent study observed that MMP9 was mainly expressed in MVP structures, neutrophils, and activated dendritic cells in GBM at initial diagnosis (Jiguet-Jiglaire *et al.* 2022).

Strong expression of MMP9 was found in cells composing some of the MVP structures in GBM samples, in accordance with previous studies (Rao *et al.* 1996, Forsyth *et al.* 1999, Ricci *et al.* 2017). In GBM, the process of angiogenesis results in an increase of the number of vessel cells with an aberrant organisation and an anarchic endothelial cell proliferation, which yields structures forming the microvasculature sometimes referred as ‘glomeruloid bodies’. These structures are comprised of sprouting vessels that form extensions of vascular cells, resembling renal glomeruli. In GBM, MMP9 is expressed in proliferative endothelial cells (microvasculature or MVP) (Rao *et al.* 1996, Forsyth *et al.* 1999, Ricci *et al.* 2017)

and induces basement membrane degradation, enabling the migration of endothelial cells for neo-vasculature formation. This is further supported by the fact that the expression of MMP9, along with MMP2 and MMP14, are correlated with VEGF expression in GBM (Munaut *et al.* 2003). Moreover, MMP9 is induced by the hypoxic factor HIF-1 and increases VEGF action by acting on its bioavailability within the extracellular space (Du *et al.* 2008). Here, MMP9 was found strongly expressed in some, but not all, MVP structures of GBM, whereas its expression in the vascular areas in lower-grade gliomas was not observed. In GBM, this was confirmed by IF with the vascular marker CD31, supporting MMP9's involvement in angiogenesis and proliferation of neoplastic vascular cells specific of GBM. Endothelial proliferation of newly sprouted vessels sparked by hypoxia (vascular hyperplasia), which form the microvasculature, is indeed observed in GBM, but not in lower grades (Rong *et al.* 2006).

2.4.4 MUC4 expression in GBM

In glioma tumour samples expressing MUC4, the protein was found to be expressed across the tumour, mostly in the cytoplasm, with a more dramatic and less restricted expression than MMP9. In those tumours, MUC4 was clearly expressed in endothelial cells of MVP structures. A previous study showed MUC4 expression in human blood vessels, indicating this may play a role in the non-adhesive property of the endothelial lumen, which protects against pathogen entrance. Moreover, MUC4 was also found intensely expressed by the inner endothelial layer of vessels in metastatic breast cancer (Workman *et al.* 2009). Additionally, it was suggested by *in vivo* studies, that MUC4 could play a role in angiogenesis. First, in a mouse model, the knockdown of MUC4 in breast tumours led to decreased blood vessel formation associated with reduced level of EMT markers and reduced phosphorylation of VEGFR2 (Rowson-Hodel *et al.* 2018). Secondly, in another study, MUC4-knockdown cells in xenografts led to decreased vessel density and VEGF expression in pancreatic cancer (Zhi *et al.* 2014).

Here, the IHC results indicated for the first time that MUC4 is found in the tumour-associated endothelial cells of GBM (MVP structures) and not in the lower-grade (II and III) gliomas. In addition, the results of this present study showed that MUC4 tissue expression was higher in GBM compared to lower grades (grades II and III).

Importantly, a recent study demonstrated that *MUC4* is part of the three most frequently mutated genes, along with TP53 and IDH1, in mutated-IDH1 astrocytoma (grade III glioma) patients progressing towards mutated-IDH1 grade 4 glioma (secondary grade IV) (Seifert *et al.* 2020). Overall, this could suggest that like MMP9, MUC4 could be involved in glioma progression and in angiogenesis/microvascular proliferation in GBM. Therefore, MUC4 should be contemplated as a potential therapeutic target and biomarker in this type of cancer. Finally, it was found in the present study that MMP9 and MUC4, as well as EGFR and MUC4, co-localised in the microvasculature. This suggests that EGFR and MUC4, both expressed on the cell surface, could interact and trigger MMP9 expression. This should be further investigated with *in vitro* assays.

2.4.5 Association of EGFR/MUC4/MMP9 protein expression and glioma patient outcome

First, this study confirmed that IDH1 is a very powerful prognostic marker for glioma patients regardless of their histological grades (Cohen *et al.* 2013). The results showed that when used individually, MMP9 and MUC4 tissue protein expression could predict a trend of worse outcome in glioma patient diagnosis, which was not observed with EGFR (no trend). However, this trend was not significant for the individual proteins, which might be due to the moderate number of samples. Therefore, the study could be underpowered and using a larger cohort of patients could show a significant trend of worse survival in patients expressing high level of MUC4 protein, which should be tested in the future. The correlation with clinicopathological characteristics indicated no significant clinical association between overall survival (OS), and MUC4 or MMP9 individual expression. However, the difference observed for the OS was significant when combining MMP9 and MUC4 together, and even more when adding EGFR. This suggests that these proteins could be useful as tissue biomarkers for glioma diagnosis and prognosis. However, future studies in larger cohort of patients will allow the consideration of additional clinical parameters, such as MGMT promoter methylation status, age, tumour enhancement and extent of resection. Finally, correlation with treatment could not be examined since most of the samples were collected for the first time, before radiotherapy and chemotherapy had begun, as it is advocated in glioma standard of care.

2.4.6 MUC4 and MMP9 levels in serum samples

MMP9 and MUC4 are expressed in glioma tissue from grade II, and at a lower level than in the more advanced grades. This means these proteins could be useful in the early detection of the disease, when prognosis and treatment response are better. From the present study, MUC4 and MMP9 proteins can be detected in solid biopsies, which are systematically available in patients diagnosed with glioma as tumour resection is part of the standard treatment. However, it would be ideal if these proteins could be measured from biofluids (liquid biopsies). Liquid biopsies have several advantages: 1) they are available from less invasive techniques, allowing more frequent collection to monitor disease progression; 2) in some instances, they may be more representative of the entire tumour in contrast to tissue sample, which only represents a restricted portion, and 3) they can be detected at an early stage of the disease (before resection) when the treatment is potentially more efficient. Since MUC4 and MMP9 were detected from the lower grades, they represent good candidates for detection in liquid biopsy.

For these reasons, the presence of MMP9 and MUC4 in the blood serum of glioma patients and control individuals (non-glioma benign tumours) was tested and protein levels of each protein were measured. MMP9 was significantly more expressed in average in the serum of GBM patients compared with the serum of grade III patients. However, MMP9 was surprisingly not elevated in grade III patients compared with the control group. When looking retrospectively at clinical parameters of the cohort, two things could be observed: 1) the grade III glioma patients were almost all oligodendrogliomas (7/10 oligodendrogliomas versus 3/20 astrocytomas), and 2) the control group was composed of different benign diseases. These two facts potentially represent important methodology limitations as these two types of gliomas (oligodendrogliomas and astrocytomas) are regarded as two distinct groups nowadays. However, the inter-patient variation was low in the grade III group, which indicated that the issue probably raised from the control group, composed of different types of non-malignant tumours. MMP9 has been shown involved in different types of cancer (Gialeli *et al.* 2011). Therefore, MMP9 could be involved at different extents in the diseases constituting the control group.

These results suggest, however, that MMP9 could help discriminating between the two malignant glioma grades (grade III and IV), although, this could also suggest that MMP9, more specifically, may help discriminate the two glioma types (oligodendrogliomas and astrocytomas). Further experiments with a different classification (with distinct oligodendroglioma and astrocytoma groups) method is required to answer that question. This type of distinction is recommended in the new 2021 classification.

Finally, MUC4 protein could be detected in the serum but only at very low levels. MUC4 is a transmembrane mucin but is believed to be secreted in order to strengthen the extracellular mucous gel in some conditions, including malignancy (Xia *et al.* 2017). The results showed that though MUC4 level was low, it was minimally elevated in higher glioma grades (III and IV) compared to controls. However, the difference was not significant and requires further investigations in a larger cohort of patients. There are currently several glycosylated proteins used as cancer biomarkers and some of them are mucins (Felder *et al.* 2014, Kudelka *et al.* 2015, Hanson *et al.* 2016, Filippou *et al.* 2018, Manne *et al.* 2021). They often correspond to transmembrane proteins that can be secreted into the circulation after cleavage, thus presenting lower levels in serum compared to secreted proteins. However, they can still represent helpful diagnostic information, which justifies combination with other cancer biomarkers. Finally, it would be interesting to verify the level of MUC4 in the cerebrospinal fluid, although the use of this type of liquid biopsy is significantly more invasive for the patient than blood collection. Given that the event of transmembrane mucin secretion is potentially rare, their detection in a fluid circulating in contact with the central nervous system might give rise to higher levels and so more precision.

2.4.7 Conclusion

The results of this study entail that MUC4 and MMP9 could serve alone or in combination as useful diagnostic and prognostic biomarkers in glioma patients as their tissue expression correlates with higher glioma grades (grades III and higher in GBM) compared to lower (grade II). Moreover, MMP9 and MUC4 were both found to be expressed in GBM-associated MVP structures, which was a new finding for MUC4 and suggests that this protein, alike MMP9, could be involved in

microvascular proliferation/angiogenesis. When combined together in an MUC4/MMP9 or MUC4/MMP9/EGFR axis expression pattern, these proteins may be used to improve the prediction accuracy of the clinical outcome of glioma patients. Finally, these makers could potentially be used in liquid biopsies, which would represent a non-invasive strategy for the patients, although further studies must be conducted to confirm the relevance of MUC4 in that case.

Chapter 3: Interplay between EGFR, MMP9, and MUC4 and regulation of expression in glioma cell lines

ABSTRACT

Background: EGFR-dependent MMP9 and MMP2 expression plays an important role in glioma progression. An upregulation of MMP9 and MUC4 expression in glioma tissue was observed previously in Chapter 2. EGFR did not correlate with higher glioma grades, yet this protein is overexpressed in most GBM cases. MUC4 is a large, glycosylated protein overexpressed at the surface of most carcinoma cells and involved in cancer progression mostly through alteration of cell-cell interactions. In carcinomas, MUC4 has been shown to interact with EGFR family members. However, the putative MUC4/EGFR/MMP9 pathway in glioma needs further *in vitro* characterisation.

Method: Transient expression with transfection of MUC4-plasmid in GBM cells before spheroid formation was performed to evaluate the impact on the cellular behaviour. The cellular localisation and protein expression of MUC4 and EGFR were investigated using immunocytochemistry (ICC). The effect of the plant-based molecule thymoquinone (TMQ), a MUC4 attenuator was also tested. Finally, gene expression of *MUC4*, *EGFR*, *MMP9*, and *MMP2* was studied by RT-qPCR, in different conditions, including the presence of known ligands of the EGFR-MUC4 oncogenic interaction.

Results: MUC4 overexpression altered the cell-cell interactions in the GBM cellular spheroid model. MUC4 and EGFR co-localised in GBM cells and were restricted at the surface in some cells. TMQ abrogated spheroid formation and MUC4 protein expression. The MUC4-EGFR ligands were able to modulate the expression of *EGFR*, *MMP9*, and *MMP2*, whereas TMQ was able to downregulate MMP9 specifically. This downregulation was more potent than the one caused by the standard chemotherapeutic drug TMZ.

Conclusion: These results support a potential role of MUC4 in GBM cell aggressiveness. MUC4 seemed to be part of a common pathway comprising EGFR and MMP9, with a parallel weaker regulation of MMP2. However, further studies will be required to confirm and define the exact regulation mechanisms of that pathway at both protein and mRNA level.

3.1 Introduction

MMP9 is known to be implicated in GBM progression, specifically through its role in angiogenesis, invasion, and proliferation (Quesnel *et al.* 2020), while EGFR is expressed in most GBM cases (Karpel-Massler *et al.* 2009). As stated previously, the correlation between EGFR and MMP9 expression has been demonstrated in several cancer types (Mandel *et al.* 2018). In the context of GBM, it is known that the level of EGF and MMP9 present in the cerebrospinal fluids of the patients both decrease after tumour resection. Moreover, blocking EGFR in this type of cancer reduced MMP9 protein level in cell lines via the PI3k/Akt and ERK1/2 pathways (Chen *et al.* 2017). Nevertheless, the underlying mechanisms in GBM remain to be better defined.

MUC4 has important implication in carcinogenesis (Carraway *et al.* 2009) and is overexpressed and implicated in pancreatic cancer. Its induction in this cancer led to metastasis and angiogenesis, as well as a shorter survival in mice. MUC4 overexpression also led to the expression of downstream angiogenic factors, such as MMP9, through the NOTCH3 pathway (Carraway *et al.* 2009, Tang *et al.* 2016)**.** MUC4 was originally discovered in a metastatic rat adenocarcinoma. After N-glycosylation in the endoplasmic reticulum and before the addition of O-glycans, MUC4 is a 300 kDa polypeptide. MUC4 has four promoter start sites and can be activated by various transcription factors. It was proposed that numerous possible MUC4 transmembrane and secreted isoforms could be expressed from the gene sequence. For these reasons, *in vitro* investigations of MUC4 have been often presented to be somewhat challenging (Tang *et al.* 2016).

Although investigations of MUC4 interaction with EGFRs have only been performed in carcinomas, seminal studies have been conducted in leading laboratories. First, it was shown that MUC4 was responsible for tumour growth and metastasis in carcinoma cell lines, through direct interaction with HER2 and induction of its activation in response to NRG1. In addition, MUC4 induced the re-localisation of HER2 and HER3 to the membrane and impeded their internalisation - and so their degradation - when NRG1 was present (Funes *et al.* 2006). A stable transfection of MUC4 in ovarian and pancreatic cell lines induced an increase of motility ability of these cells with the apparition of lamellipodia and actine rearrangement, which was

suspected to be due to the interaction between MUC4 and HER2 (Chaturvedi *et al.* 2008, Ponnusamy *et al.* 2008). In pancreatic cancer cells, the downregulation of MUC4 led to the downregulation of FGFR1, MMP9, uPA, and EMT markers. The authors proposed a model in which MUC4 interacts at the surface with an HER2/HER1 (or HER3) complex activating the uPA/MMP9 through the PI3K pathway (Rachagani *et al.* 2012). In triple negative breast cancer cells, MUC4 was found involved in the aggressiveness and EMT process through the regulation of HER1, HER2, and HER3, observed at the protein level. Specifically, MUC4 stabilised HER2 at the membrane with no effect on its transcription (Mukhopadhyay *et al.* 2013). Importantly, the use of pan-EGFR inhibitors led to a co-reduction of EGFR and MUC4 protein level (Seshacharyulu *et al.* 2015). These findings were partially observed with MUC1, which was shown to stabilise EGFR at the surface to prevent its degradation (Pochampalli *et al.* 2007). In carcinomas, the binding of Gal-3 on MUC1 induced the formation of an MUC1-EGFR complex, which in turn, activated and prolonged EGFR signalling in presence of EGF. In addition, the interaction between mucins and EGFR was proposed to reduce the access of anti-EGFR treatments (Piyush *et al.* 2017). Several studies taken together suggest that the binding of NRG1 to complexes comprising mucins and EGFR members triggers MMP9 expression in breast cancer cells (Yao *et al.* 2001, Cho *et al.* 2008, Tang *et al.* 2008, Carraway *et al.* 2009, Momeny *et al.* 2015, von Achenbach *et al.* 2018).

One of the strategies proposed to reduce and/or abolish MUC4 is the use of thymoquinone (TMQ). TMQ is a terpenoid extracted from Black seeds (Nigella sativa), known as black cumin, which has been used as a traditional remedy in the Middle East against a myriad of aches and diseases. At the appropriate dosage, TMQ is well tolerated and presents a very low risk of side effects. Anti-cancer effects of TMQ have been extensively reported *in vitro* in a variety of cancer cell lines such as in acute lymphoblastic leukaemia, pancreatic cancer, breast cancer, hepatocellular cancer, and GBM. However, no oncologic clinical trials involving TMQ as an adjuvant drug have been conducted yet. Importantly, TMQ effects in cancer cells include the downregulation of NFκB and its downstream regulators, MMP2 and MMP9, among them. It also showed anti-proliferative affects through increase of PTEN activity, and apoptotic effect via modulation of the *p53* and *STAT3* genes (Yarnell *et al.* 2011, Goyal *et al.* 2017). Importantly, TMQ has been shown to

reduce MUC4 protein via proteasomal degradation in pancreatic cancer. This reduction was only observed post-transcriptionally. MUC4 inhibition led to increased apoptosis and decreased invasion ability with reduced cellular projections (Torres *et al.* 2010)**.** Moreover, the reduction of GBM cells' invasion ability was accompanied by a reduction of MMP2 and MMP9 protein secretion and reduction of ERK signalling (Kolli-Bouhafs *et al.* 2012).

Taken together, these studies, along with the results of the previous chapter 2, support a potential common MUC4/EGFR/MMP9 pathway that remains to be demonstrated and defined, especially in brain cancers, which are different from carcinomas. Specifically, the link between MUC4 and the gelatinases (MMP9 and MMP2) still remains to be clearly demonstrated and the ligands and the exact mechanisms involved also need to be identified. This justifies extensive *in vitro* investigations in GBM cell lines. In addition, TMQ represents an innocuous and inexpensive attractive strategy to impede GBM progression by reducing the level of MUC4 and thus blocking the downstream investigated pathway, as well as rendering the cells more accessible to anti-EGFR treatments. This is represented in **Figure 3.1**. Using a set of *in vitro* assays, the present study aims at elucidating this putative pathway, its mechanisms of action, and its potential targeting.

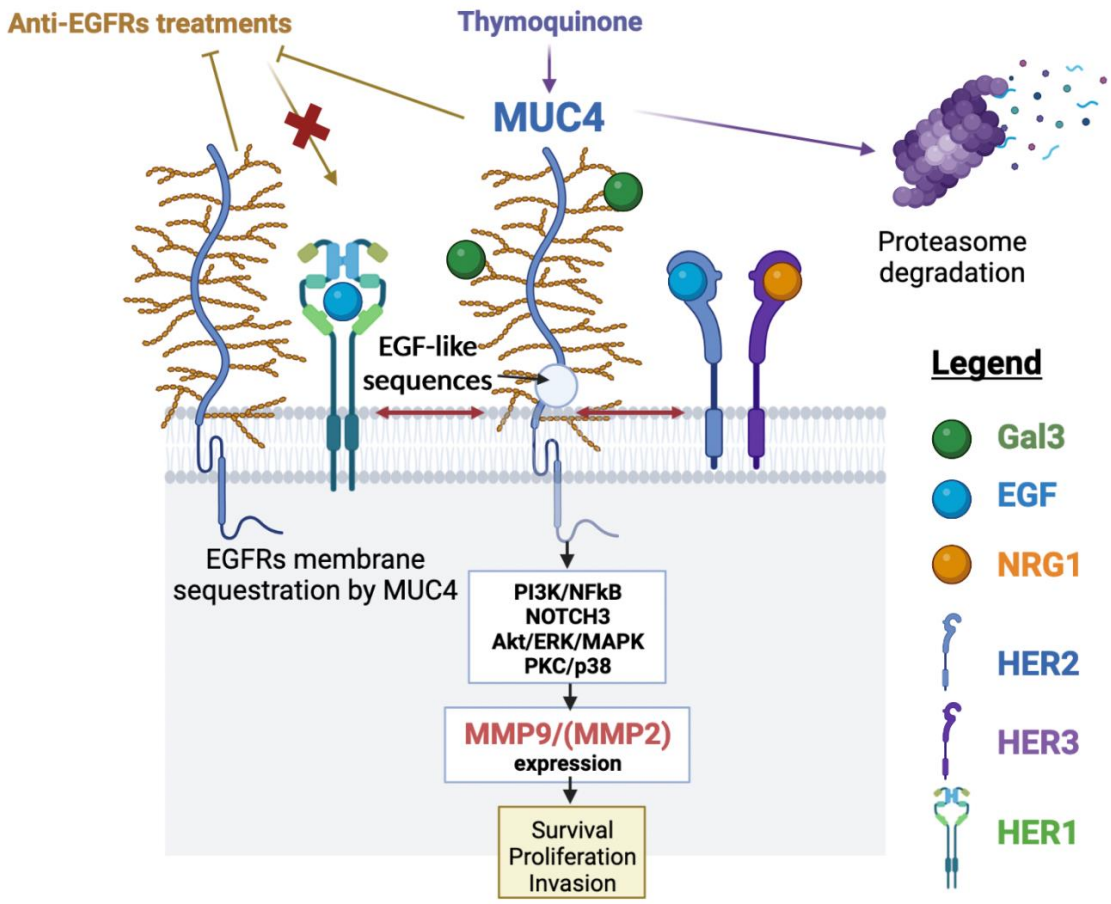


Figure 3.1 Investigated mechanisms of the MUC4-EGFR interaction in GBM progression and therapeutic strategy. Mechanisms involving MUC4, EGFRs, and MMPs in carcinoma progression are summarised together, they support the existence of the putative pathway presented, including in non-carcinoma cancers. The pathway represented is hypothetical and remains to be further specified in brain cancer. This pathway is composed of MUC4 overexpressed at the cell surface interacting with EGFRs via its EGF-like sequences, which leads to the EGFRs sequestration to the membrane and ensuing downstream signalling activation that triggers MMP9 expression, and with less evidence, MMP2 expression. This activation would require the presence of EGFRs (NRG1, EGF) and MUC4 (Gal-3) ligands. The presence of MUC4 (a large protein) at the surface impedes anti-EGFR treatments. Thymoquinone (TMQ), a very well tolerated plant-based treatment, could be used as a therapeutic strategy through proteasomal degradation of MUC4 protein. *Figure created with Biorender.com.*

The aim of this chapter was to delineate the putative MUC4/EGFR/MMP9 pathway in GBM cells with a set of *in vitro* assays. Glioma cells transfected with MUC4 and grown in spheroids were used to decipher the effect of the protein on glioma cells. Moreover, immunocytochemistry (ICC) was performed to study the proteins' expression and interaction. Finally, RT-qPCR was used to study gene expression after the addition of different ligands of EGFR and MUC4, as well as TMZ, and TMQ, to shed light on the regulation of this pathway.

3.2 Materials and methods

3.2.1 Cell lines

Cell lines used were SW1088 (#HTB-12), T98G (#CRL-1690), LN-18 (# CRL-2610), and A-172 (#CRL-1620) obtained from the American Type Culture Collection (ATCC) company (**Table 3.1**). The cell lines were all received and certified by the ATCC company by short tandem repeat DNA profiling. A negative test for mycoplasma contamination was also provided. Cells were first amplified, and stock vials from each passage (<8th passage) were kept at -150°C or liquid nitrogen for long term storage. The cells were grown in pre-warmed Dulbecco's modified eagle medium (DMEM) - high glucose (Sigma-Aldrich, #D5796) supplemented with 10% Foetal bovine serum (FBS) (Sigma-Aldrich, #F2442) and 1X antibiotic-antimycotic solution (Sigma-Aldrich, # A5955) at 37°C in a 5%-CO_2 incubator. The cells were stored in cryogenic vials with additional 5% dimethylsufoxide (DMSO) (Sigma-Aldrich, #67-68-5) freezing agent. The cells were grown in T-25 and T-75 cell culture flasks and split when reaching 70-80% confluence with a trypsin-EDTA solution (Sigma-Aldrich, #59417C) after being washed with phosphate buffered saline (PBS). 1 ml or 3 ml of trypsin-EDTA were used respectively for T-25 and T-75 flasks and the solution was left to incubate at 37°C for 10-15 minutes. Then, fresh (new) medium containing 10% FBS was added to inactivate the trypsin. Cell suspension was then spun down at 1000 rpm for two minutes and the pellet was collected, resuspended, and sub-cultured with a fresh culture medium in flasks for next passage.

Cell line	Disease	Reference	Morphology	Origin	Genome
SW1088	Astrocytoma	HTB-12, ATCC	Fibroblastic	72-year-old male	Hypertriploid
T98G	Glioblastoma	CRL-1690, ATCC	Fibroblastic	61-year-old male	Hyperpentaploid
LN-18	Glioblastoma	CRL-2610, ATCC	Epithelial	65-year-old male	p53 mutated, p16 deleted
A-172	Glioblastoma	CRL-1620, ATCC	Fibroblastic	53-year-old male	Hypertriploid

Table 3.1 Characteristics of the glioma cell lines.

3.2.2 Cell treatments

The drugs and ligands were reconstituted according to the manufacturer's instructions at the concentration recommended (**Table 3.2**). IL-6 was dissolved with a carrier protein (BSA) as recommended by the company. Cells were seeded in a 12-well culture plate (2 wells per condition) for treatment with different drugs and ligands. 400 µl of medium containing $2.5x10^5$ cells per ml was dispensed per well. When the cells reached around 70% confluence, the medium was replaced with fresh medium containing the drug or the ligand at the tested concentration (**Table 3.2**). The medium used to treat the cells only contained 5% of serum instead of 10% to reduce the interference of the serum with the drug or ligand. The control was treated with the same medium containing no drug or ligand. The final concentration of the drug or ligand in the treating medium was never above 1% and the final concentration of DMSO was 0.1% for temozolomide (TMZ) and thymoquinone (TMQ) and 0.6% for MMP9 inhibitor. After 24h and 48h, the cells were collected for gene expression analysis by using trypsin-EDTA as described previously. After being spun down at 1000 rpm for two minutes, the cells were washed in PBS and resuspended before RNA extraction.

Treatment	Stock concentration	Reconstitution buffer	Tested final concentration	Product reference
Galectin-3	50 µg/ml	PBS	0.5 µg/ml	R&D system, #8259-GA-050
EGF	500 µg/ml	10mM acetic acid	50 ng/ml	Sigma-Aldrich, #324831
NRG1	100 µg/ml	PBS	20 ng/ml	R&D system, #5898-NR-050
IL-6	100 µg/ml	PBS 1% BSA	50 ng/ml	R&D system, #206-IL-010/CF
MMP9 inhibitor	10 mM	DMSO	3 µg/ml	Abcam, #ab142180
TMZ	10 mg/ml	DMSO	10 µg/ml	Sigma, #T2577
TMQ	10 mg/ml	DMSO	10 µg/ml	Sigma-Aldrich, #274666

Table 3.2 Cell treatments specifications.

3.2.3 RNA extraction and cDNA synthesis

Prior to RT-qPCR analyses, cDNA was first synthetised from the total RNA extracted from the cells treated in each condition. RNA was extracted and cDNA was synthetised just after cell harvesting, on the same day. RNA extraction was performed using RNA extraction kits following several washing and centrifugation steps according to the manufacturer's instructions for the RNeasy mini kit (Qiagen, #74104) or the GeneJET RNA purification kit (Thermofisher scientific, #K0702). RNA concentration and quality was then determined by using a nonodrop (NanoDrop One/oneC microvolume Uv-Vis spectrophotometer, Thermofisher scientific) for each sample. The RNA was diluted in nuclease-free water at a concentration of about 50 ng/ml each time and reverse transcription was performed by using the iScript supermix cDNA synthesis kit (Bio-Rad, #1708896) according to the manufacturer's instructions. Briefly, the iScript reaction mix containing the reverse transcriptase was added to the diluted RNA solution (containing less than 1μg of RNA) at a ratio of 1 to 8. Priming was performed for 5 minutes at 25°C, reverse transcription was performed for 20 minutes at 46°C, and reverse transcriptase inactivation was performed for 1 minute at 95°C. cDNA product concentration and quality were then determined before qPCR.

3.2.4 Real-time quantitative PCR (RT-qPCR)

RT-qPCR was used to measure the relative gene expression fold change between the different conditions compared to the control condition (no treatment). qPCR was then performed using SYBR technology and specific primers designed for the targeted genes. SYBR is a dye that binds to the minor groove of the double-stranded DNA molecule, so that its intensity, which is captured by the qPCR machine, will be positively proportional to the cDNA amplification of the targeted gene.

RT-qPCR for the genes of interest (*MUC4*, *EGFR*, *MMP9*, and *MMP2*) was prepared by using the SYBR green technology from Bio-Rad. The primers were the Biorad primePCR ™ primers (MMP9 #qHsaCID0011597, MUC4 #qHsaCED0045245, EGFR #qHsaCED0045334, and MMP2 #qHsaCID0015623) optimised by the company to work optimally with the iScript cDNA synthesis kit and the Sso advanced SYBR green universal master mix (Bio-Rad, #1725271) by using the same

amplification cycle for all the genes. The primers were received in the form of a 'super mix' containing the forward and the reverse primers that could be used with no dilution step. The efficiency of the primers for each gene of interest was provided by the company and were all comprised between 98 and 101%. The reaction was prepared in a 96-well plate by dispensing 2 μl of the diluted cDNA (~100 ng/reaction) at the bottom of a 96-well plate (1 reaction), which was then covered for each gene by the mix containing the primers (1μl), the Sso Advanced SYBR super mix (10μl), and nuclease-free water (7μl). Negative control wells for each gene were prepared with the same process, except that the cDNA was replaced by nuclease-free water. Finally, the plate was centrifuged and qPCR was performed with a CFX96 Touch RT-qPCR detection system (Bio-Rad). The activation was performed for 2 minutes at 95°C, then, the denaturation (5 seconds at 95°C) and the annealing step (30 seconds at 60°C) were performed for 40 cycles.

3.2.5 RT-qPCR gene expression analysis

A housekeeping gene was used for normalisation in the qPCR since relative expression was assessed over absolute expression. The housekeeping gene was the key regulatory enzyme glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*). *GAPDH* reverse and forward primers were from GeneArt gene synthesis (Thermofisher scientific). The pair of primers was dissolved in ultra-pure water at a concentration of 100 μM and aliquoted at 10 μM and used in the reaction at the final concentration of 500 nM. To check the efficiency of the GADPH primers, cDNA from glioma cell line SW1088 was used at 5 increasing dilution factors (3, 9, 27, 81, and 243) and the mean Ct was measured by RT-qPCR from two duplicates for each dilution. The log10 (dilution factor)-mean Ct standard straight line was generated and the slope of the line was used to calculate the efficiency of the primers with the following formula: $(10^{(1/slope)-1}) \times 100$.

Relative expression was assessed using the Livak method (Livak *et al.* 2001). First, the two duplicated Ct were averaged. Then, the mean Ct was normalised for each gene of interest (GOI) in each condition to the housekeeping gene ($\Delta Ct = Ct_{(GOI)} - Ct_{(GADPH)}$), and the ΔCt was normalised for each gene and each condition to the control $\Delta\Delta Ct = \Delta Ct_{(condition)} - \Delta Ct_{(control)}$. Finally, the expression ratio was calculated

with the following formula: $2^{-\Delta\Delta Ct}$. The expression ratios of each gene were compared between the different conditions by generating a graph with GraphPad Prism version 9.3. Each condition was run once in duplicates.

3.2.6 Plasmid amplification by bacterial transformation, and purification

Insertion of the MUC4 DNA sequence into the glioma cell lines (transfection) was performed with the pCMV3-MUC4 plasmid (Sino Biology Inc, #HG16066-UT) and the empty plasmid pCMV6-XL5 (Origene) used as mock plasmid (control), Information about the plasmid maps is detailed in **Figure 3.2.** The plasmids were first amplified with bacterial transformation into ampicillin-resistant *E. coli* cells, using the One Shot TOP10 Chemically *E. coli* kit (Invitrogen, #C404010). Briefly, 1-3 μl of plasmid was added to 50 μl of competent *E. coli* for each transformation reaction, and the vial was incubated on ice for 30 minutes. Then, the vial was incubated at 42°C in a heating block for 30 seconds and placed again on ice. 250 μl of pre-warmed Super Optimal Broth (S.O.C) medium was added to the vial before a one-hour incubation at 37°C. 200 μl of the mixture was then spread on a pre-warmed Luria/Miller (LB) agar plate containing 100 μg/ml ampicillin for each transformation. The plate was incubated overnight at 37°C. The day after, the bacteria were collected with the first buffer used for plasmid purification and centrifuged for five minutes at 8000 rpm. Then, the supernatant was discarded, and plasmid purification was directly performed by using a Miniprep kit (Qiagen, #27104) following a series of elution and centrifugation steps as per the manufacturer's instructions. Plasmid quality and concentration were checked using a nanodrop as previously.

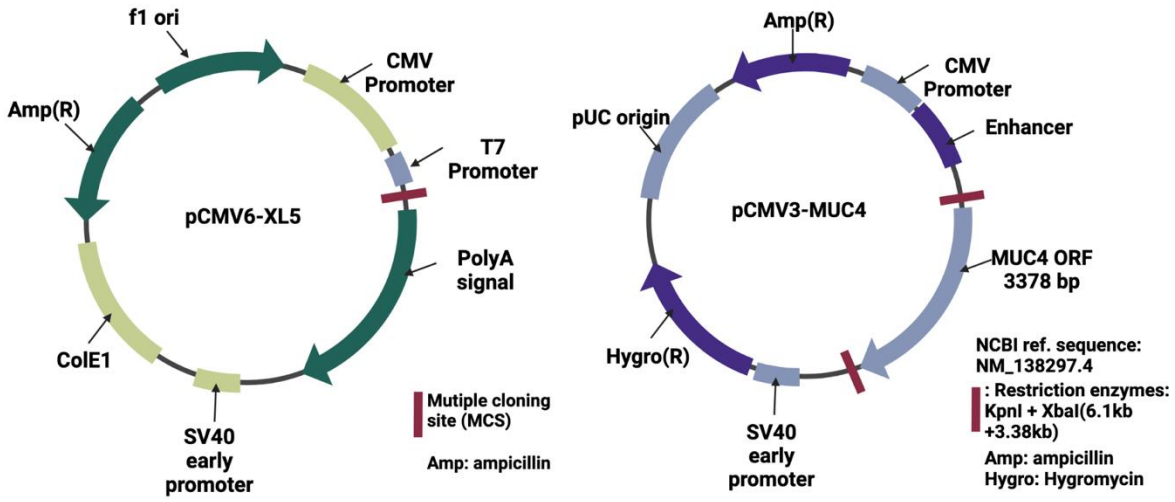


Figure 3.2 pCMV6-XL5 and pCMV3-MUC4 plasmid maps. F1 origin: bacteriophage origin of replication; CMV promoter: cytomegalovirus promoter; ORF: open reading frame; Amp(R): resistance gene for ampicillin; pUC origin: procaryote origin of replication; T7 promoter: RNA polymerase-specific promoter; SV40 early promoter: simian virus 40 early promoter for eucaryotic gene expression; ColE1: plasmid prototype vector. *Figure created with Biorender.com.*

3.2.7 Transfection of mammalian cells with pCMV6-XL5 and pCMV3-MUC4 plasmids

Glioma cells were chemically transfected with the plasmids by using Lipofectamine 2000 (Invitrogen, #1668030) when the cells reached about 70% confluence, following the manufacturer's instructions. Briefly, 15 μl of the purified plasmid and 11.3 μl of lipofectamine were diluted for 1 ml of Opti-MEM I reduced serum medium (Gibco, # 31985070) and left to incubate for 20 minutes before 3.3 ml of additional medium was added to constitute the transfecting medium. Lipofectamine works by forming liposomes around the plasmid to facilitate its integration into the cells. Incubation with the cells was performed with the transfecting medium for 5 hours. At the end of the transfection, the medium was replaced by the appropriate medium used for the following assay. The success of the transfection protocol was verified by performing RT-qPCR for *MUC4* gene expression.

3.2.8 Spheroid formation assay

Spheroid generation assay was used to assess the impact of MUC4 expression and the impact of TMQ addition on the cells' ability to proliferate, form spheroids, and interact with each other in a 3D context. The cells were seeded into two different flasks and transfected with either the pCMV3-MUC4 plasmid or the pCMV6-XL5 control plasmid as described earlier. Then, the cells were collected with trypsin/EDTA solution and washed before being diluted in fresh medium with or without 10 μg/ml of TMQ treatment. Spheroids were generated with the hanging drop method (Foty 2011, Kalli *et al.* 2018). Briefly, the cells were diluted to 2.10^4 cells/ml and 20 μl-drops of the diluted cell solution were dispensed onto the inside of a petri dish lid's surface. The lid was then placed on the dish so that the drops were hanging upside down, and the dish was filled with pure cell medium to prevent any evaporation during incubation. This technique allows the cells to grow within the drop in a spheric shape since contact with the culture plate is abrogated. The growth medium used was serum-free. After 48 hours, images were taken to analyse size and number of the spheroids with the Icy software (Institut Pasteur-France BioImaging, Copyright 2019) (**Fig.3.3**). The spheroids were left in the drops for imaging, and not transferred to a new culture dish, to preserve their integrity (number, size, and shape). 19 images were analysed for each condition (76 images in total). The mean size for each condition was calculated and compared with unpaired t-test using GraphPad Prism. The number of spheroids per image was counted for each condition and also compared with Mann-Whitney U-test. p-values significance are defined as followed: *** indicates a p-value ≤0.001; ** indicates a p-value ≤0.01; * indicates a p-value ≤0.05.

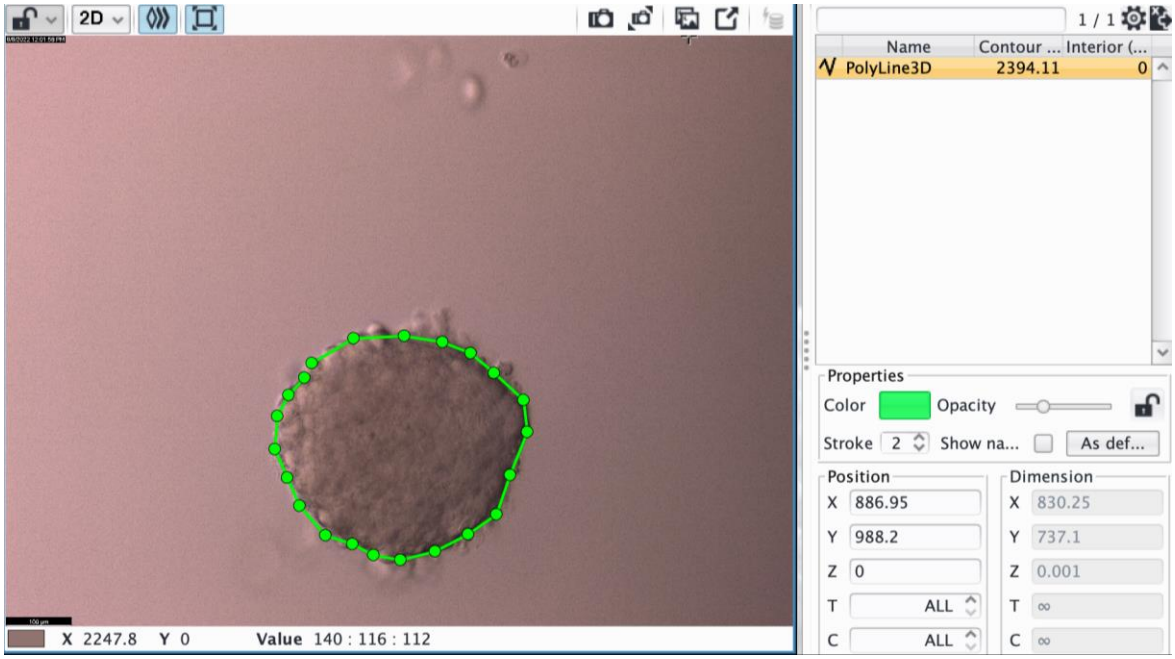


Figure 3.3 Measurement of relative spheroid size. Relative size of spheroids was measured with the Icy software (Institut Pasteur-France BioImaging, Copyright 2019), by surrounding the spheroids with the PolyLine 3D option.

3.2.9 Immunocytochemistry (ICC)

Immunocytochemistry (ICC) is a technique using the same principle as IF, except that fluorescence staining is performed on cells grown on a microscope slide. ICC was performed by first sterilizing glass coverslips with 70% ethanol. The coverslips were placed in a petri dish after being rinsed with PBS, and then covered with medium containing $2.5x10^5$ cells per ml (15 ml was used to fill one petri dish). After reaching 60-70% confluence, the cells were transfected as described before. The transfecting medium was then replaced by fresh medium and the cells were left to grow for one more day with or without treatment. Then, the cells were gently rinsed twice with PBS and fixed in a 4%-formaldehyde solution (Sigma-Aldrich, #1040021000) at room temperature for 20 minutes, washed in PBS and TBS-T, and blocked with TBS-T containing 5% BSA and 2% goat serum for 45 minutes at room temperature. Then, the coverslips were placed at the bottom of a plastic box covered by a piece of parafilm, this allowed the use of a small volume of the antibody solutions, directly dispensed onto the surface of the coverslip, with no risk of

leakage. The primary antibodies EGFR rabbit monoclonal (Life technologies #700308) and MUC4 mouse monoclonal (Abcam, ab60720) diluted in the blocking solution (respectively at 1:1000 and 1:250) were left to incubate on the coverslips overnight at 4°C. The next day, the coverslips were washed with TBS-T. Diluted secondary antibodies (see IHC section in previous chapter) were left to incubate with the cells for one hour in the dark at room temperature before being successively rinsed with TBS-T, PBS, and distilled water. The coverslips were removed from the dish (with a pipette's tip) and placed on a piece of tissue with a drop of fluorescent mounting medium containing DAPI. After one minute of incubation, the coverslips were mounted on a naked microscope slide and images were captured with a Leica DMi8 fluorescent microscope (Leica microsystem, UK).

3.3 Results

3.3.1 Optimisation and gene expression in glioma cell lines

The Bio-Rad PrimePCR SYBR Green assay was chosen for gene expression analysis. With this assay, primers for different commonly investigated genes can be used with the same amplification program, the same SYBR, and the same protocol. The assay is pre-optimised to work with the Bio-Rad qPCR thermal cycler and the Bio-Rad Sso Advanced SYBR master mix, both used in this study. Thus, MMP9, EGFR, MUC4, and MMP2 Bio-Rad PrimePCR SYBR Green assay primers were selected and used with the Sso Advanced SYBR and the cDNA synthesis kit recommended by the company (iScript). The validation data provided by the company are listed in **Table 3.3**.

Gene	Amplified sequence (Ensembl link)	Ct at 50% RFU amplification (number of cycles)	Melting curve peak (°C)	Amplification efficiency (%)
EGFR	ENSG00000146648	26	81	98
MMP9	ENSG00000100985	32	86	101
MMP2	ENSG00000087245	24	87	99
MUC4	ENSG00000145113	30	87	99

Table 3.3 Bio-Rad PrimePCR SYBR Green assay primers and their validation parameters. [Relative fluorescence units (RFU)].

To be able to interpret the qPCR results, first, the efficiency of the *GAPDH* primers needed to be validated. The expression of this housekeeping gene was used in the expression analysis to normalise the data (relative expression). The efficiency should be ideally comprised between 98 and 101% and very similar for all the other primers used in the study. The efficiency was measured by generating a standard straight line plotting the mean Ct of five decreasing cDNA dilutions against the log10 of the dilution factor. The slope of the straight line was then used to calculate the amplification efficiency, the efficiency for the GAPDH primers was 101%, which was deemed acceptable (**Fig.3.4**).

A

cDNA dilution factor	Log10 (dilution factor)	Mean Ct	Standard deviation	Efficiency
3	0.48	22.38	0.42	10^(1/3.299)-1)x100 = **101 %**
9	0.95	23.89	0.25	
27	1.43	25.17	0.12	
81	1.91	27.07	0.22	
243	2.39	28.66	0.29	

B

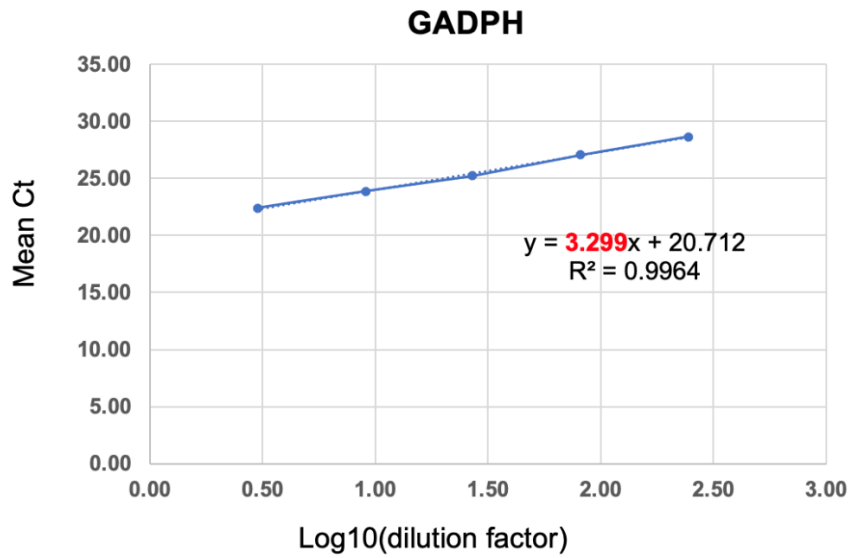


Figure 3.4 Standard line and efficiency evaluation of the GAPDH primers. (A) Five decreasing dilutions of cDNA from SW1088 cells were analysed by qPCR, mean Ct are represented with standard deviation. (**B)** Standard straight line of mean Ct for all dilutions plotted against the corresponding dilution factor. The slope of the line is used to calculate the efficiency (101%).

Next, the expression of the genes was tested with the standardised qPCR assay in the different glioma cell lines, to examine gene expression before *in vitro* assays. For cDNA samples extracted from each cell line, the housekeeping gene and the genes of interest were tested with the primers and recommended protocols. The Ct of the *GAPDH* gene was valid and very similar in all cell lines (Ct comprised between 19.3 and 20.9). *MUC4* was not detected in the grade III SW1088 cell line and expressed at low levels in the other GBM cell lines (Ct = 36.5-38). *EGFR* and *MMP2* had an average expression in all the cell lines (Ct = 23-26 and 19.5-25, respectively), and *MMP9* was expressed at low-medium levels (Ct = 31-38) in the cells. Therefore,

it was concluded that this qPCR assay could be used for the rest of the *in vitro* experiments.

Since MUC4 was shown to be expressed only at low levels, a MUC4 transfection protocol was optimised and validated. The plasmid for MUC4 mammalian cells transfection was the pCMV3-MUC4 and the empty control plasmid was pCMV6-XL5. After plasmid amplification via transformation of competent *E. coli* cells and plasmid purification, the plasmid quality was checked and the cells were transfected with the respective plasmids. SW1088 cells were used for validation of the transfection process since these cells did not express *MUC4* at detectable levels according to the qPCR results. The cells were left to incubate with the transfecting medium (containing lipofectamine and the plasmid) and toxicity was evaluated phenotypically under the microscope. The transfecting medium did not affect the viability of the cells after an overnight incubation. The cDNA synthesised from the transfected cells was analysed with the qPCR assay. An incubation for 5 hours was enough to observe a strong MUC4 overexpression in the SW1088 cells (raw data are shown in **Table 3.4**).

Condition	Gene	Mean Ct (triplicate)	Standard deviation
Non-MUC4-transfected	*GADPH*	23.3	0.69389
	MUC4	**39/no detection**	x
MUC4-transfected	*GADPH*	22.7	0.19245
	MUC4	**19.5**	0.35355

Table 3.4 Validation of MUC4 transfection protocol in SW1088 cells. Triplicate was used for each condition. MUC4 is not detected in SW1088 cells before transfection (Ct = 39 or no Ct) and more expressed than the housekeeping gene after transfection.

3.3.2 Effect of MUC4 overexpression and thymoquinone (TMQ) on spheroid formation

To study the effect of MUC4 overexpression on the proliferation ability of glioma cells, LN-18 GBM cells were transfected with either a control plasmid (pCTRL) or

the MUC4 plasmid (pCMV3-MUC4) and spheroids were then grown from both conditions. To further evaluate the effect of TMQ treatment, both types of spheroids were grown with and without 10 μg/ml of TMQ in the medium. The spheroids were generated by using the hanging drop method and images were taken in each drop after two days. The medium containing the drug was not changed and was left until all the images were taken. Both MUC4-overexpressing spheroids and control spheroids had formed after two days but in corresponding conditions grown with 10 μg/ml of TMQ, the spheroids did not form (**Fig.3.5A**). However, MUC4-overexpressing cells grown with TMQ seemed to form more small aggregates than their control equivalent and single cells appeared to be less numerous (**Fig.3.5A**).

In the conditions containing no drug, the control spheroids formed were larger, while the MUC4-overexpressing spheroids were smaller but more numerous. The number and the size of all the spheroids in the images were reported and compared. Since the distribution was not normal, Mann-Whitney U-test was used. This observation was confirmed and the difference was highly significant ($p<0.0001$ for both size and number) (**Fig.3.5B**).

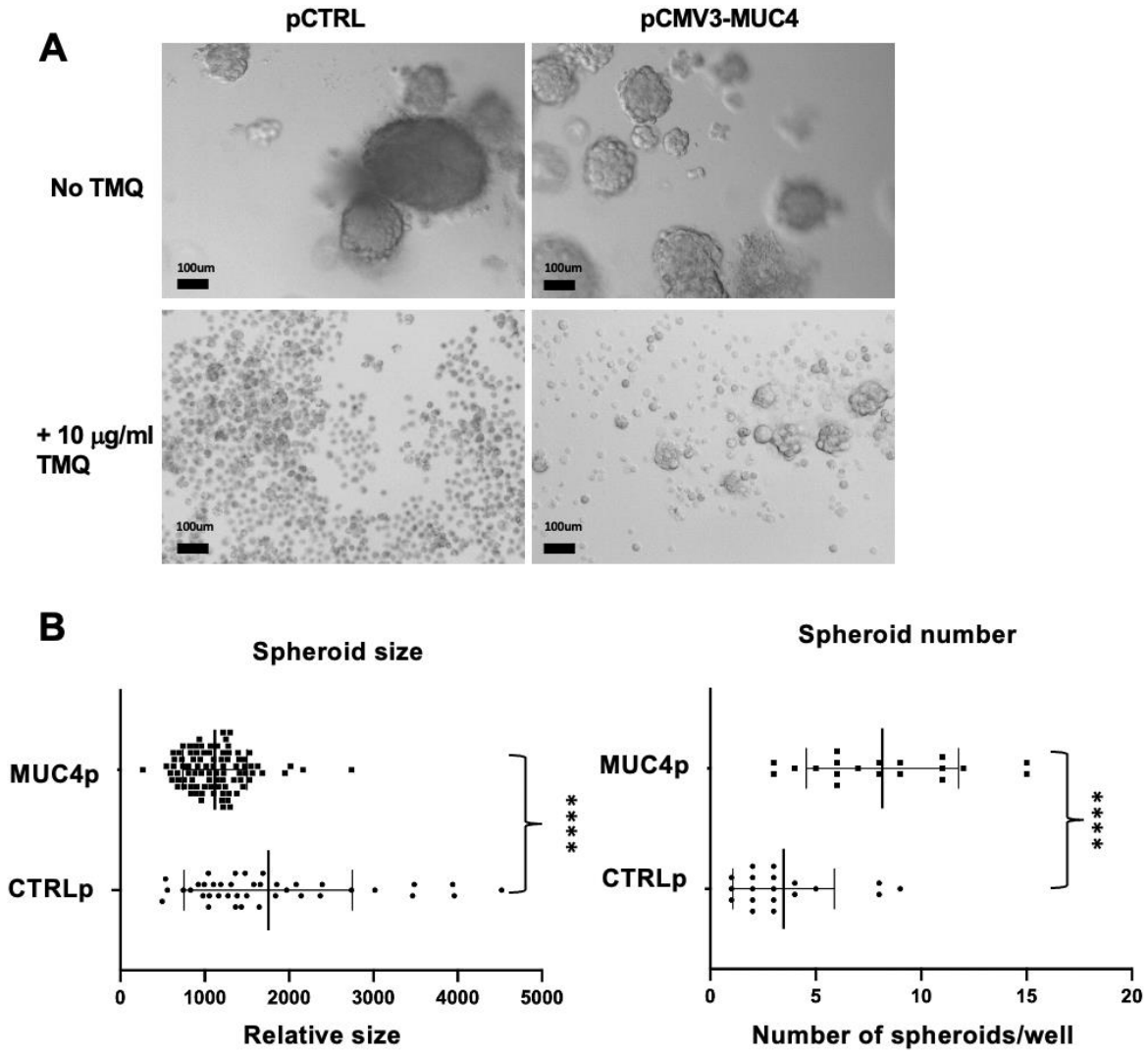


Figure 3.5 Spheroid formation assay on MUC4-overexpressing GBM LN-18 cells. (A) Images representing MUC4-overexpressing spheroids and control spheroids after 48h with and without 10 μg/ml of TMQ. Spheroids did not form with TMQ, especially from cells not expressing MUC4 (pCTRL). **(B)** The size of all spheroids formed (with no TMQ) was measured (relative units) and compared with Mann-Whitney U-test, spheroids formed from MUC4-overexpressing cells were significantly smaller ($p<0.0001$). The number of spheroids per well was also calculated, there was significantly more spheroids formed from MUC4-overexpressing cells than from control cells ($p<0.0001$). Scale bar = 100 μm.

3.3.3 Effect of MUC4 overexpression and TMQ treatment on MUC4 and EGFR protein localisation

Next, EGFR and MUC4 protein expression and localisation in the cells was evaluated with immunocytochemistry (ICC) in LN-18 cells under three different conditions: 1) control condition with no MUC4 transfection and no drug, 2) MUC4-overexpressing cells, and 3) MUC4-overexpressing cells with 10 μg/ml of TMQ. EGFR protein was expressed in GBM cells before MUC4 transfection, as expected, and the expression was observed in the cytoplasm/surface of the cells, similar to what was observed in the tissue. EGFR expression seemed to be considerably higher than MUC4 in the control cells (**Fig.3.6**). This was also expected since *MUC4* was also only expressed at relatively low level in the qPCR results before transfection.

As expected, after pCMV3-MUC4 plasmid transfection in LN-18 cells, MUC4 and EGFR were expressed and MUC4 exceeded EGFR expression, especially in the cytoplasm. The expression of these proteins appeared to be polarised within some cells, with EGFR and MUC4 co-localizing at the surface of one side of the cells (orange arrows) and an isolated MUC4 expression in the cytoplasm of the opposite side (green arrows) (**Fig.3.6**). The co-expression at the surface was associated to an exclusion of the proteins from the cytoplasm. This result suggests that all the proteins on this side of the cells might co-localise at the surface and interestingly this was not observed in all the cells, suggesting that this might happen only at a certain stage of the cell cycle.

When treating the MUC4-overexpressing cells with 10 μg/ml of TMQ for 24 hours, the protein level of MUC4 was reduced as assessed by the very low signal (**Fig.3.6**) and the cells looked more similar to the control condition with a dominance of EGFR expression in the cytoplasm and with MUC4 signal that was not visually detectable. In these cells, MUC4 appeared to be expressed at an even lower level than in the control condition before MUC4 overexpression, although this would need confirmation with another technique as subtle differences are difficult to visualise with fluorescence.

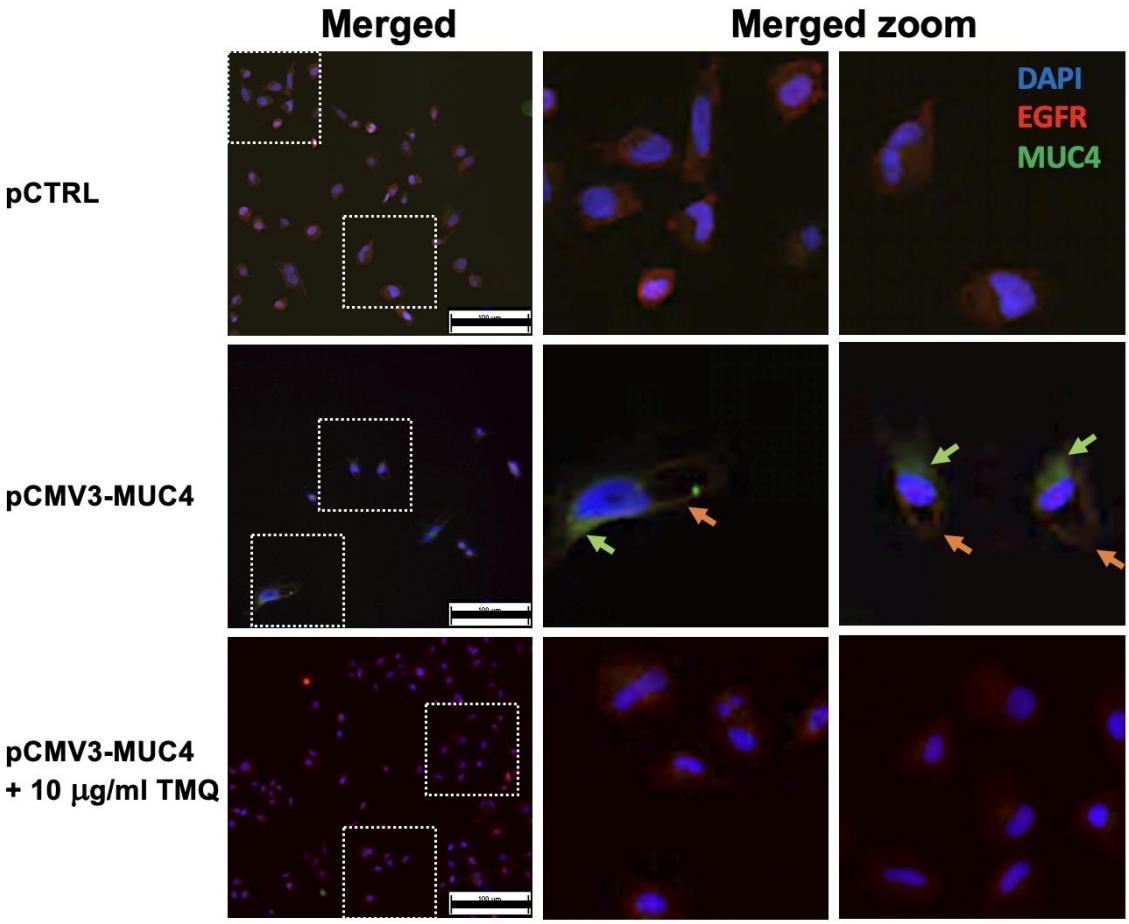


Figure 3.6 Immunocytochemistry (ICC) on MUC4-overexpressing LN-18 cells. EGFR (in red) and MUC4 (in green) protein localisation and expression was evaluated with ICC in control cells, and MUC4-overexpressing cells with and without TMQ. The images show merged representations of the three colours (nuclei are stained with DAPI and appear blue). The dotted white squares on the first panel are zoomed in the second and third panel. In the middle row, the green arrows indicate isolated MUC4 expression at one side of the cells and the orange arrows indicate co-localisation of MUC4 with EGFR. Scale bar = 100 µm.

3.3.4 Gene expression analysis in glioma cell lines at different time points

To investigate the putative EGFR/MUC4/MMP9 molecular pathway, RT-qPCRs were performed in the GBM LN-18 cell line. As previously noted, *MUC4* expression was only expressed at low level in the cells, compared to the *EGFR*, *MMP9*, and *MMP2* in the control conditions. *MUC4* was first included in the RT-qPCR analysis to see if some conditions would significantly increase its expression, however, since *MUC4* was only expressed at low level in these cells, Ct values obtained were too high to obtain good replicability. Therefore, *MUC4* was not included in the RT-qPCR analysis.

In addition to the control condition, other conditions with ligand and drug treatments were tested. Different ligands whose binding with EGFR/MUC4 is expected to have an effect on the investigated pathway were selected: EGF, the most common ligand involved in EGFR activation; neuregulin-1/heregulin (NRG1), the ligand that binds to the MUC4/EGFR complex in carcinomas; and galectin-3 (Gal-3), that binds to MUC4 glycan chains in carcinoma with a role in carcinogenesis. In addition, interleukin-6 (IL-6) was selected since this cytokine has been shown to activate MUC4. A permeable and reversible MMP9 inhibitor was added to the selection as well as two treatments: TMZ, which constitutes the standard treatment in GBM; and the plant-based TMQ treatment, expected to reduce MUC4 protein expression.

The results were generated with the Livak method (results are normalised to the same housekeeping gene in each sample and treated cells are normalised to the control condition). The results are given at two time points after treatment (RNA extraction at 24h and 48h post-treatment) on GBM LN-18 cells. *MMP2* was also investigated in addition to *EGFR* and *MMP9* since MMP9 and MMP2 are the two gelatinases among the MMP family. These two proteins present high homology and very similar substrates and are often found co-regulated in the context of EGFR signalling (Menashi *et al.* 2003, Majumder *et al.* 2019). Including *MMP2* in the gene expression analysis will give indication on whether this gelatinase is part of the pathway under investigation or rather, that MMP9 is specifically linked to EGFR. Results for all the RT-qPCRs are represented in **Figure 3.7** (in **A** for the ligands and **B** for the drugs/inhibitors).

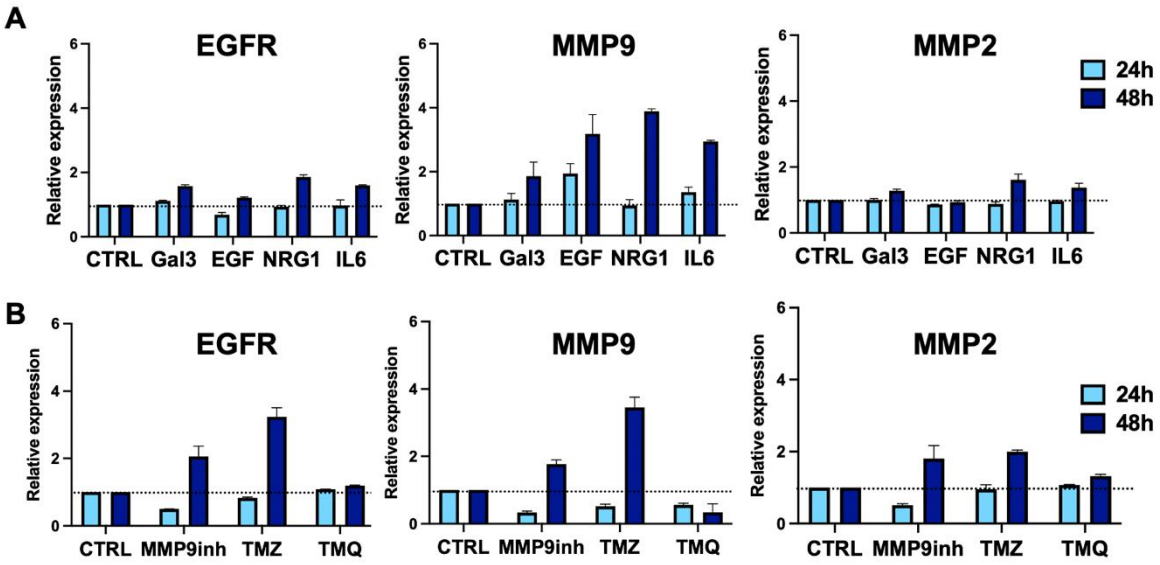


Figure 3.7 *EGFR*, *MMP9*, and *MMP2* gene expression analysis in glioma cells. Gene expression was assessed in LN-18 cells at 24h and 48h after different treatments with relative RT-qPCR assay. The results shown are relative to the control. **(A)** RT-qPCRs after treatments with different ligands and **(B)** different drugs/inhibitors. Control (CTRL); galectin-3 (Gal-3), 0.5 μg/ml; neuregulin 1 (NRG1), 20 ng/ml; thymoquinone (TMQ), 10 μg/ml; MMP9 inhibitor (MMP9inh), 10 mM; temozolomide (TMZ), 10 μg/ml. Error bars represent sampling variation (standard deviation) between the duplicates of the same experiment. The highest standard deviation was 0.35.

At 24h post-treatment, MMP9 inhibitor led to a decrease of *MMP9* (3-fold reduction), and to a lesser extent, *EGFR* and *MMP2* (2-fold reduction for both), confirming that these three genes code for proteins that might be part of a common pathway. The addition of EGF ligand was responsible for an increase of *MMP9* expression (~2-fold increase). TMZ treatment led to a decrease of *EGFR* and even more *MMP9* (~2-fold reduction), whereas *MMP2* was not affected. TMQ led to a decrease of *MMP9* only, at a similar level that the standard drug TMZ (~2-fold for both).

EGFR and *MMP9* expression appeared to be more affected by the different treatments than *MMP2* at 48h. Interestingly, this gene seems to behave (react to the treatments) like *MMP9* but at a much lesser extent. This suggests that *MMP2*

might be linked to the EGFR/MMP9 pathway but this link must be indirect or weaker that the EGFR/MMP9 signalling interaction. NRG1 and IL-6 - which can both act on MUC4 - increased the genes' expression, and more particularly *MMP9* (~4-fold and ~3-fold increase, respectively). MMP9 inhibitor led to a sharp increase of the genes at 48h, which is expected since this inhibitor target MMP9 protein, which probably induces the induction of the gene expression as a response to re-establish the expression of the protein. The same observation was made with TMZ, with a sharp augmentation at 48h (~3.5-fold increase for *MMP9*), while the genes were decreased at 24h. Finally, TMQ was the only drug that did not lead to a responsive augmentation but instead lead to an important decrease of *MMP9* specifically (~3.5-fold reduction), and not *MMP2*.

3.3.5 Gene expression analysis in MUC4-overexpressing GBM cells

Finally, LN-18 cells were transfected with the MUC4 plasmid to see the potential impact of MUC4 overexpression on the other genes and to investigate the impact of the treatments on *MUC4* expression. As previously specified, the transfection protocol was validated in SW1088 (because these cells do not express *MUC4* at detectable level) and gave a very high expression (superior to the housekeeping gene). RT-qPCR was performed for all the genes in LN-18 GBM cell line on cDNA synthesised from RNA extracted at 48h.

Induced *MUC4* expression completely eliminated *MMP9* and *EGFR* gene expression in all conditions, but not *MMP2*. Given that this result was unexpected, this was verified in the grade III SW1088 cell line separately and the RT-qPCR gave exactly the same results (data not shown). Since the transfection led to a high MUC4 expression (overexpression), this result could be due to an ensuing high protein expression and a negative feedback retro-control on the other pathways. These results support that MUC4/EGFR/MMP9 are linked in a common regulatory pathway, whereas MMP2 is linked to that pathway to a much lesser extent as suggested previously. In respect of what is the precise effect of MUC4 on EGFR and MMP9 expression, these results are not fully conclusive. Therefore, adjustment and optimisation of the transfection protocol (transient and/or stable) should be performed to allow the analysis of the exact detailed mechanisms of regulation between EGFR, MUC4, and MMP9.

3.4 Discussion

3.4.1 Effect of MUC4 overexpression on spheroid formation

A standard spheroid formation assay was first performed to evaluate the effect of MUC4 overexpression on GBM cells. The spheroid formation was observed in both control and MUC4-overexpressing cells. However, the number and size of these spheroids were significantly different in both conditions. From MUC4-overexpressing cells, spheroids formed were higher in number but smaller compared with control cells (not overexpressing MUC4). This could be due to a higher fusion rate of spheroids when MUC4 is not present/or present at low levels at the surface of the cells, or higher dissociation of the spheroids due the MUC4 overexpression. In both cases, this may be explained by the possible anti-adhesive properties of MUC4 protein. Anti-adhesion is believed to be one of the main functions of MUC4 in pathological conditions, including cancer where the induction of the protein has been shown to make the cells detached from each other in *vitro* (Komatsu *et al.* 1997). This is believed to be due to the glycan chains sterically impeding the interaction between surface ligands (for example integrins) and their receptors. In addition, in pancreatic cancer, where MUC4 is the most prevailing oncogene, its overexpression provokes disruption of cell-to-cell interactions (Xia *et al.* 2017). This phenomenon has been shown to be dependent on the glycosylation extent of the protein (Chaturvedi *et al.* 2008). It is important to note that MUC4 is also believed to facilitate cell-cell interactions in another context, by providing new interaction sites, adhesion sites (AMOP domain), and potentially via ligand sequestration facilitating the binding to their receptors (Chaturvedi *et al.* 2008). Further experiments are needed with more tested conditions for further validation, but this result supports a role of MUC4 in cancer cell dissemination, which was not shown previously in GBM and especially in a 3D spheroid cellular model.

3.4.2 Effect of TMQ on MUC4-overexpressing and control spheroids

The spheroids did not form when TMQ was added to the medium at the concentration of 10 μg/ml. TMQ has been shown to reduce the level of MUC4 protein expression by proteasomal degradation potentially via modifying the glycosylation pattern on the protein (Torres *et al*., 2010). The concentration used

correspond to one of the highest doses typically used in other *in vitro* studies. Since TMQ is a very well tolerated drug – not granted with plant-based treatment - the use of an even higher dose is conceivable. When looking more carefully, some very small spheroids still formed in the MUC4-overexpressing condition, while single cells were hardly present. The opposite was observed in the control (not transfected with MUC4) cells, with no spheroids formed and high number of single cells. This result suggests that MUC4 overexpression still could give an advantage to spheroid formation abilities or proliferation. Further experiments need to be conducted to decipher this result. Moreover, some of the cells treated with TMQ seemed to be bigger and more translucent when overexpressing MUC4, which suggests that MUC4 might confer a survival advantage. Finally, the fact that TMQ was able to impede spheroid formation even when MUC4 was only expressed at low level from the RT-qPCR (control spheroids), suggests that the effect of this drug is potentially very potent and confirms that it might act on other onco-molecular targets along with MUC4 (Goyal *et al.* 2017).

3.4.3 Protein localisation and effect of TMQ

The ICC assay first confirmed the overexpression of MUC4 at the protein level after transfection, as well as an average EGFR protein expression and a low MUC4 protein expression in the cells before transfection, that aligned with the RT-qPCR results. After transfection, the pattern of EGFR localisation was altered and, in some cells, a clear co-localisation at one pole of the cells was observed with little possible doubt. However, in most of the cells the co-localisation did not appear to be polarised. In the cells showing the polarised co-localisation, EGFR was totally absent from the cytoplasm (appearing black in the image), which clearly suggests a sequestration of the receptor to the membrane induced by the interaction with MUC4. This suggests that MUC4 could impede the internalisation – and so the degradation and recycling - of EGFR (HER1), supporting the prolongation of its activation through exposition to ligands. This result aligns with a similar mechanism to what has been described in carcinoma cell lines with HER2 and HER3 receptors in previous studies (Funes *et al.* 2006, Mukhopadhyay *et al.* 2013). Therefore, this phenomenon could exist in GBM tumours, since MUC4 was shown to be expressed in most GBM patients in the previous chapter. These results could also suggest the activation of downstream signalling pathways, which should be investigated in future

studies. Finally, MUC4 protein was not observed in the cells anymore after TMQ was added to the cell growth medium. This result shows that TMQ leads to a strong MUC4 downregulation, at the protein level, which has been also observed in pancreatic cancer cells (Torres *et al.* 2010).

3.4.4 Regulation of EGFR, MUC4, MMP9, and MMP2 at the mRNA level

The gene expression analysis on non-transfected GBM cells was then performed and confirmed a co-regulation of EGFR and MMP9 in glioma cells (Chang *et al.* 2015, Chen *et al.* 2017) associated to a less obvious regulation of the other gelatinase MMP2. The ligands tested included EGF, NRG1, and Gal-3, which have all been shown to activate complexes formed by mucins and EGFRs (Funes *et al.* 2006, Piyush *et al.* 2017). The concentrations for the treatments were chosen to be relatively high but still within the standard range. This analysis also showed that TMQ, a non-specific MUC4 protein inhibitor, was able to downregulate MMP9 at the mRNA level. This downregulation was observed in previous studies, but only at the protein secretion level (Kolli-Bouhafs *et al.* 2012). This downregulation was compared with the one caused by the standard treatment TMZ.

The reduction of expression was at first very similar (24h), but after 48h, TMQ maintained a downregulation of *MMP9* gene expression, while TMZ seemed to induce what looked like a positive retro-control on that gene, which could be due to a strong inhibition at the protein level. This observation may be explained by the different mechanisms of action of these two drugs. TMZ has been shown to downregulate MMP9 and MMP2 at the protein level (Mirabdaly *et al.* 2020). It will be important to conduct extensive dose-dependent assays at different time points with these two drugs on different molecular actors of GBM progression, MMP9 being an important one. *MMP2* seemed to be regulated in a similar fashion to *MMP9*, although at a lesser extent. There are only very few differences between the two homologous proteins and previous studies have more often highlighted the co-regulation of EGFR signalling with MMP9 only, although MMP2 is sometimes shown to be co-regulated with MMP9. MMP9 has been shown to play the most important role among other MMPs in GBM progression (Quesnel *et al.* 2020). Taken together, these results suggest that MMP2 has remaining function and similar regulation to MMP9 (aligned with the fact that they are almost identical) but that this protein has

more attenuated function and EGFR-independent responses than its homologous MMP9 counterpart.

In MUC4-overexpressing cells, the *EGFR* and *MMP9* gene expression levels were undetectable. However, from the ICC, EGFR was still present at the protein level after the same incubation time. Further optimisation of the transfection assay needs to be done to solve this drastic effect and elucidate the exact molecular mechanisms of this pathway. However, this phenomenon was consistent between two different cell lines, and shows a strong influence of MUC4 on EGFR and MMP9 regulation.

3.4.5 Conclusion

By using *in vitro* assays, this study first confirmed the EGFR/MMP9 co-regulation in glioma cells. In addition, several potential regulatory mechanisms observed were new and coherent with previous findings made in carcinoma cells with other approaches:

1) MUC4 protein seemed to display an anti-adhesive role and had an impact on cell-cell interactions in cells grown in spheroids, further supporting its potential role in cancer cell dissemination.

2) MUC4 and EGFR (HER1) co-localised at the surface of the cells after a MUC4 transient overexpression, which led to EGFR restriction to the membrane.

3) TMQ led to a drastic downregulation of MUC4 protein expression and abrogation of spheroid formation.

4) TMQ downregulated specifically MMP9 (and not MMP2) at the mRNA level more efficiently than the standard drug TMZ.

Chapter 4: Monitoring biomolecular changes with a focus on glycosylation using Raman spectroscopy on glioma cells

ABSTRACT

Background: Raman spectroscopy (RS), a technique used since relatively recently in the biomedical field, may lead to fast and objective cancer diagnostics, but also provides information about changes in the bio-composition at the cellular level. The aim of this chapter was to examine the application of RS in glioma cell models for discrimination and identification of biological changes, with a focus on glycosylation. This study further assessed the sensitivity and detection limit of this methodology as a proof-of-concept.

Method: Glioma cell lines originated from two grade IV patients and one grade III patient were compared using RS combined with principal component analysis (PCA) and classification learning method. Another grade IV cell line was then transfected with MUC4 protein to assess whether this method could detect more specific changes introduced by the overexpression of the heavily glycosylated protein. The cells were grown on a steel slide in monolayer and spheroids were generated with the hanging drop method and dispensed on a steel dish for RS acquisitions. The dataset was processed and analysed with PCA and classification learning. Peak assignment was performed with an in-house glycosylation database and a general literature-based database, both generated for this study.

Results: Grade III and IV cell lines clustered after PCA when using the largest principal components (PCs) and could be discriminated in both models. An increase of DNA and proteins in grade IV could explain the highest difference between both grades, as well as an increase in lipid content. A decrease of glycans in grade IV was notably observed in the monolayer model and less marked in the spheroid model. Finally, the subtle discrimination between MUC4-overexpressing glioma cells and normal glioma cells was challenging, but still feasible as a clusterisation was observed within the five largest PCs.

Conclusion: Different glioma grades and cell lines could be easily discriminated, while more subtle differences due to the overexpression of a single highly-glycosylated protein was possible but would require more acquisitions for more detailed analysis.

4.1 Introduction

Raman spectroscopy (RS) is a label-free spectroscopic tool that can be combined with standard microscopy and provides a molecular signature of any type of sample, including live or fixed cells. First, the use of Raman on cells have been suggested and optimised for an *in-vivo* intra-operative application. Manoeuvrable RS optical probes have been tested during surgery, these probes can be used in direct contact with the brain and detect glioma cells that are not detectable with MRI only. This method identified invasive glioma cells with good accuracy beyond three centimetres of the MRI margin and therefore could potentially be implemented to refine surgical resection (Jermyn *et al.* 2016). Other implications embrace single cell diagnostic and identification/monitoring of biological changes occurring during cancer development at the cellular level.

In vitro studies have used RS on cancer cell lines to demonstrate diagnostic applications and to identify biomolecular changes. These studies were able to detect broad differences between different types of cells and changes observed during different cancerous events such as transformation and metastasis. Fixed prostate cancer cell lines were successfully used to study the effect of irradiation on the lipid content and concentration in the cytosolic lipid droplets (Roman *et al.* 2020), while metabolic alterations were identified between breast cancer cell lines with different metastatic tropisms (Marro *et al.* 2018). Some cancers have been studied extensively with this technique. In breast cancer, it was found and confirmed by independent studies that the most important difference between the highest aggressive type and other cell types (transformed cells, normal cells) could be assigned to an increase and alteration of the lipid content (Damayanti *et al.* 2013, Chaturvedi *et al.* 2016, Iwasaki *et al.* 2021), but also a decrease of polysaccharides (Damayanti *et al.* 2013). Therefore, RS has proven to be a useful tool for biomolecular identification in cancer cells.

Other elegant *in-vitro* studies attempted to identify smaller and more subtle differences. For instances, Raman micro-spectroscopy was used to monitor treatment resistance in non-small cell lung cancer cell lines. Patients with known resistance to an anti-EGFR drug displayed fewer Raman changes after treatment with that drug than with other anti-EGFR treatments. Importantly, this was shown to

be dependent on the presence of one EGFR mutation, which was in accordance with western blot results (Hammoud *et al.* 2018).

From extensive comparison of the literature on RS performed on cells, a pattern of potential limitations could be identified, such as for example, the use of different substrates. The traditional glass and quartz substrates are affordable and have been used extensively, but these types of substrate give relatively low signal and the high noise contribution to the spectrum requires additional data correction (Kerr *et al.* 2015). Certain substrates seem potentially too costly for clinical routine application (e.g., gold, and calcium fluoride CaF_2). Other limitations include the use of different complex and in-house bio-informatic methods for data analysis, different cell treatments before acquisitions (fixation, drying process), different numbers of acquisitions per cell, the use of automated acquisitions that prevents a good control of the laser focus, the use of quick acquisitions that reduces the spectral resolution and finally, a bias during spectral assignments to different biological components (Damayanti *et al.* 2013, Chaturvedi *et al.* 2016, Hammoud *et al.* 2018, Marro *et al.* 2018, Roman *et al.* 2020, Iwasaki *et al.* 2021). For example, changes in the 1000-1200 cm^{-1} spectral region at the cellular level, has been assigned to proteins (Corsetti *et al.* 2018) or fatty acids only (Marro *et al.* 2018), while only a few studies (Chaturvedi *et al.* 2016) have assigned this region to carbohydrates, potentially overlooking the importance of glycosylation.

The aim of this chapter was to assess whether RS could be used for cell discrimination between grade III and grade IV glioma cell lines and for biological component identification at the cellular level in glioma. In addition, MUC4-overexpressing LN-18 cells were compared with normal glioma cells to further assess the sensitivity of RS at capturing more subtle changes induced by the overexpression of a heavily glycosylated protein upregulated in GBM patients. For this, cells were grown both in monolayer and in spheroids and RS acquisitions were performed on mirrored stainless-steel slide. This type of substrate was shown to present several advantages in comparison with CaF_2 slides that has been used frequently with cell culture: it gives a lower standard deviation, a better spatial resolution permitted by a double beam reflection, and a better signal-to-noise ratio (Lewis *et al.* 2017).

Moreover, in addition to the use of steel substrates, this study was designed so as to bypass different other potential limitations: 1) the cells were targeted only once by the laser to ensure the viability and integrity of the cells, 2) live cells were targeted while being in their native state (non-fixed), 3) manual and precise acquisitions were used to preserve the reproducible focus between each cell and the spectral resolution, 4) the analysis performed on the data consisted of a simple method comprised of standard Raman processing, PCA, and classification learning already built in MATLAB, and finally, 5) the combined use of both an in-house glycosylation database and a general literature database were used as an attempt for objective spectral assignments.

4.2 Materials and methods

4.2.1 Principles of Raman spectroscopy

Raman spectroscopy (RS) is a type of vibrational spectroscopy that relies on the interaction between incident photons from a monochromatic laser beam and the molecules contained within the targeted sample. In general, when light interacts with matter, the incident photons are mostly scattered following the Rayleigh scattering effect, meaning with no frequency shift, while a small fraction (1 photon out of 10^8) is scattered inelastically following the Raman effect (i.e., with a change of frequency), producing what is called a 'Raman shift'. When the scattered photon is lower in energy than the incident photon, the emission is shifted toward the red wavelengths and is called the Stokes effect, while the converse phenomenon is called the anti-Stokes effect (**Fig.4.1A**). The Stokes effect is the most likely effect to occur at room temperature and therefore is the effect observed when using RS. One of the biggest advantages of this technique is that no label is required since the Raman effect is based on the intrinsic vibrational properties of the biomolecules. Different vibrational modes represent different positions of the atoms within the molecule, namely, stretching, which can be asymmetrical or symmetrical: and bending, for instance, twisting and wagging (Ferraro, 2003). The association of all the intensities of the Raman shifts (represented by different spectral peaks corresponding to different wavenumbers) constitutes a molecular fingerprint of the sample that can be studied to evaluate the relative presence and ratio of all the

different biomolecules of life (proteins, lipids, DNA, etc) in the interrogated organic samples. Typically, a reduced intensity for a wavenumber indicates a reduced concentration of the biomolecule assigned to this specific wavenumber. Consequently, RS has been used in biology for sample identification, classification, and tissue delineation, even though it was not specifically intended for biological purposes (Crow *et al.* 2003, Mamede *et al.* 2021, Iturrioz-Rodríguez *et al.* 2022). As the Raman scattering effect is very weak, it needs to be coupled with a powerful laser, a sensitive charge-coupled device (CDD) detector, and different optical components to successfully capture subtle changes (**Fig.4.1B**). In addition, it can be coupled to a traditional optical microscope to allow visualisation and precise targeting of the sample. Finally, excitation with a near-infrared laser overcomes the problem of sample alteration as well as interference caused by autofluorescence (Daniel C. Harris 1990, Ferraro, 2003, Srinivasan 2010).

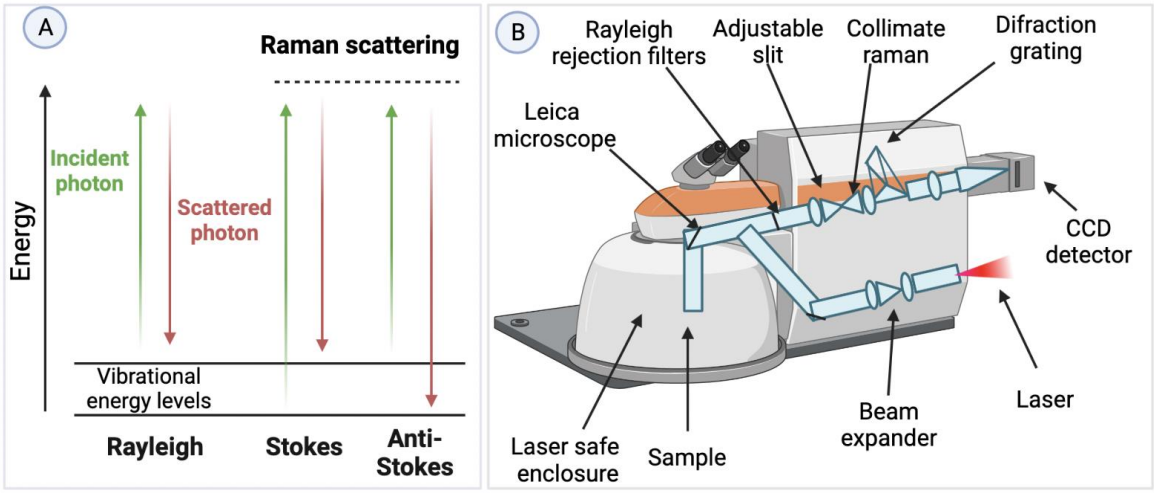


Figure 4.1 Principles of Raman spectroscopy. (A) Schematic representation of Raman scattering effect (Stokes and anti-Stokes versus Rayleigh). **(B)** The different parts of Raman spectrometer (in Via Qontor, Renishaw, Gloucester, UK) are represented. *Figure created with Biorender.com.*

4.2.2 Rationale and methodology overview

RS can give the molecular fingerprint of any organic sample in the form of a spectral signature, thereby, RS can be used to capture significant and subtle changes between samples. Unlike with infrared spectroscopy, water does not generate a strong Raman signal and thus, RS is suitable to study aqueous solutions and

interrogate *in vitro/in vivo* samples (Auner et al. 2018). The frequency of the Raman shift corresponds to the x-axis of the spectral signature and is measured in wavenumbers (cm^{-1}), while the intensity is represented in the y-axis in relative units as a function of the frequency. The Raman peaks can afterwards be assigned to biological components by comparison with an empirical Raman database. Database extracted from the literature can be used, but an in-house database for specific biomolecules can also be generated by using reference spectra generated from standards of these biomolecules. This is justified by the fact that different machines and differences in the method processing can generate slightly different Raman shifts for the same components. Therefore, since this study focuses in part on the identification of glycosylation changes in glioma cells, an in-house glycosylation database containing the seven most common glycan motifs was generated to serve as a database for peak assignment.

The general overview of the experimental pipeline for this chapter is schematically represented in **Figure 4.2.** Careful technical considerations need to be addressed before performing RS analysis including the choice of substrate used for interrogation; but also, choice of excitation length, sample preparation method, and types of processing and statistical analysis. All the acquisitions were performed on stainless-steel slides with a 785 nm laser and after processing, the dataset was analysed with PCA and machine learning classification. Then, the mean spectra were compared and the differences between them were assigned by comparison with the database. The rationale for the methodology and processing steps will be further explained in each of the following sections.

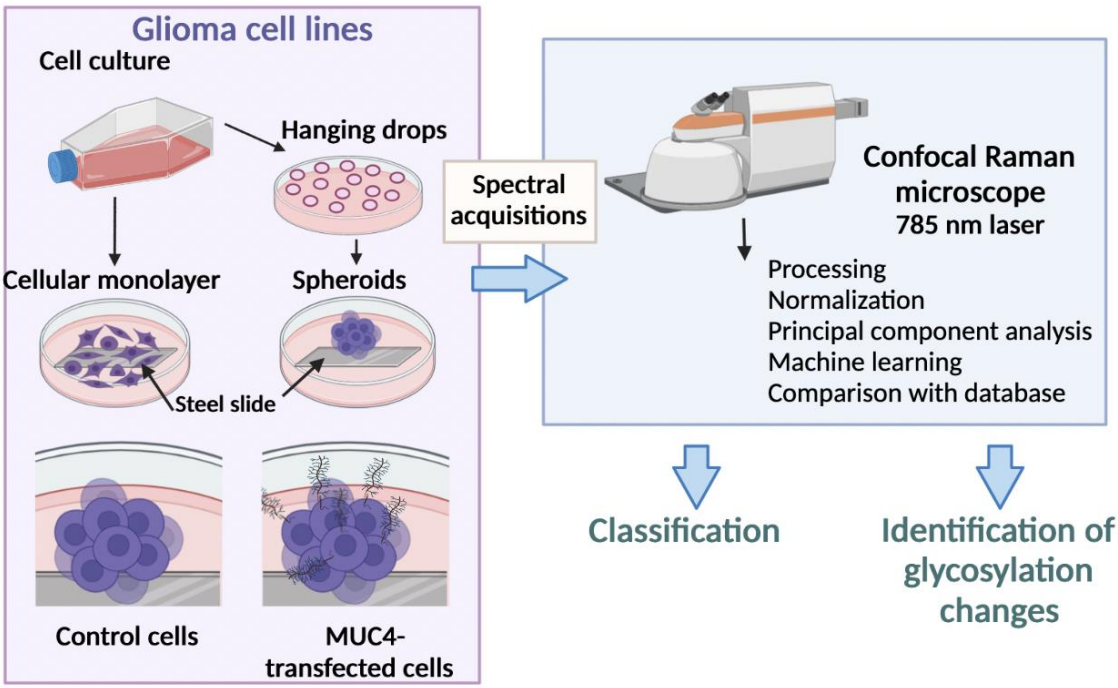


Figure 4.2 Methodology overview of Raman spectroscopy on glioma cell lines. Cell lines from glioma patients, grown in monolayer or spheroids, were interrogated with a Raman confocal microscope. After processing, normalisation, principal component analysis (PCA), and machine learning methods, the study tested whether the different cell lines and cell conditions (monolayer versus spheroids and control versus MUC4-overexpressing) could be discriminated/classified. Spectral changes were assigned to biological changes, by comparison with an in-house glycosylation database for glycosylation patterns and with a more general biological database extracted from the literature. *Figure created with Biorender.com.*

4.2.3 Raman spectrometer and acquisition parameters

All single spectral acquisitions were taken manually with a Raman spectrometer inVia Qontor (Renishaw, Gloucestershire, UK). All spectra were taken between 400 and 1800 cm^{-1} with a 50X objective, corresponding to the biological fingerprint region (Chau *et al.* 2008). A 785 nm (near infra-red) laser was used as it usually gives a low fluorescence background while generating a good Raman intensity with minimal sample alteration (Kerr *et al.* 2015). An integration time of 10 seconds was used with a 50% (55mW) power and a pinhole was applied to focus the laser. Each manual measurement was about 2-3 minutes long. The laser was calibrated by using the 520 cm^{-1} peak of a silicon wafer before acquisition as that peak is well

characterised and has been shown useful for calibration (Roman *et al.* 2020, Surmacki *et al.* 2022). Mirrored stainless-steel slides were chosen as the substrate for all acquisitions since this type of substrate has been shown to enhance the signal-to-noise ratio compared to the CaF_2 substrate that is more traditionally used for Raman. Stainless-steel slides also represent an advantage for potential clinical use in routine as they are significantly less expensive and susceptible to damage (Lewis *et al.* 2017). In addition, glass substrates produce high fluorescence background when using a 785 nm laser (Kerr *et al.* 2015). Since a 785 nm laser was used, the stainless-steel slide was considered a better option for this study.

4.2.4 Generation of the in-house glycosylation database

To generate the glycosylation database, Raman spectra of several glycan standards were collected with the Raman spectrometer (InVia Qontor, Renishaw, Gloucester, UK), following the parameters described before, which were the same used for sample acquisitions. Glycan standards (in powder form) used were Mannose (Sigma, #92683), Fucose (Sigma, #931887), N-acetyl-galactosamine (Sigma, #A2795), N-acetyl neuraminic acid (Sigma, #19023), Galactose (Sigma, #G0750), Glucose (Sigma, #G8270), and N-acetyl-glucosamine (Sigma, #A8625). Standards were dissolved in ultra-pure water at least at two different concentrations (50 mg/ml, 25 mg/ml, and/or 12.5 mg/ml) to verify that the characteristic peaks would change intensity proportionally to the concentration to assure specificity of the peaks. Since glucose has a well-characterised Raman signature, its spectrum was used as a reference control to check the quality of the protocol (Arboleda *et al.* 2000). A 20 µl-drop of the solution was dispensed on a mirrored stainless-steel slide for each concentration of each standard. Duplicates were then averaged for each glycan and each concentration. Only the highest peaks (reaching at least 100 intensity counts) were considered as specific and were selected to be included in the listed database.

4.2.5 Data processing

WiRE software (Renishaw, Gloucestershire, UK) was used for automatic baseline subtraction and cosmic ray removal. A 5^{th} order polynomial smoothing (Stavitzky-Golay) of the curve was applied to suppress the noise in order to make the spectra easier to interpret and compare. Cosmic rays are due to particles with extra-atmospheric origin that randomly hit the detector producing a spurious signal that

can interfere with the data. They appear as very narrow and sharp peaks that are quite easy to detect by an experienced eye (Anthony *et al.* 2017). After automatic removal of cosmic rays, additional manual removal was performed if their presence was still detected.

The dataset was processed in a matrix format of around 1300 wavenumber points given by the software (in txt. Format) with their corresponding intensity for each acquisition. The following processing steps were performed in Excel software. First, the acquisitions were normalised to be rescaled between 0 to 1 (Minimum-maximum normalisation) with the formula (x-min(data))/Max(data)-min(data)), where 'min' is the minimum value contained in the dataset. This type of normalisation rescales the values across the dataset to make them easier to compare and interpret. Applying this type of normalisation before analysing Raman data is relevant, since erratic variations in the signal intensity can be observed between each acquisition. Thus, the 0 to 1 normalisation is used to prevent one acquisition from over-influencing the final analysis. Obvious outliers (repeats presenting high deviation from the characteristic mean signature) were removed from the dataset before statistical analysis, since they were believed to be non-specific to the sample.

4.2.6 Principal component analysis (PCA)

PCA is a method that can be used to reduce the dimensionality of a large dataset. This technique makes the interpretation of the dataset easier without removing important statistical information by transposing the variances between the samples into new variances called principal components (PCs), which are smaller and easier to visualise. PCA allows the instant visualisation of the variance between the samples as well as their clusterisation as a function of the different PCs plotted (Jolliffe *et al.* 2016). The largest variances explaining the differences observed between the samples correspond to the first PCs (PC1, PC2, PC3). Each PC explain a certain percentage of all the variances between the samples and so the larger they are, the more they potentially contribute to a meaningful classification of the samples. The PCA plot can be represented in different dimensions. n dimensions (nD) PCA plots can help visualise the clusterisation of the samples according to n PCs. The two or three first PCs are generally plotted since they account for the largest variations that discriminate the samples. However, other more subtle

variances can also have a biological significance and thus different PCs can be explored.

Consequently, this technique is particularly helpful to analyse dataset generated with RS. First, because the dataset is large and significant information needs to be extracted from it. Secondly, because the PCA plots generated make all the potential sample classifications appear. The goal of this chapter was firstly to determine if glioma cells could be classified with the RS according to their cell line/grade. Therefore, if the samples are clustered on the PCA plots according to their category, the technique of RS combined with PCA can represent an objective tool for glioma classification with biological meaning. For these reasons, the PCA technique is commonly chosen to interpret Raman data in biology (Crow *et al.* 2003, Auner *et al.* 2018, Kang 2020, Kopec *et al.* 2021, Iturrioz-Rodríguez *et al.* 2022).

After processing, PCA was performed and 2D/3D plotted for the averaged and normalised samples for tissue and serum by exporting the dataset matrix into MATLAB R2021b software (The Mathworks, Inc., Massachusetts). Then, a simple sequence of code entries was determined to generate the PCA plots with colours that corresponded to the category for each sample. All the variables corresponding to one discrete wavenumber point, were used for the PCA to represent the entire spectral range between 400 and 1800 cm^{-1}.
The code lines used in the software were the following:

`>A=table2array(data)` converts the table containing the data in a homogenous array on which PCA can be performed.

`>[coeff,score,latent,explained] = pca(A)` performs the PCA on the array A and gives the coefficients (loadings), the scores, the variances, and the percentages of variances between the samples explained by each PC.

`>x=score(:,1)` assigns the x-axis of the 3D/2D PCA plot to PC1
`>y=score(:,2)` assigns the y-axis of the 3D/2D PCA plot to PC2
`>z=score(:,3)` assigns the z-axis of the 3D PCA plot to PC3

>scatter3(x,y,z,size,A(1:n,1:1));colormap(jet(colour number code));
axis equal
xlabel('name')
ylabel('name')
zlabel('name') generates the 3D PCA plot with the three largest PCs assigned previously (x, y, and y). 'size' is the size of the dots on the plot, '1:n' refers to the lines in the data table that needs to be plotted, n being the number of samples (the number of lines) so that all the samples are plotted in this example. '1:1' corresponds to the first column that contains the information about the category of each sample and that gives a different colour for each group. 'name' is the name chosen for each axis title.

>gscatter(x,y,A(1:n,1:1))
axis equal
xlabel('name')
ylabel('name') generates a 2D PCA plot in the same manner.

To generate the loading plot, the 'plot' function was used in MATLAB as follows: plot(1:1270,coeff(:,1:3),'-'); where 1270 was the total number of variables in this example and 1:3 defined PC1 to PC3.

4.2.7 Classification learning and statistics

Then, the built-in classification learner application was used in MATLAB to generate the accuracy rates of each classification model. All classification models were tested for each comparison. The three largest components were chosen for the classification as they were enough to discriminate the patients according to their grade on the PCA plots. A 5-fold cross-validation was used for the classification to protect the classification against overfitting. p-values significance are defined as followed: *** indicates a p-value ≤0.001; ** indicates a p-value ≤0.01; * indicates a p-value ≤0.05.

4.2.8. *In vitro* acquisitions on live cells grown in monolayers

For acquisitions performed on cells, the cell lines used were the glioma cell lines A-172, SW1088, and T98G (ATCC) (for cell lines discrimination) and LN-18 (ATCC)

(for MUC4 transfection). For the monolayer cell culture, the cells were seeded in a petri dish containing a stainless-steel slide covered with 10 ml of DMEM medium containing the $2.5x10^5$ cells (< passage 10) supplemented with 10% FBS and 1X antibiotic-antimycotic. The stainless-steel slide (Renishaw) was sterilised beforehand with 70% ethanol and UV light exposure. After 48 hours of incubation (37 ^oC, 5% CO_2), the slide covered by the cells was collected and rinsed with PBS. Acquisitions were then performed while the cells were still viable, and the slide was kept wet by adding PBS onto the surface of the cells every 10 minutes to prevent dryness caused by the microscope's laser and light. Viability of the cells was evaluated by eye. All the acquisitions were performed within 30 minutes for each slide and one cell was targeted for each acquisition. 12 cells were interrogated for grade III and 28 for grade IV (n=9 for A-172, and n=19 for T98G).

4.2.9 *In vitro* acquisitions on live spheroids

Spheroids were generated with the hanging drop method as previously described using $2x10^4$ cells per ml. After 48 hours, the drops containing spheroids were collected with a pipette and washed once with PBS. The spheroids were left in a small volume of PBS and dispensed inside a 48-well stainless-steel culture plate just before acquisition. One spheroid was targeted for each acquisition. 12 spheroids interrogated for grade III and 32 for grade IV (n=10 for A-172, and n=22 for T98G).

For spheroid cell cultures using transfected cells, the GBM LN-18 cells were transfected as previously described, with either the empty plasmid (pCTRL) for the control cells, or the MUC4 plasmid (pMUC4) for the MUC4-overexpressing cells as previously described. The cells were grown in culture flasks until 70% confluence was reached and then, the growing medium was replaced by the transfecting medium (containing the plasmid and the lipofectamine). The effect of the transfecting medium on the cells' viability was first evaluated visually and was deemed non-toxic for the cells. The day after, the cells were trypsinised and seeded for spheroid generation with the hanging drop method as described previously. Raman acquisitions were performed on 13 different spheroids in control and MUC4-overexpressing spheroids respectively.

4.3 Results

4.3.1 Characterisation of an in-house glycosylation database

An in-house glycosylation database was first generated for glycosylation monitoring. The seven glycan motifs most commonly found on heavily glycosylated proteins, such as in mucins (Chugh *et al.* 2015, Tondepu *et al.* 2022), were chosen to be included in this database, namely, glucose, fucose, galactosamine, galactose, glucosamine, mannose, and neuraminic acid (sialic acid). Standards of the different glycans dissolved in water, were interrogated by RS. To ensure that the different peaks generated by the Raman were specific to the glycan, each standard was interrogated at two different concentrations. Peaks that did not decrease in intensity when lowering the concentration were not considered specific. Only a peak at about 1645 cm^{-1} was non-specific in all the glycan signatures because the intensity of this peak was not proportional to the concentration (**Fig.4.3**). This peak corresponded to the liquid water contribution according to the literature (Turcotte *et al.* 1992). All the other peaks reaching at least 100 intensity counts had an intensity proportional to the concentration and were included in the glycosylation database. These peaks are listed in **Table 4.1**. Overall, the protocol was therefore associated with very low background noise. All the glycans clearly shared a common increased intensity in their signature in the 950-1200 cm^{-1} regions (**Fig.4.3**).

The mean spectra for each glycan (representative signatures) are represented in **Figure 4.4A**. The glucose signature was in accordance with the reference spectrum from previous studies (Arboleda *et al.* 2000). Glycan signatures shared common increased intensities, in the 800-910 cm^{-1} and especially in the 950-1200 cm^{-1} region (pink area), but also in the 400-600 cm^{-1} region (except for neuraminic acid). As a demonstration of concept of the PCA approach performed on biological spectroscopic data, PCA of the different glycans (in duplicates and used at the same concentration) was performed. Since the glycan standards are giving different signatures and are corresponding to different organic components, the PCA must be able to discriminate these standards and the first principal component (PC1) should account for almost the totality of the discrimination since the difference is expected to be obvious and large and the noise should be very low. The results are presented in **Figure 4.4B**. The clusterization was clear: the different glycans

clustered by pair of duplicates and PC1 accounted for 96.5% of the variance between the samples. The Raman and PCA approach on spectroscopic data with possibility of glycan identification was thus demonstrated with success.

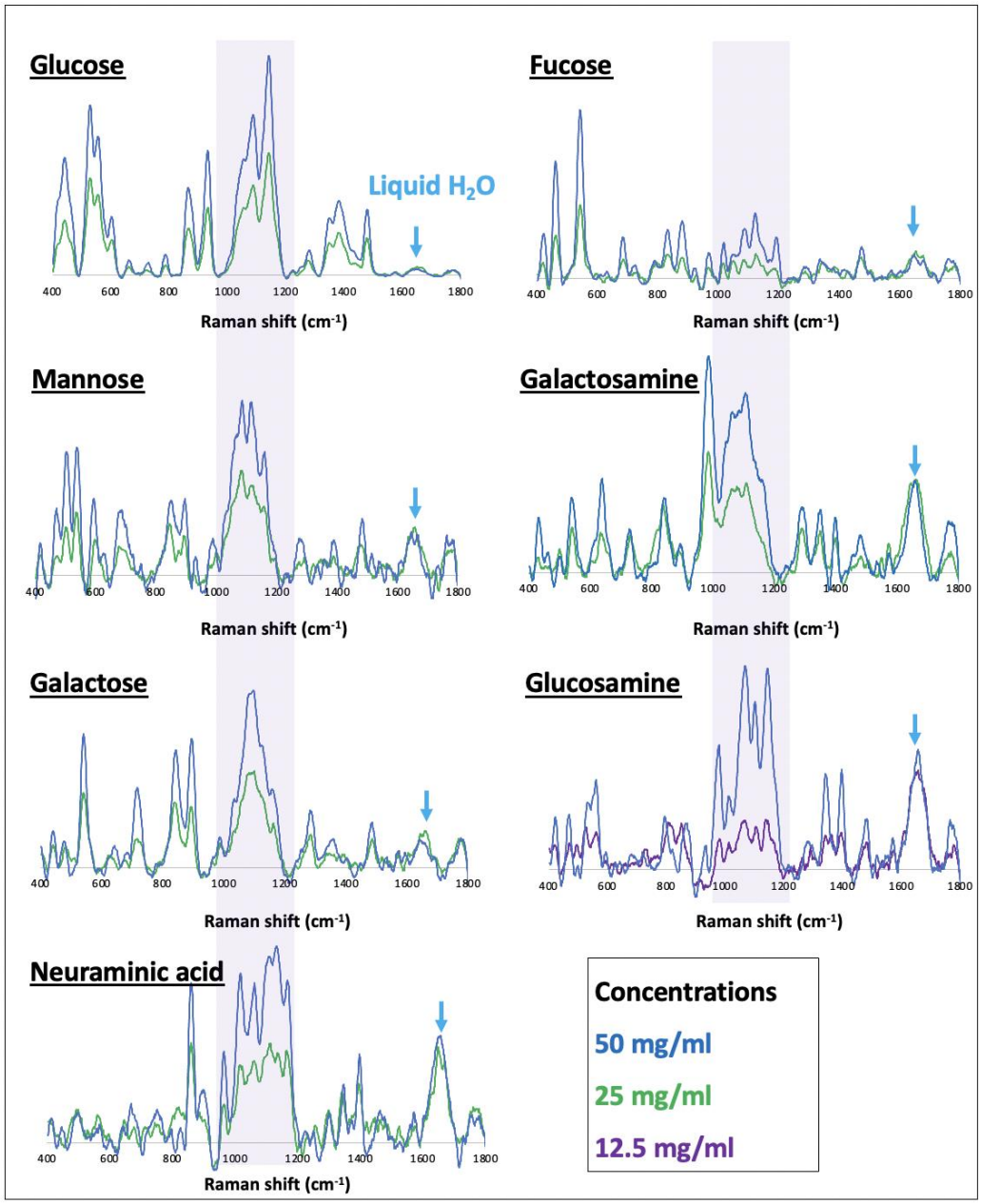


Figure 4.3 Specificity of glycan Raman signatures. The standards for the seven most common glycans found on glycoproteins were interrogated at two different concentrations to ensure the specificity of the peaks. Shaded pink area highlights the most significant intensity increase shared by all the glycans. Blue arrows indicate the liquid water peak that was not specific to the glycans and not included in the database.

Peak WN (cm^{-1})	Component(s) assigned	Peak WN (cm^{-1})	Component(s) assigned
417	Glucose	**1023**	Fucose
434-436	Galactosamine, fucose	**1036**	Neuraminic acid
440	Glucose	**1046**	Glucosamine
445	Glucosamine, galactose	**1055**	Galactosamine
476	Mannose	**1059**	Glucose, mannose
510	Glucosamine, galactose, mannose	**1063**	Fucose
517-519	Galactosamine, galactose, fucose	**1078**	Glucosamine
536	Glucosamine	**1085**	Galactosamine
615	Galactosamine	**1094**	Mannose
662	Fucose	**1000-1200**	Galactose
700	Galactose	**1100-1109**	Fucose, neuraminic acid
706	Galactosamine	**1122**	Glucose, glucosamine
819	Galactosamine	**1130**	Mannose
836	Neuraminic acid	**1143**	Neuraminic acid
857-859	Fucose, glucose	**1169**	Fucose
863	Neuraminic acid	**1261**	Glucose, galactose
873	Galactose	**1269**	Galactosamine
912	Glucose	**1323-1326**	Glucosamine, galactosamine
944	Neuraminic	**1369**	Glucose
958	Glucosamine	**1377**	Glucosamine, neuraminic acid
819-969	Galactosamine	**1460**	Glucose
990	Glucosamine	**1772**	Galactose
998	Neuraminic acid		

Table 4.1 In-house glycosylation database. [wavenumber (WN)].

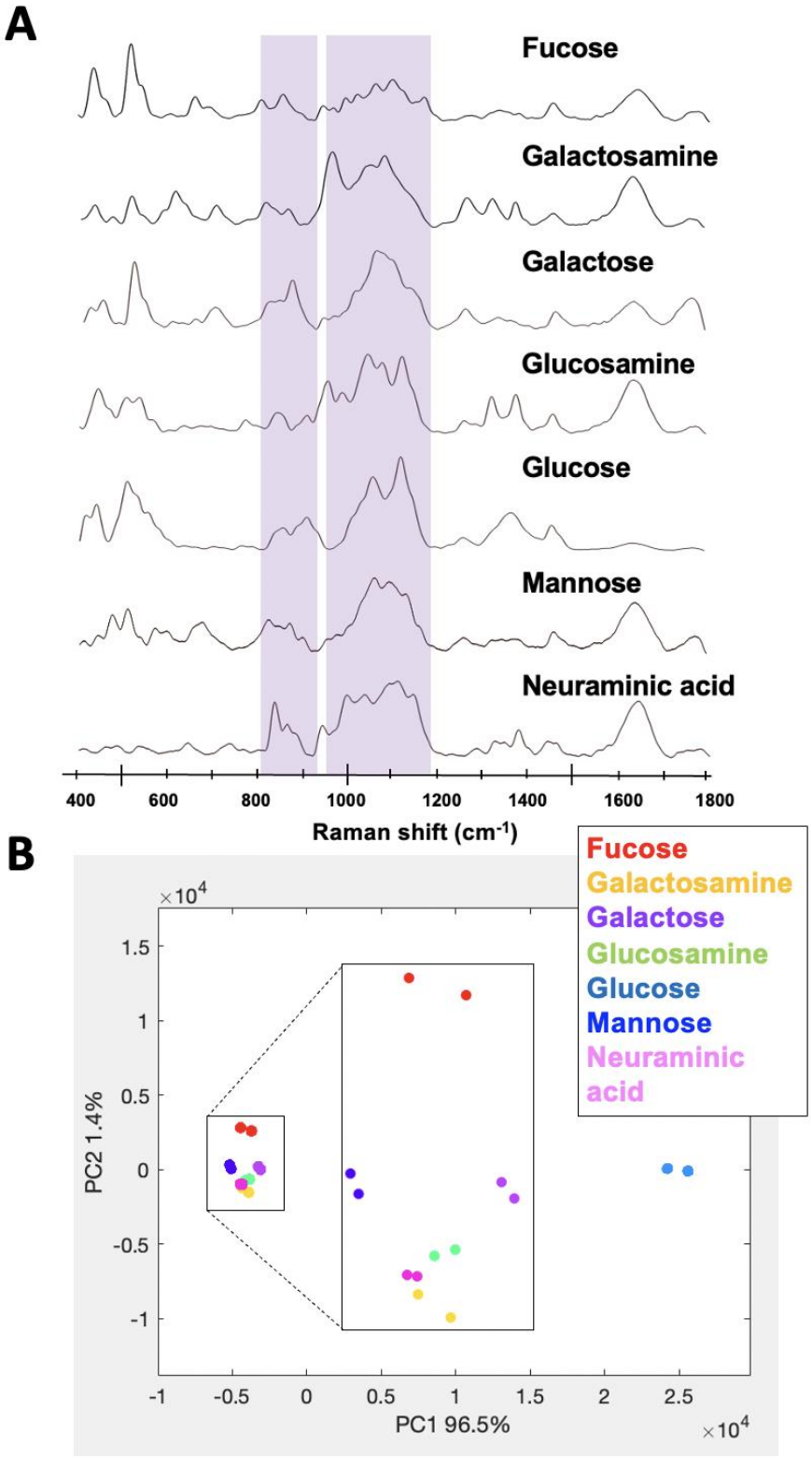


Figure 4.4 Representative signatures of glycans and their clusterization after PCA. A) Spectral regions highlighted in pink give strong Raman intensity and are shared by the different glycan signatures. **B)** The different glycans (2 acquisitions per glycan) clusterize clearly after PCA.

4.3.2 Discrimination of glioma cell lines from different glioma patients and grades

To investigate the changes observed between grade III and grade IV at the cellular level, different *in vitro* cellular models (monolayers and spheroids) were used. Specifically, this study aims at assessing whether grade III and grade IV viable cell lines could be discriminated, whether there is a difference in spectra between single cells and multicellular spheroids, and, further, whether biological changes between the different types of cells could be monitored by RS. Therefore, grade III (SW1088) and grade IV (T98G and A-172) cell lines were grown in monolayers and spheroids in the same growth media (DMEM). Cells were seeded directly on a stainless-steel slide at the bottom of a petri dish containing the growth medium. Cells were able to grow on the stainless-steel slide, meaning that the acquisitions could be performed directly on the cells adhering to the steel substrate, which permits a good signal-to-noise ratio.

Spheroids were first generated with the hanging drop method, then collected and dispensed in a stainless-steel dish before acquisitions. Both cell monolayers and spheroids were washed once with PBS before acquisition and a low volume of PBS was left to cover the cells that needed to be enough to avoid dryness. However, the volume of PBS needed to stay low to avoid interference with the laser and so that the laser could be focused on the cells easily. The method was optimised a few times before collecting the final data. Representative spheroids images for each cell line are shown in **Figure 4.5A**. For each cellular model, different cells or spheroids were targeted once with RS microscopy (**Fig. 4.5B**).

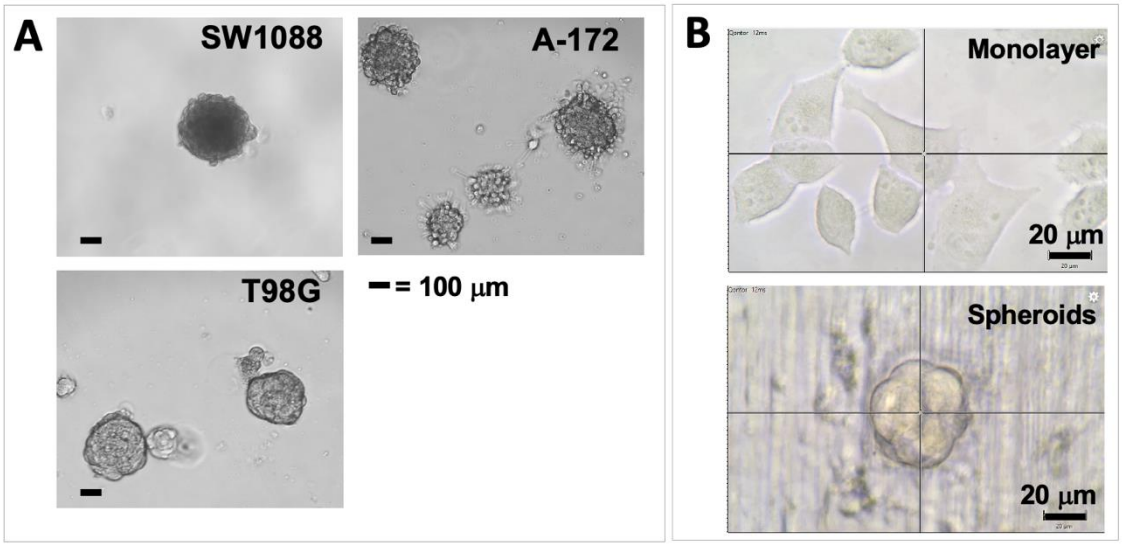


Figure 4.5 Visualisation of cellular models. (A) Images representing the spheroids generated by the hanging drop method for the three cell lines using standard inverted microscopy**,** total magnification, x200. **(B)** Images representing cells grown in monolayers and in spheroids under the Raman confocal microscope on the steel substrate, before acquisition.

To be able to assign peaks to biomolecular components, a general database relevant for cancer monitoring was extracted from the literature, to be used in addition to the glycosylation database (**Table 4.2**).

Peak (cm^{-1})	Component(s) assigned	Peak (cm^{-1})	Component(s) assigned
446-476	Cholesterol, proteins	**1095**	Mannose
498	Nucleic acid	**1120-1121**	GAG, glucose
500-550	Proteins	**1127**	Proteins
589	Amide I band	**1129**	Lipids
636-638	Tyrosine, lactose	**1131-1139**	Mannose
640-646	Proteins	**1157**	Carotenoids
670	Haemoglobin	**1206-1207**	Proteins
676	DNA	**1225**	Haemoglobin
700-703	Proteins, cholesterol	**1242**	GAG
725-729	Proteins, nucleic acid	**1230-1306**	Amide III, proteins, collagen, lipids
743-790	DNA, proteins, haemoglobin	**1300**	Proteins
818	Collagen	**1263**	Lipids
823	Tyrosine	**1313**	Phospholipids
843	Glucose	**1322**	Collagen, proteins
857	Proteins, collagen, glycans, GAG	**1331-1338**	Nucleic acid, proteins, glucose
870	Glucose	**1358**	Cytochrome C
880-890	Tryptophan, galactosamine	**1365**	Tryptophan
910-911	Glucose	**1378**	DNA
933	Proteins	**1380**	Glucose
935-940	Proteins, collagen	**1402**	Phospholipids
959	Proteins	**1441-1445**	Collagen, phospholipids
997	Glucose	**1450**	Proteins
1003-1005	Phenylalanine, collagen	**1521**	Carotenoids
1000-1200	Glucose	**1541-1546**	Amide II, cytochrome C
1032	Proteins	**1573-1585**	Nucleic acid, proteins, haemoglobin
1064-1068	Lipids	**1602-1607**	Proteins
1074	Phospholipids, collagen	**1654-1657**	Amide I, collagen, lipids
		1732	Lipids

Table 4.2 General Raman database from the literature. (Arboleda *et al.* 2000, Feng *et al.* 2010, Kast *et al.* 2014, Lin *et al.* 2014, Wiercigroch *et al.* 2017, Abramczyk *et al.* 2018, Aubertin *et al.* 2018, Galli *et al.* 2019, Zhang *et al.* 2020, Riva *et al.* 2021).

The cell lines grown in monolayers (**Fig.4.6A and C**) or spheroids (**Fig.4.6B and D**) were first compared using PCA. For single cells grown in monolayers, the three largest PCs explained together 67.6% of the variance between the samples. The discrimination rate between the cell lines was 85% when using a 5-fold cross-validation (linear discriminant analysis). For cells grown in spheroids, the first three PCs explained 77.3% of the variance and the discrimination rate between the cell lines was 68.2% (up to 79.5%, quadratic discriminant). The mean spectra of grade III and grades IV cells were compared for both monolayers (**Fig.4.6C**) and spheroid models (**Fig.4.6D**). To generate the grade IV mean signature, all the grade IV cells (from the two different cell lines) were averaged. The intensities were rescaled between 0 and 1. Between 400 and 900 cm^{-1}, grade IV cells displayed much higher overall intensity in both models. The higher intensity observed in grade IV (GBM) cells in this region could reflect a higher level of protein content since this region is widely associated with proteins (446-476, 500-550, 640-646, 700-703, 725-779, 859 cm^{-1}) as well as DNA (498, 676, 725-729, and 743-790 cm^{-1}).

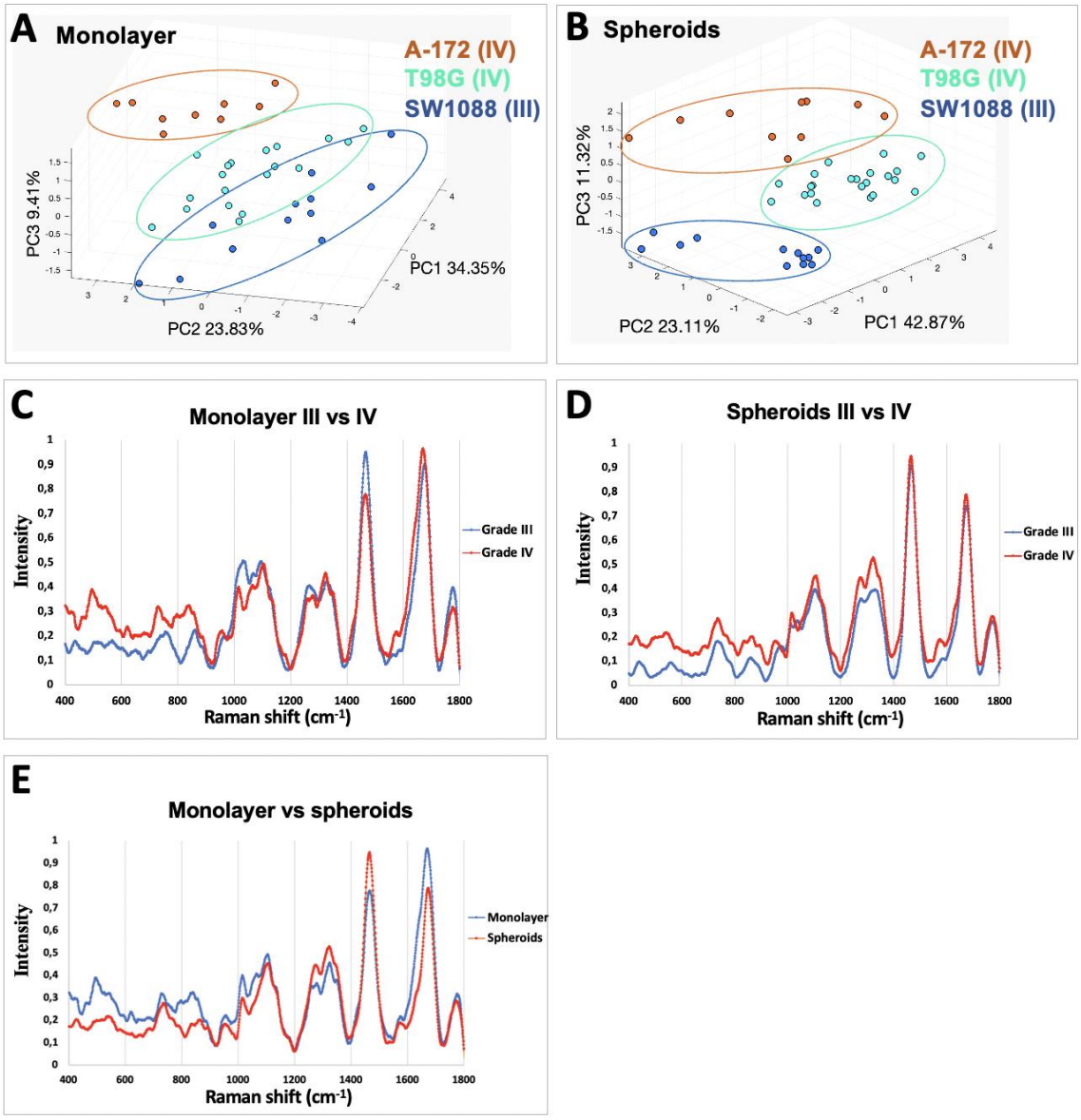


Figure 4.6 Discrimination in glioma cells. (A) PCA plot of the individual cells (one dot represent one cell) grown in monolayer using the three largest PCs. **(B)** PCA plot of all the individual spheroids using the three largest PCs. **(C)** Pair-wise comparison between the mean spectra of grade III and grade IV individual cells. **(D)** Pair-wise comparison between the mean spectra of grade III and grade IV spheroids. **(E)** Pair-wise comparison between the mean spectra of grade IV monolayer and grade IV spheroids. Circles drawn on PCA plot highlight trends assessed subjectively by eye.

For monolayers, within the 1000 and 1100 cm^{-1} region, which is part of the glycan-shared signature, a decrease in intensity was observed in grade IV compared to

grade III, as well as in the 1760-1775 cm^{-1} region, assigned to galactose. Finally, the augmentation in the 1400, 1610-1660, and 1700-1750 cm^{-1} could easily correspond to an increase in the lipid composition.

Targeting the spheroids, the signature was similar between 1000 and 1800 cm^{-1} with some differences. An increase in the 1200-1400 cm^{-1} region in grade IV compared with grade III was observed, which could be assigned to lipids and the amide III vibration of proteins, while the difference of this region in the monolayer model was less marked. The decrease of the glycan region in grade IV was however less pronounced in the spheroid model, the signature of the two grades in this region was instead very similar. Overall, the grade IV signature was higher in intensity than the grade III in the spheroid model, except in the glycan-rich region where the signatures seemed to overlap.

The grade IV signatures from the monolayer and spheroids were also compared (**Fig.4.5E**). The intensity given by the spheroids was lower between 400 and 1200 cm^{-1} in regions largely assigned to proteins and glycans, while the intensity was higher in the 1200-1400 cm^{-1} region, corresponding to lipids, DNA, and amide III.

4.3.3 Identification of MUC4-induced glycosylation in GBM spheroids

Next, another grade IV cell line (LN-18) was used to further study the impact of MUC4-induced glycosylation increase on the Raman signature. The Raman acquisitions were collected and analysed as previously described on two groups of LN-18 spheroids, control LN-18 spheroids, that were formed from LN-18 cells transfected with the empty plasmid (pCTRL); and MUC4-overexpressing LN-18 spheroids, formed from LN-18 transfected with the MUC4 plasmid (whose efficiency was checked previously by RT-qPCR). Spheroids were used over traditional cell cultures, as they represent a multicellular model with cellular interactions in all three dimensions.

MUC4 is expected to be expressed at the surface of the cells, potentially highly affecting cellular interactions and the level of glycosylation at the cell surface. This study aims at evaluating to which extent the Raman is a sensitive enough tool, able

to detect a slight change due to the dysregulation of one single protein among several thousands. MUC4 is a heavy and highly glycosylated protein with potential impact on the expression of other proteins, as well as cell surface organisation, proliferation, and spheroid size. The transfection performed here led to a high overexpression of MUC4 protein. Therefore, capturing the change is expected to be challenging but still conceivable with the appropriate analysis. Finally, this study may be useful to confirm that glycosylation changes will mostly account for a spectral change localised between 950 and 1200 cm^{-1}, a region particularly assigned to glycans.

The results are presented in **Figure 4.7**. The seven largest PCs are usually considered enough to explain the variation between the samples in a PCA (Chen *et al.* 2015). Since this time, the difference was expected to be significantly more subtle than previously, different combinations of all seven largest PCs were tested. The first largest PCs did not seem to give clusterisation of the two groups (pCTRL and pMUC4 spheroids), only the fifth PC (PC5), explaining only ~3% of all variances between the observations, seemed to give a good clusterisation of the two groups visually (**Fig.4.7A**). The most obvious clusterisation was obtained by plotting PC3, PC5, and PC7 (**Fig.4.7B**), but all 3D plots containing PC5 was enough to separate the two groups (data not shown). When comparing the averaged spectra of the two groups (n=13 spheroids averaged in each group), the two signatures were very similar, with the exception that pMUC4 gave higher intensities in the 950-1200 cm^{-1} region compared with pCTRL. Since the difference studied here is expected to be subtle and that only the fifth PC gave a visual discrimination, standard deviation throughout the spectrum was examined. The standard deviation in the 950-1200 cm^{-1} region was relatively high and not exactly consistent between both groups (**Fig.4.7C**). Peaks at 1280 cm^{-1} and 1470 cm^{-1} also showed a slight difference in intensity (**Fig.4.7C**). Therefore, unpaired t-tests for these peaks were performed to compare the intensity values between both groups. By comparing the intensities values between the two groups at 1015, 1050, 1080, 1280, and 1470 cm^{-1}, no significant difference was detected with that strategy (data not shown).

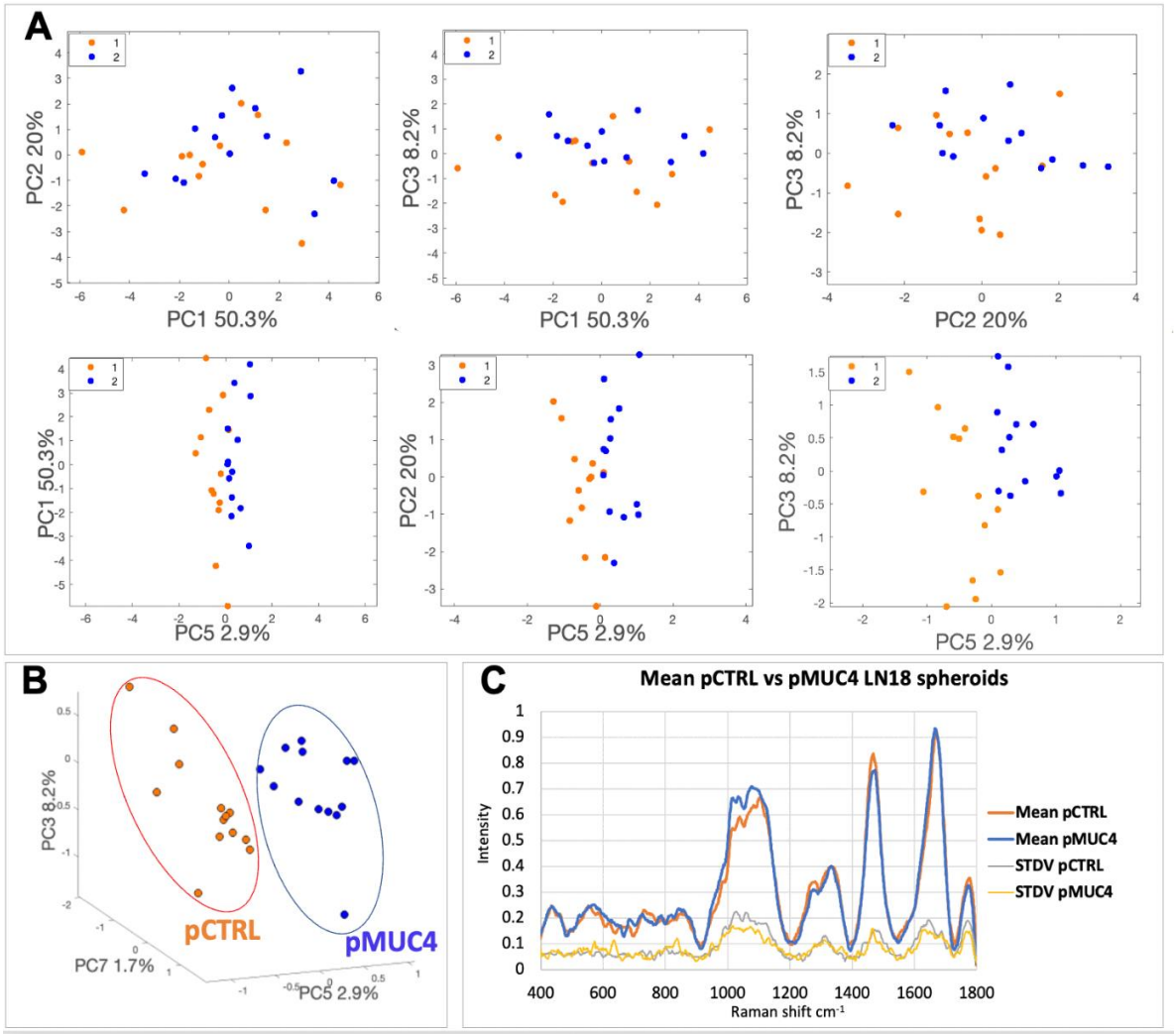


Figure 4.7 Discrimination between MUC4-overexpressing and control LN-18 spheroids. (A) PCA 2D plots using different combinations of PCs are represented for the comparison between individual spectra obtained from LN-18 spheroids. Percentages of the variance explained by each PC is represented on the axes. One dot represents one acquisition obtained from one spheroid. Comparison is made between spheroids formed from LN-18 cells transfected with MUC4 plasmid (pMUC4, in blue, n=13) and spheroids formed from normal LN-18 cells (pCTRL, in orange, n=13). The first largest PCs, accounting for the highest variance between the acquisitions, are not enough to separate pMUC4 cells and pCTRL cells. PC5 (2.9% of the variance) discriminate successfully between the two groups of spheroids. **(B)** 3D PCA plot using PC3, PC5, and PC7 gives the best visual clusterisation between the two groups. **(C)** Mean spectra of both groups are compared with the standard deviation values represented (STDV).

PCA loadings were therefore plotted across the spectrum (**Fig.4.8**) to identify what variables (Raman shifts) could be associated to PC5. PCA loading is a representation that allows the visualisation of the contribution of each PC across the variables (across the spectrum). If the loading value for a particular peak is highly positive or negative, then fluctuation in that peak contributes widely to the PC. The largest PC (PC1), which was identified as not being responsible for the discrimination of the two cell types, seemed to account, at least partly, for the difference observed at 950-1200 cm^{-1}. This suggests that the high difference in this region observed in this particular comparison was mostly due to noise, and other variations between the cells that were not due to the presence of MUC4. PC5, identified as the variation of interest between the absence and the presence of MUC4, seemed to be very dispersed across the entire spectrum, making interpretation complex. For that reason, all Raman shifts that correlated with an important contribution of PC5 (reaching at least 0.03 or -0.03 of loading counts) were aligned with the spectrum to make the visualisation easier (**Fig.4.8**). The most important differences identified were in the amide I and amide III peaks (1650 and 1230-1300 cm^{-1}), which correspond to a change in the peptide backbone structure, and therefore a change in secondary structure of proteins. This aligns with the fact that the two groups of cells should differ mostly by a change of MUC4 protein level expression at the surface of the cells. Other changes were more subtle. To a relatively small extent, the 950-1200 cm^{-1} region could account for the cell discrimination, more specifically the peaks at 1042, 1100, and 1165 cm^{-1}. From the database used previously, these peaks could correspond to a contribution of glycans and proteins. Other peaks could correspond to glycans (420, 660, 700, 880, 1400 cm^{-1}). The peaks at 700 and 1042 cm^{-1} could also reflect a change in proteins, while the peak at 660 cm^{-1} could be associated to a change in DNA. Finally, the 1270, 1680, and the 1042 cm^{-1} peaks have also been associated to a rearrangement of lipids.

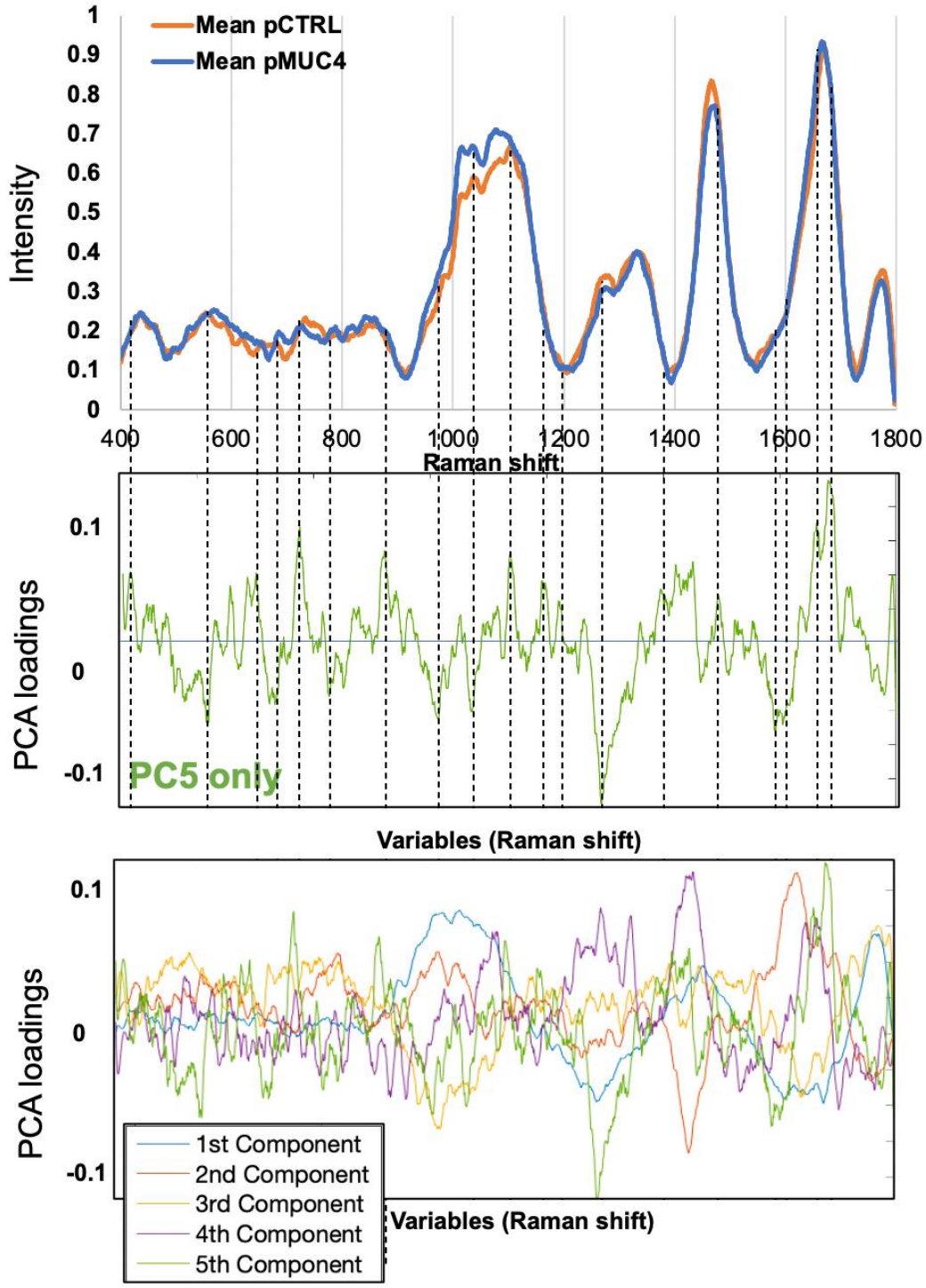


Figure 4.8 PCA loadings plot of comparison between MUC4-overexpressing cells and control cells. PCA loadings of PC5 only and all largest PCs are represented and compared with the spectral comparison between the two types of cells. For PC5, the most important changes are in the 1650 and 1230-1300 cm^{-1} peaks, which correspond to amide I and amide III and therefore a change in protein level.

4.4 Discussion

4.4.1 Discrimination of glioma cells

In this study, the ability of RS at discriminating grade III and grade IV glioma at the cellular level was investigated. Different cell lines from grade III and IV glioma patients could be discriminated by PCA when using monolayer and spheroids with good accuracy, indicating that grade discrimination may be achievable on single live cells with one single acquisition. However, repeating this approach on a higher number of cell lines or patient-derived cell lines could be more representative. Both methods gave similar but different biomolecular signatures for GBM cells. An increase of DNA, proteins, and lipids was observed in grade IV compared to grade III in both models. There was a decrease of glycans in grade IV compared to grade III, which was more pronounced in the monolayer model. However, in the spheroid model, the overall intensity of the grade IV signature was higher than the one of the grade III signature, with the exception that the glycan-rich region (950-1200 cm^{-1}) was instead very similar.

As expected, targeting a single cell or a multicellular spheroid did not provide the same biomolecular signature reflecting their differences in cellular metabolism and other biomolecular pathways. These results suggest that targeting a single cell or whole tissue for cancer diagnosis might lead to different clinical predictions. This result can be explained considering the complex tumour microenvironment of the tissue compared to the single cell. The increased intensity in the spheroid grade IV signature could reflect an increased level of lipid, DNA, and amide III explained by the targeting of several nuclei by the laser but also a change of lipid and protein configuration due to cellular interactions. More importantly, targeting single cells with the proposed RS approach could provide an additional benefit to the current single-cell technologies in cancer personalised medicine.

4.4.2 Detection of MUC4-induced glycosylation in GBM spheroids and limitations

Discrimination between MUC4-overexpressing spheroids and control spheroids was then performed. Only the fifth PC, explaining around 3% of the overall variation, gave a clear clusterisation of the two groups, which implies that the largest PCs

(PC1 to PC4, explaining together ~83% of the overall variation) did not account for significant difference between the two groups, since *a priori*, the only significant difference between the two groups is the induced MUC4 overexpression that was used to generate them. One explanation could be that the difference is so subtle that the signal corresponding to this difference is covered by the noise, higher in intensity. For the comparison at the spectral level, it is thus impossible, since no difference was significant by using t-tests, to be certain that the relevant difference (PC5) corresponds to a specific region of interest, but rather to an accumulation of small differences spread across the spectrum, as confirmed by the PCA loading plot. Among these small differences, changes in glycosylation could be identified and the most important differences identified could be assigned to a change in the protein level (amide I and amide III peaks).

In this chapter, a number of decisions were made to overcome potential limitations usually encounter with this technique. In addition to the choice of the substrate that was discussed before, the cells were targeted while being viable on the substrate with only one acquisition per cell. This decision was made to avoid the inclusion of acquisitions on cells that could have been altered by the laser exposure. Indeed, since high spectral resolution and therefore long acquisitions were used, multiple acquisitions on the same cells could have led to cell damage. It is yet not clear to what extent long or repeated laser acquisitions could damage the cells. RS is a technique that does not lead to sample destruction, although, the impact of long laser exposure on a live cell have not been extensively studied. Since viability assessment can only be performed by eye under the Raman microscope during the process, it was decided here to only target each cell once. This technique was sufficient to obtain very good discrimination between the different cell lines. Nonetheless, better discrimination between significantly different cell types, or identification of smaller difference between cells might be improved by multiple acquisitions. It might be thus important in the future to assess the exact impact of laser exposure on live cells.

Finally, it is important to point out that the application of RS in the biological and medical field remains under relatively recent investigation with the number of publications each year increasing rapidly. RS was originally not designed for

biological application. Therefore, many of RS studies have focused on proof-of-concept and optimisation, rather than using RS as a tool to answer biological questions. This leads to several limitations when interpreting these results, especially in terms of identifying biological components. First, the method and analytical processes are inconsistently used between different laboratories and one can observe the heterogeneity in the choice of materials used (substrate, types of machines…), protocols (number of repeats, enhancement technique…), and the numerous statistical analyses available to extract differences. Despite this inconsistency, the reproducibility of the results is relatively preserved since different studies led to the same conclusions and that different machine learning processes successfully discriminated the investigated samples. This supports the power of RS in biological and medical investigations.

However, a bias in the peak assignments is still present. Unlike other comparable techniques, such as mass spectrometry, there is still no official Raman database that would permit an automatic and objective peak assignment. This yields a relative freedom in the biological interpretation of these changes that renders the process quite complex, with the need of good knowledge of the previous studies, and therefore potentially biased. This is exacerbated by the fact that different components overlap the same Raman shifts. In this study, an in-house database with a particular focus on glycans was used to answer this challenge. This represented an asset since glycans seem to be - more than other biocomponents - overlooked in the literature. However, this could also have accentuated a bias toward glycan identification, a limitation that was partly bypassed by the use of a more general database extracted from previous studies.

4.4.3 Conclusion

To conclude, Raman spectroscopy (RS) is showing promise of high sensitivity in the discrimination and clusterization of cancer cells, and PCA allows the extraction of relevant differences. However, subtle alterations probably require a very high number of observations so that the relevant variable of interest can overcome the noise in intensity. Moreover, the current lack of a universal Raman database hampers objective biological interpretation. This study contributed to the understanding of the RS sensitivity and limit of detection in the biological field

application, using a reasonable acquisition time. It can be concluded that a long acquisition time would be necessary to detect subtle changes. This would represent technical challenges at the cellular level since, contrarily to fixed tissue or freezable samples (serum), whether it be live cells, tissue, single cells, or real-time biopsy, multiple acquisitions would have to be performed quickly and so the number of repeats will constitute a limit. As for acquisitions performed on cellular models, such as in this study, the number of acquisitions possible on a short time period (necessary for rigorous comparison with RS) is limited by different factors: the complexity of spheroids formation, the life expectancy of the cells under the Raman microscope and laser, and the time length of manual acquisitions. Improvement of RS specifically intended for clinical use may overcome this problem, for example, intelligent automatisation of laser focus between different localisations and more rapid precise acquisition ability.

Chapter 5: Monitoring biomolecular signatures with a focus on glycosylation, using Raman spectroscopy in glioma tissue and serum

ABSTRACT

Background: Raman spectroscopy (RS) may lead to automated cancer diagnostics but also provides information about changes in the bio-composition of the sample. Aberrant glycosylation is one of the most important post-translational modifications in brain cancer cells and can impact several steps in tumour progression. Cancer-associated aberrant glycosylation represents an opportunity for cancer monitoring and was captured successfully with different accuracies by RS in the previous chapter. This chapter aimed at testing the efficacy of RS at classifying glioma clinical samples, with a particular focus on monitoring the changes in the glycosylation pattern.

Method: FFPE tissue sections and serum samples from glioma patients diagnosed with different glioma grades were interrogated using RS and classification was tested by using PCA combined with classification learning methods. In addition, Raman spectra were compared with the in-house glycosylation database generated for the most common glycans found on mucins, as well as with the broader literature-based database.

Results: RS combined with PCA and machine learning classified glioma grades from dewaxed FFPE tissue with an 85% accuracy and was even successful at discriminating grade III and IV malignant glioma samples with an accuracy of 90% using a drop of liquid serum. Changes in the glycosylation pattern seemed to largely account for the successful discrimination between the different grades in the clinical samples using the in-house glycosylation database. Peaks assigned to glycans could be used individually to show significant difference. The serum method presented several advantages compared to the diagnostic method using tissue, which included better discrimination but also, significant reduction of the acquisition time.

Conclusion: Glycosylation-based RS monitoring could be used to classify glioma grades in patient samples, including serum samples that are collected in a non-invasive way.

5.1 Introduction

Grade III, and IV gliomas remain challenging to discriminate as they share overlapping characteristics (Wood *et al.* 2019). The diagnosis currently lacks objectivity as it relies partly on the investigator's eyes and pathology staining methods; thus, molecular biomarkers and/or more objective and label-free imaging techniques that rely on the biological composition would help improve glioma accurate diagnosis (D'Amico *et al.* 2017). Raman spectroscopy (RS) presents many advantages that makes it a good candidate diagnostic tool to be routinely used: 1) it does not lead to sample destruction; 2) it does not require any external labels (label-free technique); 3) unlike infrared spectroscopy, it is not influenced by water molecules and hence, can be used to study aqueous solution; and finally, 4) the sample preparation is not demanding (Larkin 2011, Auner *et al.* 2018).

RS has been used in cancer diagnosis as it can generate a specific spectral signature for any tissue and capture any subtle changes in their composition. Cancerous tissues and their healthy counterparts produce different Raman signatures as they do not contain the same composition in biomolecules (Auner *et al.* 2018). RS can have different applications in cancer diagnosis (Krafft *et al.* 2012). First, it can be used to objectively and rapidly predicts the type and the grade of a cancerous tissue (Crow *et al.* 2003, Kopec *et al.* 2021, Mamede *et al.* 2021). Normal brain, meningioma, glioma, and brain metastasis have been discriminated from dewaxed FFPE tissues (Gajjar *et al.* 2012). In addition, RS can be used for application in liquid biopsies: discrimination between low-grade (I and II) and high-grade (III and IV) glioma has been suggested (Chenxi Zhang 2020), however, the accuracy and the approach remain to be optimised, more particularly, the discrimination between grade II and III and between grade III and IV remains to be demonstrated.

In addition to diagnostic improvement, RS can be used to assess general changes in the bio-composition throughout cancer progression. Raman studies investigating the differences between GBM and healthy brain tissue have shown that brain cancer tissue had altered collagen and lipid levels compared to normal brain. Haemoglobin has also been found to be increased in GBM, which is in accordance with the high perfusion rate observed in this type of cancer (Krafft *et al.* 2012). More recently, RS

has also been used to monitor post-translational modification (PTM) on proteins (Ma *et al.* 2020). PTM pattern of proteins is highly altered in cancer compared with healthy tissues. Glycosylation pattern in the brain is important and is expected to be highly modified in cancer. Indeed, half of all proteins found in the brain are glycosylated with impact on their functions and interaction with other proteins. Glycosylation is involved in cell motility, cell-to-cell adhesion, and cellular signalling. Glycans (the carbohydrate portion of glycoproteins) can be O-linked (to serine and threonine) or N-linked (to asparagine) and different transferases are involved in these modifications (Veillon *et al.* 2018).

Proteins found dysregulated in cancer tissue and secreted in the serum are very often glycosylated and are therefore being considered as potential biomarkers and overall, glycans are currently used as suggested biomarkers in cancer (Silsirivanit 2019). There are currently only a few glycosylation-based cancer biomarkers and most of them are mucins (**Fig.5.1**) (Tabarés *et al.* 2006, Felder *et al.* 2014, Kudelka *et al.* 2015, Hanson *et al.* 2016, Munkley 2019, Manne *et al.* 2021). However, these biomarkers tend to be present only at low level in the serum since they do not always have a high secretion rate. Therefore, detection of the overall glycosylation changes by RS would be beneficial to monitor disease progression and to understand metabolic changes happening during cancer progression with more significance. In the preceding chapter, it was found that changes in the glycan pattern found on mucins and other glycoproteins as well as the more subtle changes induced by overexpression of MUC4 in glioma cells, could be captured and exploited with RS. The impact of laser exposure on the data interpretation in fixed tissue and serum samples is expected to be lower than in live cells.

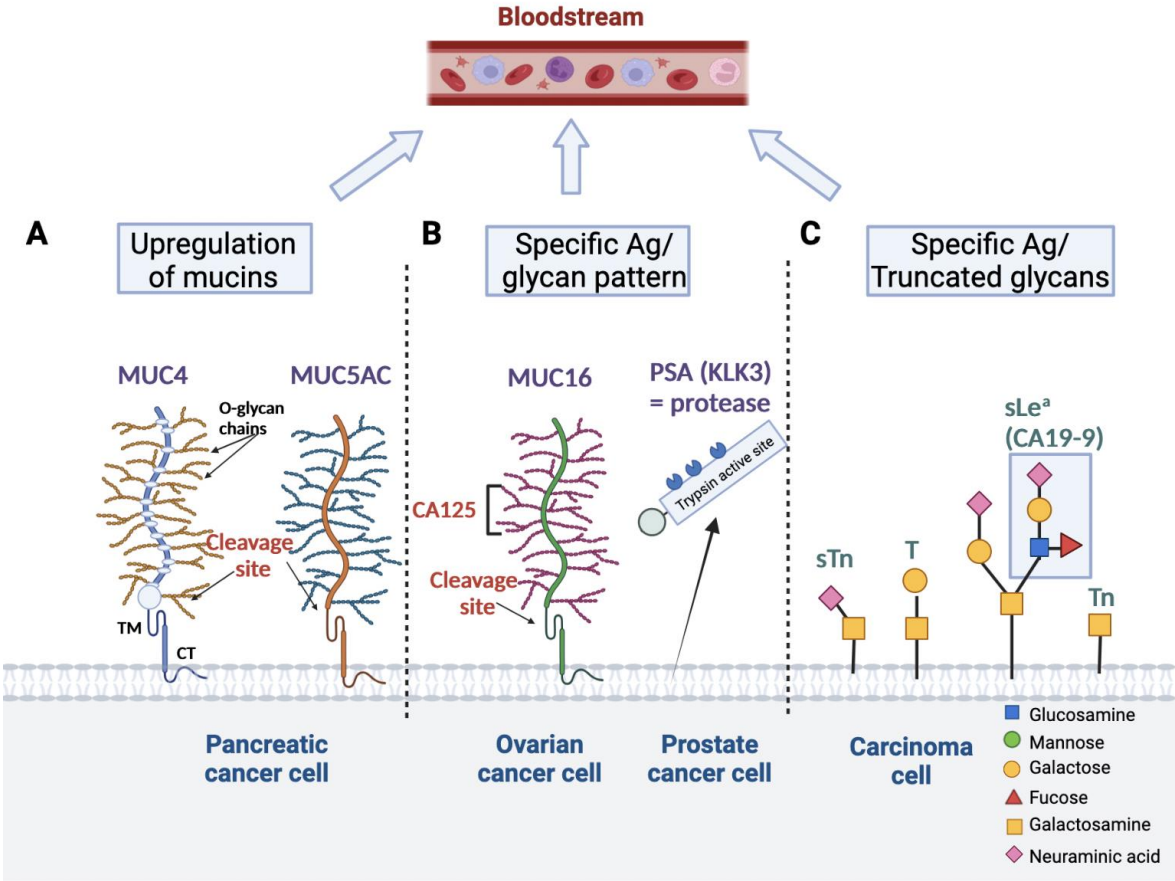


Figure 5.1 Glycosylation-based tissue and serum cancer biomarkers. Glycoproteins expressed in cancer cells, which can be then present in the blood serum after secretion or cleavage, have the potential to serve as liquid biomarkers to monitor cancer progression. **(A)** MUC4 and MUC5AC are heavily glycosylated transmembrane proteins that contain a cleavage site and are overexpressed by pancreatic cancer cells. **(B)** The transmembrane MUC16 and the secreted Kallikrein 3 both contain a specific glycan antigen (CA125 and PSA, respectively) that have been shown to correlate with ovarian and prostate cancer aggressiveness, respectively. **(C)** Carcinoma cells overexpress specific truncated or aberrant glycans that can be detected in the serum and serve as cancer biomarkers, such as the CA19-9 antigen. [Antigen (Ag), Kallikrein (KLK)]. (Tabarés *et al.* 2006, Felder *et al.* 2014, Kudelka *et al.* 2015, Hanson *et al.* 2016, Munkley 2019, Manne *et al.* 2021). *Figure created with Biorender.com.*

For these reasons the same approach tested earlier was applied on clinical samples. The aim of this chapter was to test the ability of RS to accurately

discriminate glioma grades from FFPE tissue and blood serum, and more specifically the malignant grade III and grade IV, which remain challenging to discriminate. Successful discrimination between different types of brain cancer, and between normal brain and brain cancer (Gajjar *et al.* 2012) or benign and malignant glioma has been previously achieved on dewaxed fixed tissue (Zhou *et al.* 2019). However, to consider RS as a valuable diagnostic tool in clinics, its ability to discriminate with very good accuracy grade III and grade IV glioma with solid biopsies, but also potentially with less invasive liquid biopsies, needs to be clearly demonstrated. In parallel, changes in the bio-composition of the samples during glioma progression were explored using the in-house glycosylation database generated for this thesis, as well as the reference database containing different types of components from the literature.

5.2 Materials and methods

5.2.1 Rationale and methodology overview

The approach of the previous chapter was applied on clinical samples (fixed tissue and serum). Eventually, the method was designed to help answering the following two main questions:

1) Whether RS is an objective tool for glioma grade discrimination, and most specifically for discrimination between grade III and IV (and/or the two IDH1 genotypes) from glioma samples.

2) Whether changes in the glycosylation pattern commonly found on mucins and other glycoproteins accounts for this discrimination.

The methodology overview of this chapter is presented in **Figure 5.2**.

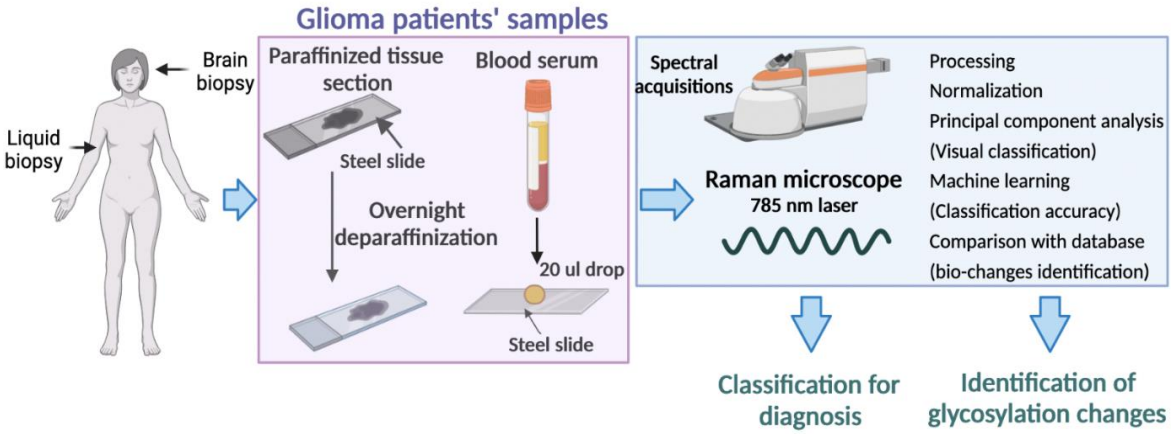


Figure 5.2 Methodology overview of RS on tissue and serum samples. Deparaffinised FFPE tissue sections (brain biopsy) and blood serum (liquid biopsy) from glioma patients were interrogated with a Raman confocal microscope. After processing, normalisation, principal component analysis (PCA), and classification learning methods, the study tested if the different glioma grades could be discriminated for diagnostic purposes. Spectral changes between the grades were assigned to biological changes, by comparison with the in-house glycosylation database for glycosylation patterns and with the general biomolecular database extracted from the literature. *Figure created with Biorender.com.*

5.2.2 Patients and clinical samples

Serum samples were stored at −80 °C prior to analysis. FFPE samples, cut on stainless-steel slides (Renishaw) for Raman acquisitions, and on glass for parallel histological analysis, were obtained from tumour debulking surgery or biopsy. 30 glioma FFPE slides (10 grade II, 10 grade III, 10 grade IV) and one malignant melanoma slide, were obtained from NovoPath Biobank Newcastle (Newcastle-upon-Tyne, UK). Grades for each case were determined by a neuropathologist beforehand, following the 2016 WHO histological classification. The grade II and III patients from the tissue sample set were diagnosed with astrocytic glioma (respectively diffuse astrocytomas and anaplastic astrocytomas, but no oligodendrogliomas). 30 blood serum samples (10 non-glioma benign tumour (controls), 10 grade III, 10 grade IV) were obtained from the MCRC biobank (Manchester, UK).

5.2.3 Ethical approval

The Research Ethics Boards approved sample collection of the respective Biobanks (Manchester Cancer Research Centre (MCRC) (REC Ref 18/NW/0092) and NovoPath Biobank Newcastle (REC Ref 17/NE/0070), UK) and the Teesside University Research Ethics Committee upon receipt of the ethical approval. Informed consent was collected for each patient and all procedures followed the Declaration of Helsinki.

5.2.4 Serum and tissue acquisitions

For measurements on glioma tissues, the FFPE stainless-steel slides were first chemically dewaxed with an 18 hours-incubation in xylene, as described in previous studies (Faoláin *et al.* 2005), followed by a serial incubation in decreasing concentrations of ethanol to remove paraffin contribution as much as possible. The samples were then rinsed in distilled water and left to air-dry at room temperature for at least 30 minutes before Raman measurements were taken. 25 single spectral acquisitions were taken in each tissue sample individually at different arbitrary spots corresponding to neoplastic regions (25 replicates, 750 acquisitions for all the patients in total). To monitor the laser focus between each acquisition as best as possible, manual acquisition was preferred over automated acquisition. To ensure that only neoplastic regions were targeted, comparison with corresponding glass slides stained with haematoxylin/eosin, glial fibrillary acidic protein (GFAP), and EGFR (performed previously) was made by using the same staining protocol as described previously. GFAP antibody was Sigma #G3893, produced in mouse and used at a 1:750 dilution.

For blood serum, the samples were left to thaw on ice and a 20 μl-drop was dispensed on a mirrored stainless-steel slide and was either left to air-dry for 30 minutes or directly processed while being still liquid. To verify that the spectra were correct, presence of several characteristic peaks for human serum was assessed (Ryzhikova *et al.* 2015, Parlatan *et al.* 2019). Each drop was repeatedly measured five times in different locations (5 replicates) within the centre of the drop by using the same focal distance each time.

5.2.5 Statistical analyses

Data were processed and analysed exactly as described in the previous chapter (processing, normalisation, PCA, and classification learning) with the difference that several acquisitions were made per sample. The different replicates from different locations within the sample were averaged to generate the mean spectra of each patient, which were then averaged to generate the arithmetic mean spectra of each grade before grade comparison.

To calculate the significance of the differences observed between two grades at different manually selected peaks, two-tailed unpaired t-test was performed with GraphPad Prism version 9.3 (GraphPad Prism software, California) on averaged and normalised intensities of the samples.

5.3 Results

5.3.1 Samples and patient characteristics

The FFPE tissue sections were cut on mirrored stainless-steel slides and were from 30 glioma patients (10 grades II, 10 grade III, and 10 grade IV) and the serum samples were from 30 patients (10 non-glioma controls; and 10 grade III, and 10 grade IV glioma). The serum samples were the same as the ones used in the clinical study of the present thesis (from the Manchester Biobank) and the glioma tissue sections were part of the Newcastle NovoPath set of patients used in the clinical study. The melanoma sample, used as a control, was also from Newcastle NovoPath. Sex, mean age, number of recurrent cases, and IDH1 genotypes are summarised in **Table 5.1** for all the patients interrogated with the Raman in this chapter.

Tissue	Gender		Mean age	Recurrent cases	IDH-WT	Serum	Gender		Mean age	IDH-WT
Grade II Total=10	**F**	4	47	3	1	**Control Total=10**	**F**	6	62	x
	M	6					**M**	2		
Grade III Total=10	**F**	1	54	1	4	**Grade III Total=10**	**F**	5	43	1
	M	9					**M**	5		
Grade IV Total=10	**F**	3	60	5	9	**Grade IV Total=10**	**F**	4	62	10
	M	7					**M**	6		
Total=30						**Total=30**				

Table 5.1 Tissue and serum samples, patient characteristics. [Female (F), Male (M), Wild-type (WT)].

5.3.2 Grade classification from fixed glioma tissue sections

The diagnostic power of RS was first tested using FFPE tissue sections from grade II, grade III, and grade IV patients in groups of equal size (n = 10 per grade). Manual single spectral acquisitions were performed at different spots corresponding to neoplastic tissue within the samples with a near-infrared laser. Three parallel glass sections stained with haematoxylin/eosin, EGFR, and the glial fibrillary acidic protein (GFAP), respectively, were used as guides when the neoplastic regions were challenging to identify under the Raman microscope. This staining was useful to label neoplastic cells in glioma tissues (**Fig.5.3**).

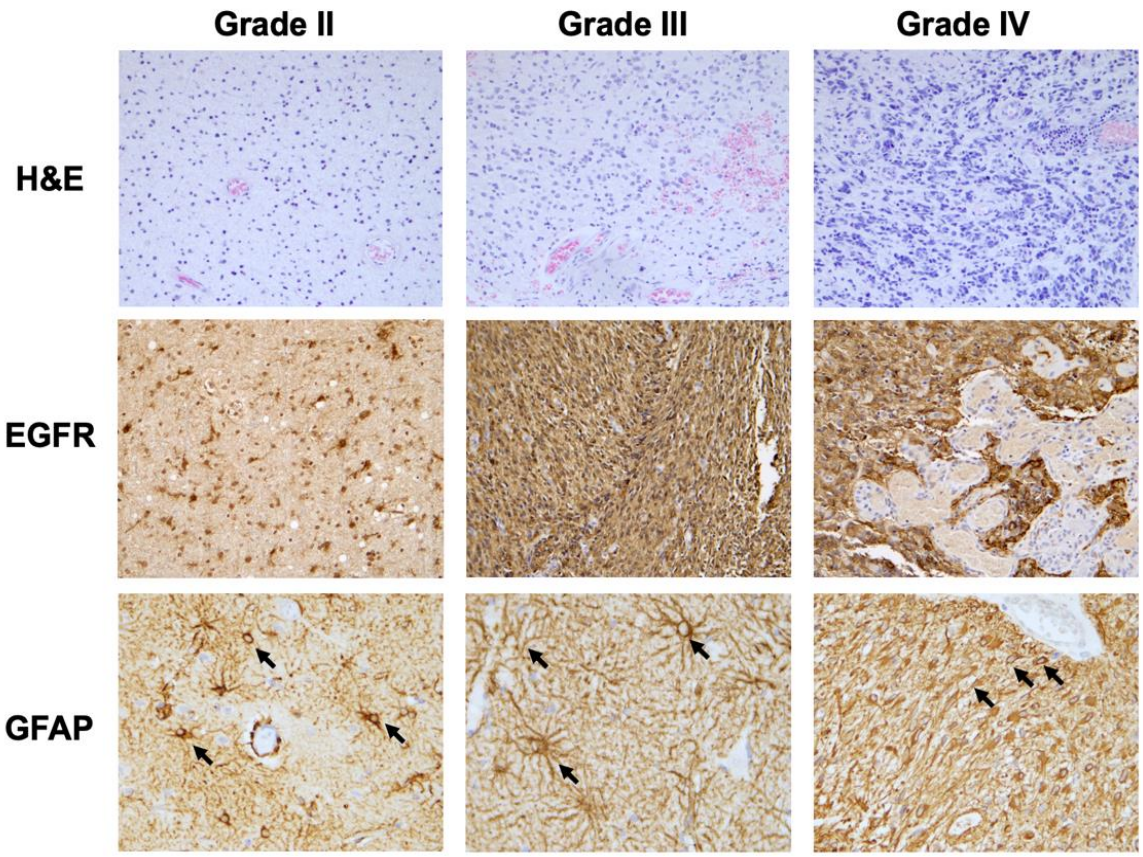


Figure 5.3 Neoplastic region labelling in parallel glass FFPE slides. Haematoxylin/eosin (H&E), EGFR, and GFAP staining were performed on three parallel FFPE slides to help identifying neoplastic regions within the sections cut on stainless-steel slides, for Raman acquisitions. The slides were stained for EGFR and GFAP with the IHC method. Black arrows indicate expression of GFAP in glial cells, revealing their shape.

Haematoxylin and eosin staining reveals the cellular density and so can be used to identify neoplastic regions (**Fig.5.3**). EGFR can also be used to label neoplastic regions since it is expressed by about 90% of glioma tumours and makes the MVP regions of GBM more visible as most of them are EGFR-negative (**Fig.5.3**). GFAP is a marker of the cytoskeleton of normal astrocytes and is also overexpressed by glioma cells. The expression of this marker reveals the morphology of the cells within the brain tissue and thus, allows the visual discrimination of normal and neoplastic glial cells. Astrocytes lose their characteristic healthy shape in GBM. Instead, GFAP expression in GBM reveals a dense network of aberrantly shaped glial cells (**Fig.5.3**) (van Bodegraven *et al.* 2019).

Visualisation of the tissue structures was possible under the microscope after dewaxing so that vessels, which are abundant in grade IV glioma, and non-cellular compartments (extracellular matrix) were avoided as best as possible (**Fig.5.4A**). Next, the method was tested by using two samples from each grade category and one melanoma sample, used as an external control to verify that a different type of cancer would give a significantly different signature. This is important since RS is not traditionally used with positive or negative controls. As the required number of repeats per samples still needed to be determined, a relatively high number of repeats (n=30 per sample) was used and averaged for each sample. Then, the average spectral signature of the two samples for grade II, grade III, and grade IV glioma, as well as the signature for the only melanoma sample, were compared. The results are presented in **Figure 5.4B** and shows that visually, the melanoma sample displayed a very different signature than the glioma signatures. The three glioma signatures (corresponding to the three grades) were very similar.

Therefore, PCA was performed with the seven samples to assess the discrimination between the sample types. The results looked promising before normalisation (data not shown). A maximum-minimum normalisation (0 to 1 rescaling) was then applied to the data and improved only slightly the clusterisation. The results are presented in **Figure 5.4C**. The melanoma sample indeed behaved very differently. The largest PC (PC1) explained 94% of all the variance between the samples and was responsible for the clear separation between the melanoma sample and the glioma samples. The glioma samples were closer to each other but the two samples from each grade clustered by pair. PC2 and PC3 were responsible for the more subtle difference observed between the glioma samples. It was therefore concluded that the method used was promising for glioma grade discrimination and that a number of 25 repeats, which reduced slightly the acquisition time, may be enough to observe the clusterisation.

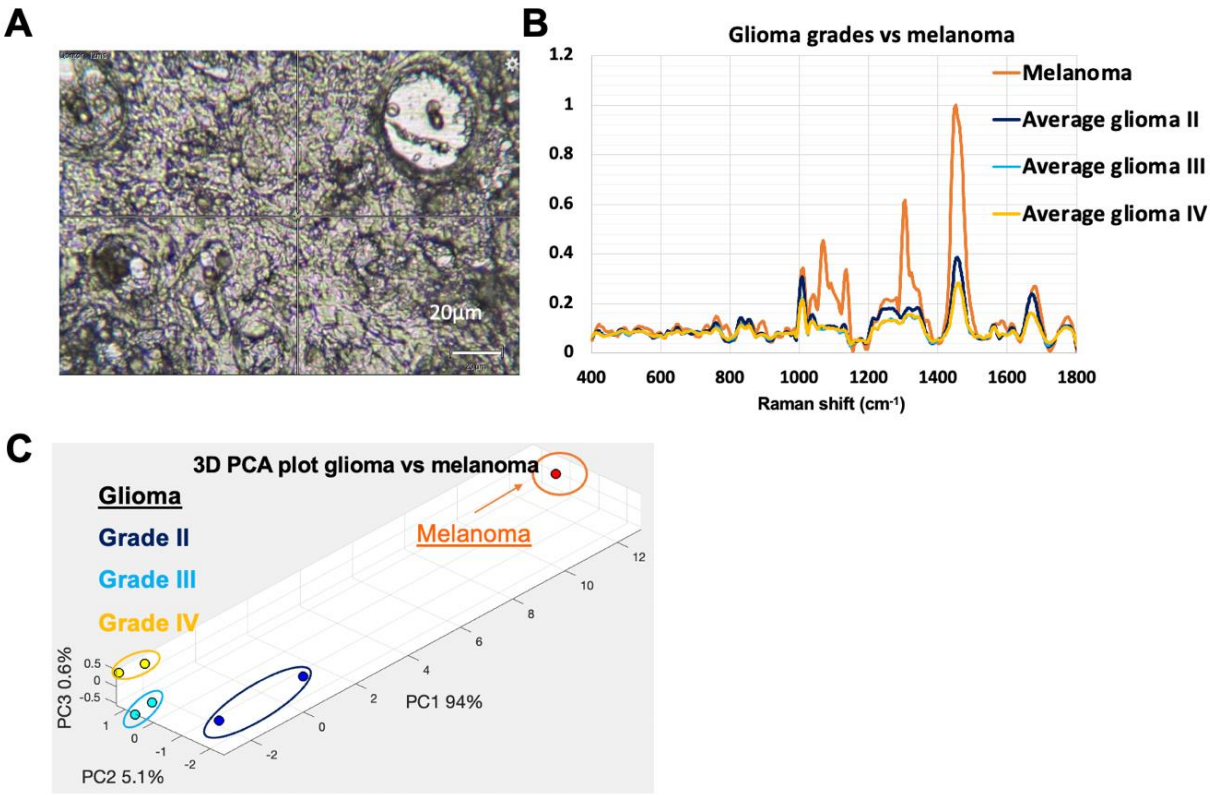


Figure 5.4 Evaluation of the method for glioma grade discrimination. (A) Representative image of one dewaxed glioma tissue visualised with the Raman microscope before acquisition. **(B)** 30 Raman acquisitions per slides were taken and averaged. The two samples picked randomly for each glioma grades were averaged and compared with the signature of the melanoma sample. The melanoma signature behaved like an outsider and was different from the glioma signatures that all looked very similar. **(C)** 3D PCA plot on all the samples (averaged and normalised data) confirmed that the Raman method was specific enough and could distinguish very clearly two types of tissue and potentially even the different glioma grades. Percentages of variance explained by each PC is indicated on the axis.

The method was therefore applied next to the entire set of glioma samples (10 samples per grade, n=25 repeats minus obvious outliers, per sample). An averaged spectrum generated from the different replicates for each patient after processing and 0 to 1 rescaling. The averaged mean spectra from all the patients are represented in **Figure 5.5** for grade II**, 5.6** for grade III**, and 5.7** for grade IV, each spectrum is the arithmetic mean of all the replicates for the patient in **A**, and the spectrum in **B** is the arithmetic mean of all the patients for the grade.

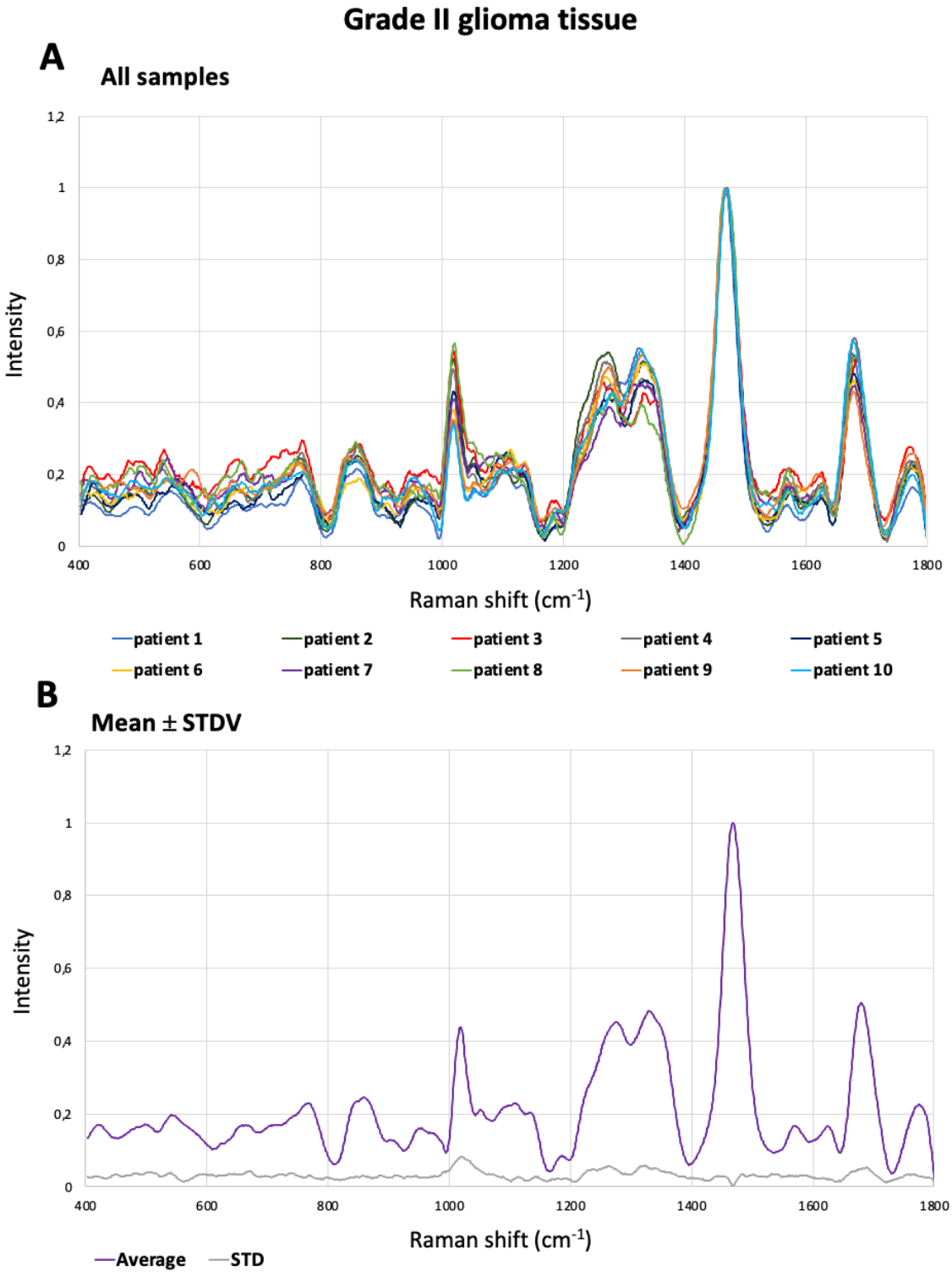


Figure 5.5 Mean tissue spectra of grade II glioma patients. (A) All mean spectra for each patient (each sample). **(B)** Mean spectra (in purple) for grade II with the standard deviation (in grey).

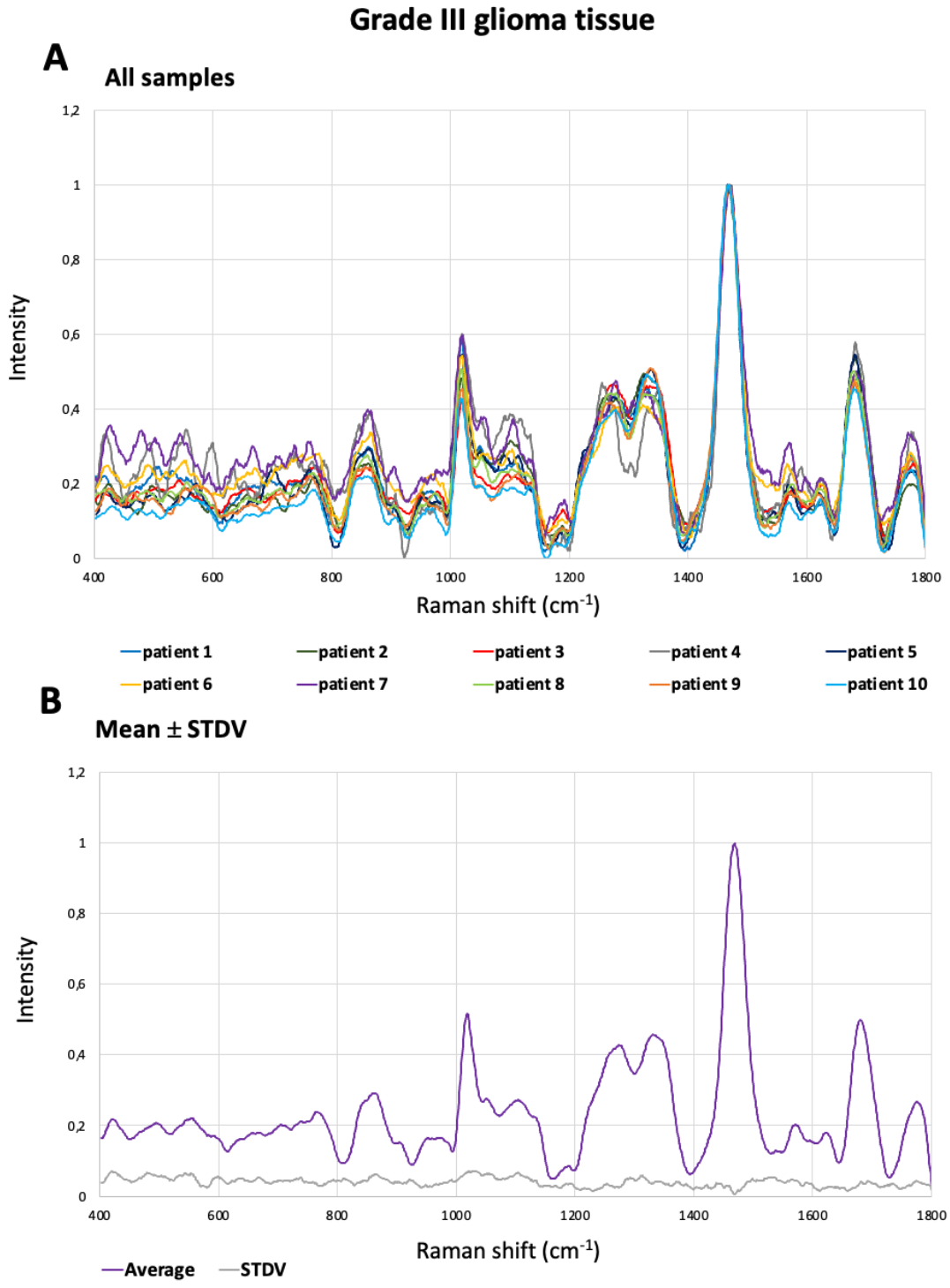


Figure 5.6 Mean tissue spectra of grade III glioma patients. (A) All mean spectra for each patient (each sample). **(B)** Mean spectra (in purple) for grade III with the standard deviation (in grey).

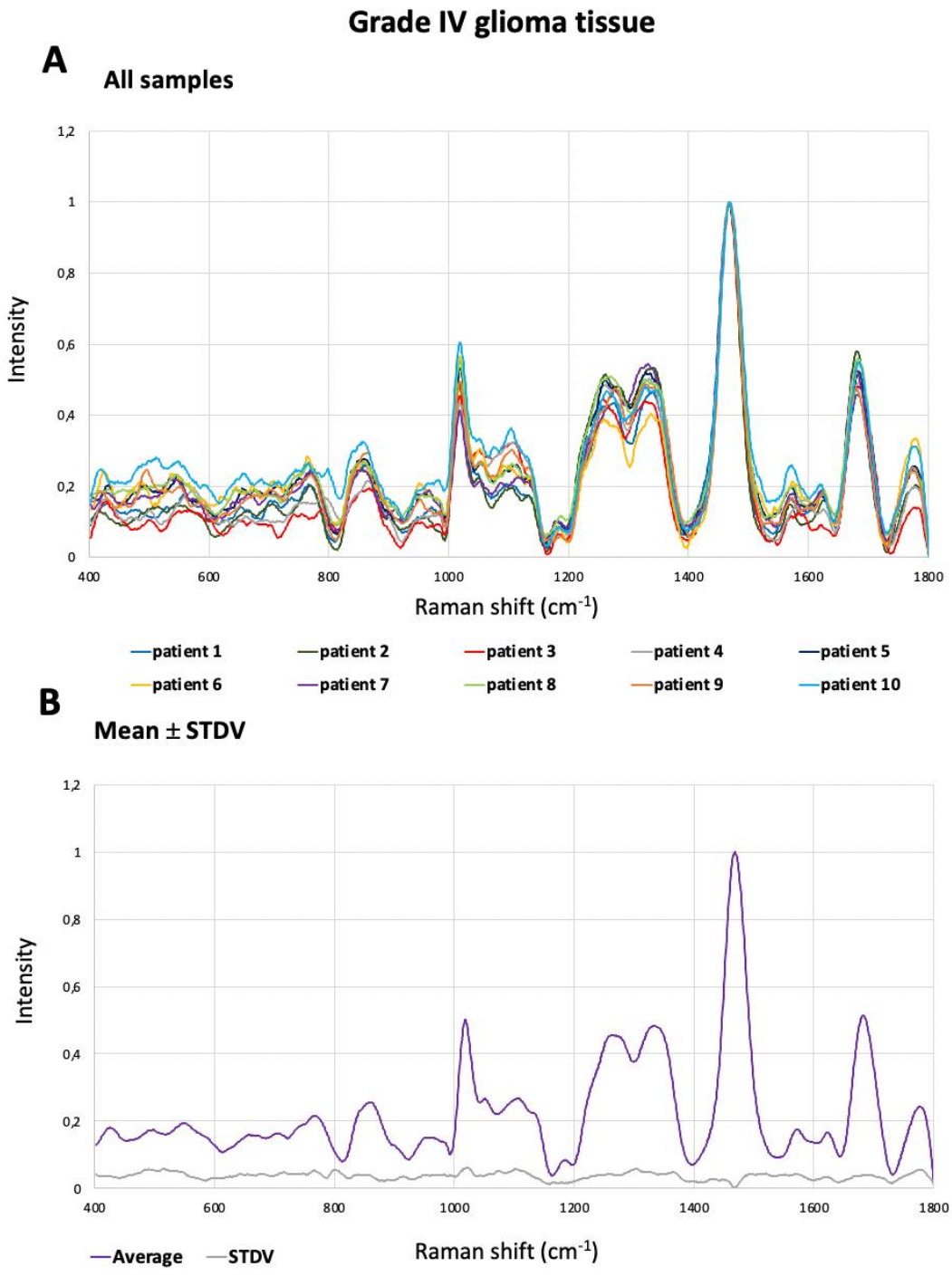


Figure 5.7 Mean tissue spectra of grade IV glioma patients. (A) All mean spectra for each patient (each sample). **(B)** Mean spectra (in purple) for grade IV with the standard deviation (in grey).

PCA and classification were then performed to evaluate whether the main variations between the samples could be assigned to their histological grades. The 2D PCA plot for the two largest PCs is represented in **Figure 5.8**. The two largest components PC1 and PC2 were enough to clearly discriminate all the patients as a function of their histological grades and explained respectively 41.7% and 27.2% (68.9% in total) of all the variance between the samples. The percentages of the variance that each PC explains between the samples are summarised in **Table 5.2**.

Classification learning was then used to calculate the accuracy rate of this discrimination by using a 5-fold cross-validation on different model types and with the three largest PCs. The linear SVM model gave the best accuracy rate to discriminate the tissue: 80% between grade II and III, 85% between III and IV, and 75% between II and IV. This suggests strongly that RS can predict the histological grade of clinical samples with high accuracy from fixed glioma tissues and importantly can discriminate grade III and grade IV glioma grades. All the classification results are summarised in **Table 5.3** and the confusion matrices for the linear SVM classification are represented in **Figure 5.9**.

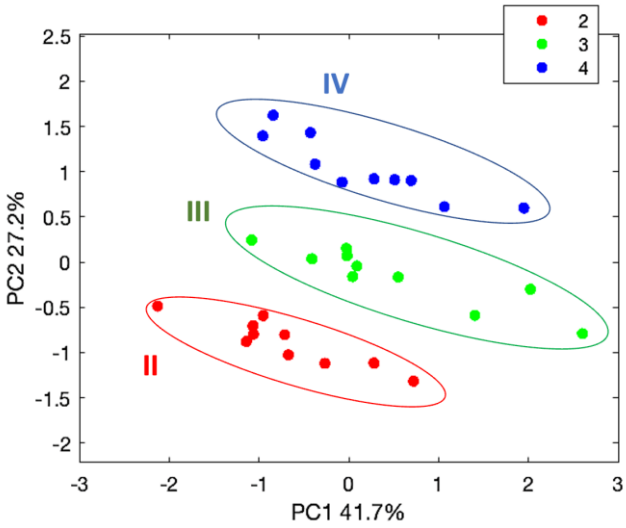


Figure 5.8 PCA analysis of grade II, III, and IV glioma tissue samples. All the patients (averaged and normalised data) were plotted for the two largest PCs (PC1 and PC2) and seemed to cluster as a function of their grade.

	% of variance explained			
Principal component	Tissue			
	II vs III	II vs IV	III vs IV	All grades
PC1	56.44	43.18	52.57	41.68
PC2	11.74	26.88	14.14	27.2
PC3	7.78	9.98	11.84	8.68
PC4	5.64	6.26	5.32	5.45
PC5	5.24	4.3	4.16	4.23
PC6	3.54	3.14	3	2.85
PC7	1.97	1.28	1.83	2.08
PC8	1.74	1.05	1.48	1.49
PC9	1.2	0.9	1.09	1.17
PC10	0.95	0.63	1.07	0.82
PC11	0.83	0.51	0.88	0.6
PC12	0.61	0.36	0.53	0.52
PC13	0.52	0.31	0.48	0.44
PC14	0.38	0.28	0.41	0.37
PC15	0.37	0.23	0.3	0.32
PC16	0.3	0.21	0.26	0.28
PC17	0.25	0.17	0.21	0.25
PC18	0.22	0.14	0.18	0.2

Table 5.2 Percentage of variance explained between glioma tissue samples. The percentage of variance between the samples for each PC and each comparison between the three grades (grade II-IV) are given.

Classification accuracy (%)	Linear SVM	Linear discriminant	Cosine KNN	Logistic regression	Bilayered neural network	Narrow neural network
Tissue						
II vs III	80%	75%	70%	70%	40%	50%
II vs IV	75%	70%	60%	70%	45%	55%
III vs IV	85%	60%	75%	60%	55%	50%
All grades	63.3%	60%	53.3%	x	36.7%	30%

Table 5.3 Accuracy rates of glioma tissue classification. Classification accuracies given by different models are represented. Accuracy rates higher than 70% are highlighted in green and rates equal or under 50% are highlighted in orange. [Support vector machine (SVM)].

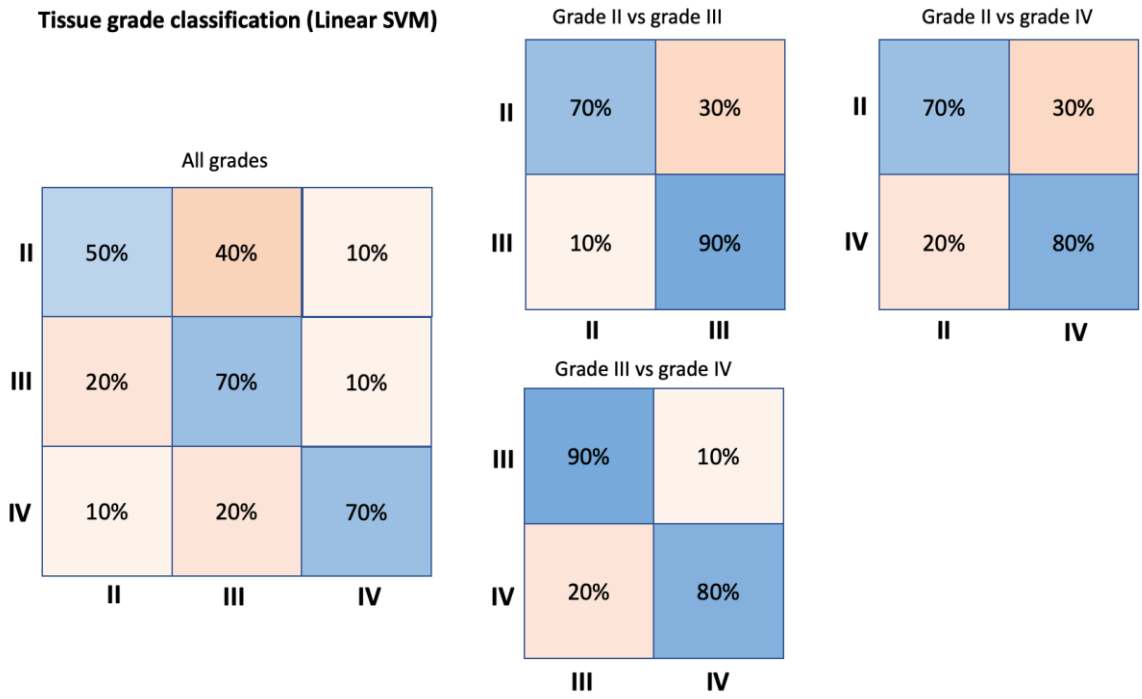
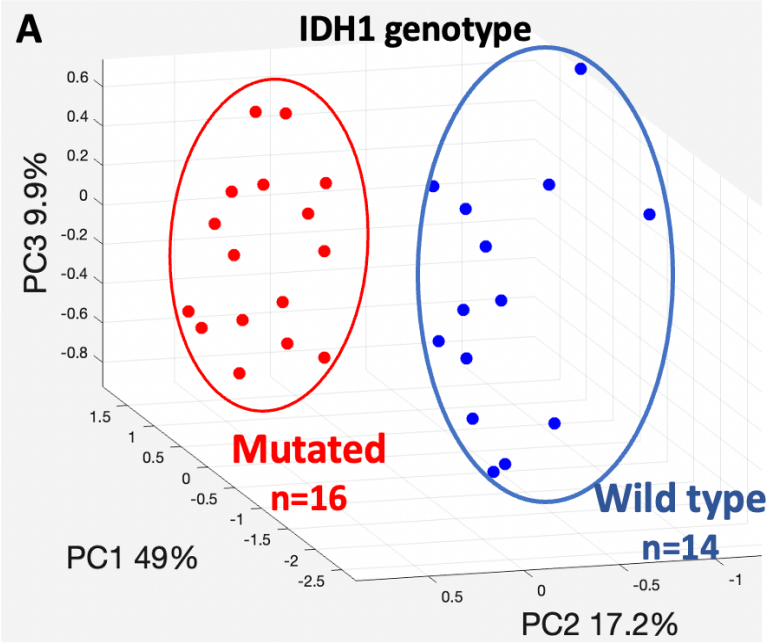


Figure 5.9 Confusion matrices of glioma tissue comparison with linear SVM classification. Rates of misclassified samples are represented in beige and rates of correctly classified samples are in blue. The intensity of the colour is proportional to the rate.

5.3.3 Discrimination between IDH1 mutated and IDH1 wild type patients from glioma tissue

Since IDH1 genotype is crucial in the current classification and diagnostic procedure (updated in 2021), the discrimination of the two genotypes (wild type and mutated) was tested to see if the methodology assessed in this chapter could be applied in clinics for objective and up-to-date diagnostics. As expected, – since IDH1 genotype was already very efficient at predicting the histology – the two genotypes discriminated with high accuracy, like the histological grades. The three first PCs explained 76% of the variance between the samples and the accuracy rate was 80% (Linear SVM) (**Fig.5.10**). The wild-type group, which is strongly associated with higher histological grades (Melhem *et al.* 2022), was slightly more dispersed on the PCA plot. The mutated group, strongly associated with lower glioma grades presenting better prognosis, displayed smaller variance between the samples.



B

Principal component	Tissue IDH-m vs IDH-WT
PC1	48.9752
PC2	17.2098
PC3	9.8729
PC4	5.1212
PC5	3.6553
PC6	3.0782
PC7	2.6785
PC8	1.7372
PC9	1.3536
PC10	1.0255
PC11	0.7635
PC12	0.6537
PC13	0.5400
PC14	0.4445
PC15	0.3938
PC16	0.3540
PC17	0.3072
PC18	0.2726

C

Classification accuracy (%)	Linear SVM	Linear discriminant	Cosine KNN	Logistic regression	Bilayered neural network	Narrow neural network
IDH m vs WT	80%	76.7%	76.7%	80%	63.3%	63.3%

Figure 5.10 Discrimination between m-IDH1 and wt-IDH1 patients in tissue. (A) 3D PCA plot of m-IDH1 and wt-IDH1 patients (grade II, III and IV). The circles drawn on the plot represent the trend determined subjectively by eye. **(B)** Percentages of variance explained by each PC. **(C).** Classification accuracy for different models of classification learning. Accuracy rate above 75% are highlighted in green.

5.3.4 Analysis of the biomolecular changes in tissue

Next, a pair-wise comparison of the representative spectra of each histological grade was performed. As the histological grades (and IDH1 genotypes) discriminated well after PCA, it could be assumed that the main spectral differences observed between the grades could have biological significance. Comparison of grades II and III and grades III and IV are represented in **Figure 5.11** since these pairs represent two consecutive stages of glioma that are challenging to discriminate in clinics because of overlapping histological characteristics (Wang *et al.* 2019). As it has been shown in previous studies (Krafft *et al.* 2005, Aubertin *et al.* 2018, Zhang *et al.* 2019), it is possible to identify changes in the tissue

composition between two consecutive stages. Changes in intensities between grades II and III were mainly present in the 400-790, 790-900, 950-1000, 1000-1150, 1250-1350, 1500-1620, and 1750-1800 cm^{-1} spectral regions (**Fig.5.11**, grade II vs III). To assign the regions and peaks that vary between the two grades to specific components, the in-house glycosylation database and the literature-based database described before were used. A detailed summary of all peak assignments in tissue are presented in **Table 5.4**.

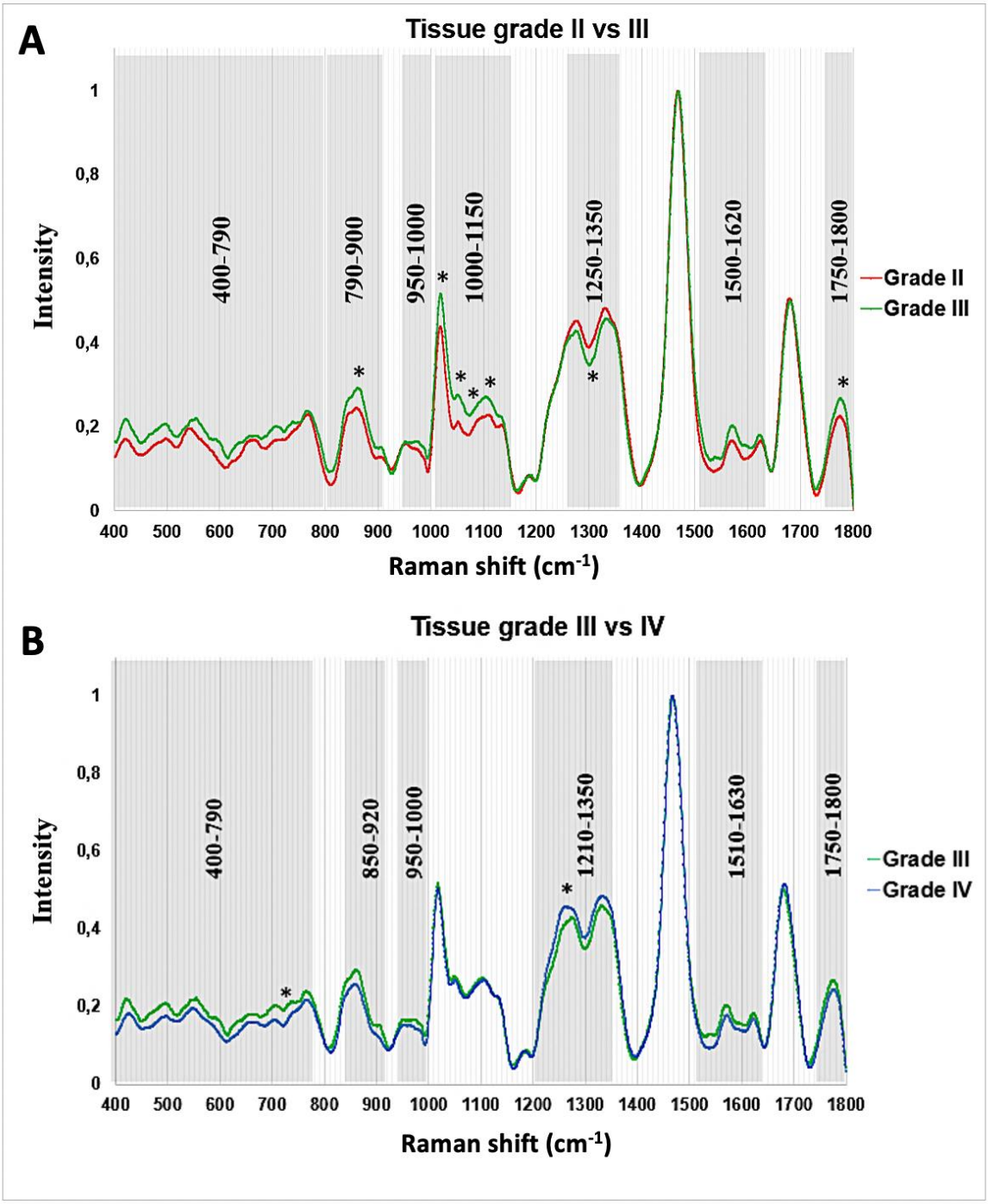


Figure 5.11 Spectral pair-wise comparison of tissue glioma grades. Pair-wise comparison between the averaged spectra of grade II and grade III **(A)**, and grade III and grade IV **(B)**. Asterisks indicate individual peaks for which the intensity values were significantly different using unpaired t-test.

Tissue observations (cm^{-1})		Tentative assignments
II<III	400-790	All glycans except NA
	790-900	Galactosamine, galactose, fucose, NA, glucose
	950-1000	Glucosamine, galactosamine, NA, fucose
	1000-1150	All glycans
	1750-1800	Galactose
	859	Glucose (843)
	912	Glucose (910-911)
	1094	Mannose (1095)
	400-780	Cholesterol and proteins (446-476, 700-703), proteins (500-550, 636-646), DNA (498, 676), proteins and DNA (725-729), haemoglobin (670), DNA, proteins, and haemoglobin (743-790)
	790-900	Collagen (818), proteins (823, 880-890), proteins, collagen, and GAG (857)
	950-1000	Proteins (959, 1003)
	1000-1150	Proteins (1032, 1127), lipids (1064-1068, 1129), phospholipids and collagen (1074),
	1500-1620	DNA, proteins, haemoglobin (1573-1585), proteins (1602-1607)
III<II	1250-1360	Amide III (1230-1306), cytochrome C (1358), lipids (1263), phospholipids (1313), collagen (1323)
IV<III	400-790	All glycans except NA
	850-920	Fucose, glucose, NA, galactose
	950-1000	Glucosamine, galactosamine, NA, fucose
	1750-1800	Galactose
	1510-1550	Carotenoids (1521)
III<IV	1210-1350	Haemoglobin (1225), GAG (1242), proteins, collagen, lipids (1230-1306), lipids (1263), phospholipids (1313), collagen, proteins (1322), DNA, proteins (1331-1338)

Table 5.4 Summary of biomolecular changes observed in tissue and their tentative assignments. Assignments were made from the in-house glycosylation database (in blue) and from the literature-based general database (in grey). [Neuraminic acid (NA)]. (Arboleda *et al.* 2000, Feng *et al.* 2010, Kast *et al.* 2014, Lin *et al.* 2014, Wiercigroch *et al.* 2017, Abramczyk *et al.* 2018, Aubertin *et al.* 2018, Galli *et al.* 2019, Zhang *et al.* 2020, Riva *et al.* 2021).

The frequency ranges that dominated in intensity in grade III compared with grade II could be assigned from the in-house glycosylation database to different glycans across the spectrum: all the glycans except neuraminic acid (400-790 cm^{-1}); galactosamine, galactose, fucose, neuraminic acid, and glucose (790-900 cm^{-1}); glucosamine, galactosamine, neuraminic acid, fucose (950-1000 cm^{-1}); all glycans (1000-1150) cm^{-1}; and galactose (1750-1800 cm^{-1}). The assignment of some glycans was confirmed by the literature database, for instance glucose (859, 912 cm^{-1}) and mannose (1094 cm^{-1}) (**Table 5.4**). The most important difference seemed to be localised in the 950-1200 cm^{-1} region, which corresponds to the high intensity regions shared by all glycans. In addition to the glycosylation changes, assignments to other biocomponents present in several spectral areas could be made from the literature database (**Table 5.4**): cholesterol, proteins, haemoglobin, DNA, GAG, collagen, proteins, lipids, and phospholipids. There was a slight decrease observed in grade III compared with grade II in the 1250-1360 cm^{-1} region that contained peaks that could be assigned to the amide III and cytochrome C bands and that also contained peaks assigned to lipids, phospholipids, and collagen from the literature database (**Table 5.4**).

Changes in intensities between grades III and IV were smaller and mainly present in the 400-790, 850-921, 950-1000, 1210-1350, 1510-1630, and 1750-1800 cm^{-1} spectral regions (**Fig.5.11**, grade III vs IV). The 950-1200 cm^{-1} showed no difference, meaning that the glycan increase observed in grade III compared with grade II in this region persisted in grade IV. Outside this region, there was a slight decrease in grade IV compared with grade III and most spectral ranges that differed could be assigned to glycans (400-790, 850-920, 950-1000, 1750-1800 cm^{-1}) from the in-house glycosylation database; but also, other biomolecules such as carotenoids (1510-1630 cm^{-1}) from the literature database. Increase of intensity in the 1210-1350 cm^{-1} in grade IV compared to grade III could also be interpreted as an increase in collagen, proteins, lipids, haemoglobin, GAG, and DNA. Comparison between grade II and grade IV spectra (data not shown) showed only an increase in grade IV compared to grade II in the 950-1200 cm^{-1}, which corresponded to the high intensity spectral region shared by all glycans. From the literature, this region could also be assigned to an increase in proteins, lipids, phospholipids, and collagen, whereas all the other regions had very similar intensities (**Table 5.4**).

To further assess the significance of these changes observed between the different grades and to see if some specific individual peaks could be more useful for diagnosis than others, peaks that showed the largest intensity difference between two grades were selected. T-tests were then performed for these peaks, meaning, intensity values for each specific peak were compared between both grades (**Fig.5.12**, significance is indicated by asterisk in **Fig.5.11**). Differences between grades II and III were overall more significant than differences between grade III and IV and notably, the 950-1200 cm^{-1} region, assigned to glycans, showed significant difference for the four peaks selected (1015, 1042, 1050, 1070 cm^{-1}). These peaks could not be assigned to any other components than glycans from the literature database. The other peaks that showed significant difference could be assigned to other well-characterised biocomponents from the literature database (**Fig.5.12 and Table 5.4**). Overall, changes in glycosylation were obvious and significant between grade II and the higher grades, while these changes were smaller between grade III and IV.

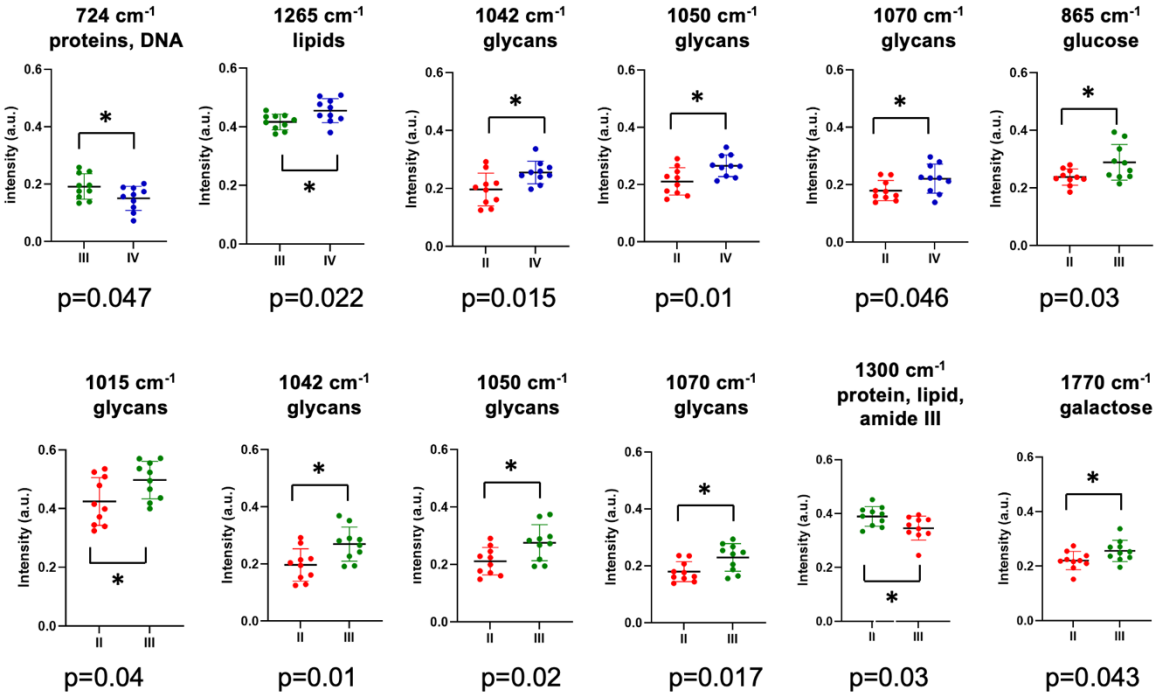


Figure 5.12 Pair-wise comparison of individual intensities between tissue glioma grades and assignments. Scatter plots of intensity values (and mean ± standard deviation) at peaks showing significant difference using two-tailed unpaired t-test, with their corresponding assignments. a.u. = arbitrary unit (normalised spectral intensity). Grade II, III, and IV samples are represented in red, green, and blue, respectively.

The PCA loadings plot was next generated to visualise the contribution of each PC to the variances observed between the grades across the spectrum. In addition to the t-tests performed, this type of representation can further confirm what areas of the spectrum contribute the most to the significant differences between the grades. Since the three largest PCs allow the clusterisation of the grades, therefore these PCs - representing the largest variance observed - are the ones that can be correlated to the histological differences between the samples. These PCs can be related to the different variables to identify the ones that account the most to the histological differences. The results are represented in **Figure 5.13**. The different PCs are usually widely distributed throughout the variables, which makes the interpretation of these graphs complex. However, the plot showed clearly that two

regions had major contribution in the grade distinction, namely, the 950-1200 cm^{-1} and the 1250-1350 cm^{-1} regions of the spectrum, which further confirmed the importance of glycosylation changes discussed earlier but also the relevance of this method.

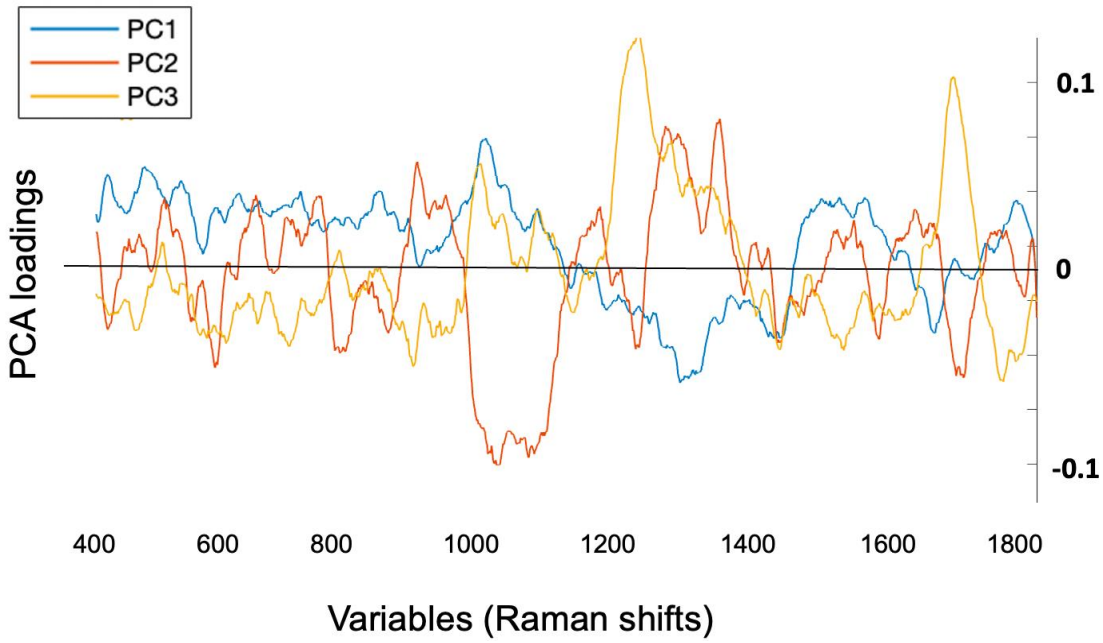


Figure 5.13 PCA loadings plot of glioma tissue discrimination. The PCA loadings of the three largest PCs are plotted against the variables (Raman shifts). This representation permits the identification of each PC's contribution to the variance observed between the grades at each point of the spectrum. The importance of the 950-1200 cm^{-1} and 1250-1350 cm^{-1} regions are further confirmed.

Finally, the Raman signatures of m-IDH1 and wt-IDH1 patients were compared. The comparison between the two signatures looked very similar to the comparison between the grade II (containing almost only m-IDH1 patients) and grade IV (containing almost only wt-IDH1 patients) as expected. Therefore, the wild-type signature showed only an increase in the 1010-1140 cm^{-1} (within the glycan-rich region) in comparison with the mutated signature (**Fig.5.14**). This implies that the level of overall glycosylation could account widely for the high accuracy of Raman discrimination between m-IDH1 and wt-IDH1 patients.

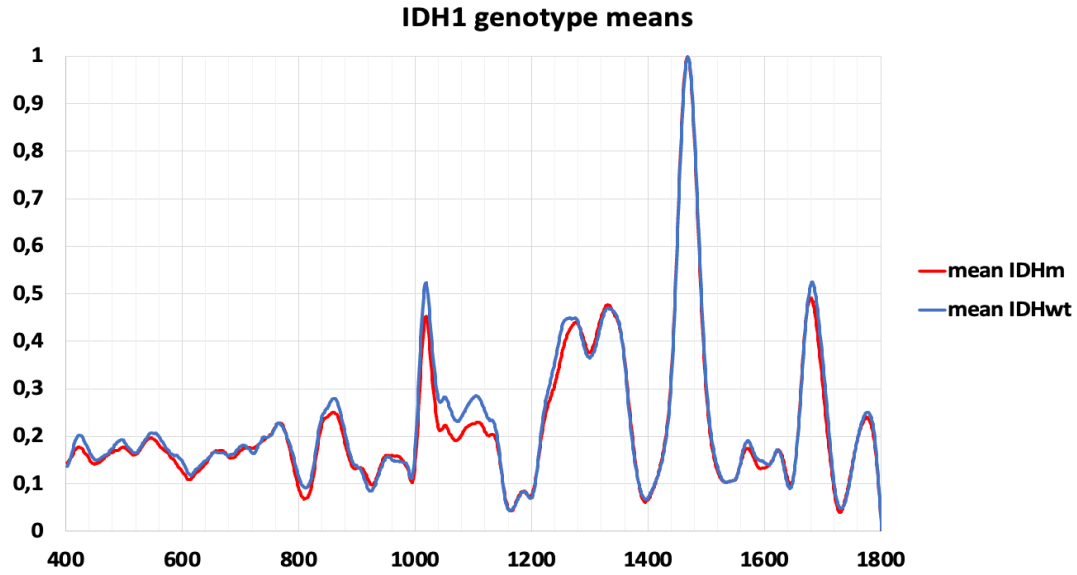


Figure 5.14 m-IDH1 and wt-IDH1 mean Raman spectra.

5.3.5 Grade classification from liquid serum samples

Next, the potential use of blood serum for grade III and grade IV distinction was further evaluated as it can be collected less invasively and more frequently than solid biopsies. Three groups of equal size (grade III, grade IV, and benign non-brain cancer controls, 10 per group) were included in the study. Summary of the patients' characteristics is presented in **Table 5.1**. The same technique, parameters, and processing were used as described earlier for the tissue approach. A drop of liquid serum was dispensed on the stainless-steel slide for RS acquisitions. The averaged spectrum of the different replicates was generated for each patient after processing and 0 to 1 rescaling. The number of replicates was reduced to 5 since liquid serum is expected to not contain any intrinsic heterogeneity, unlike the tissue samples. The Raman spectra signatures for all the patients are represented in **Figure 5.15** (controls), **5.16** (grade III), and **5.17** (grade IV), categorised by grade with the mean of the patients for each grade and the standard deviation between the patients. The standard deviation between the serum samples was visibly higher than for the tissue, meaning, the inter-patient heterogeneity seemed to be high for all the groups (control group and glioma grades). This may indicate that RS can detect more subtle biological changes from liquid serum than from dewaxed fixed tissue, which could be explained by the nature of this type of samples (non-paraffinized and in their native state).

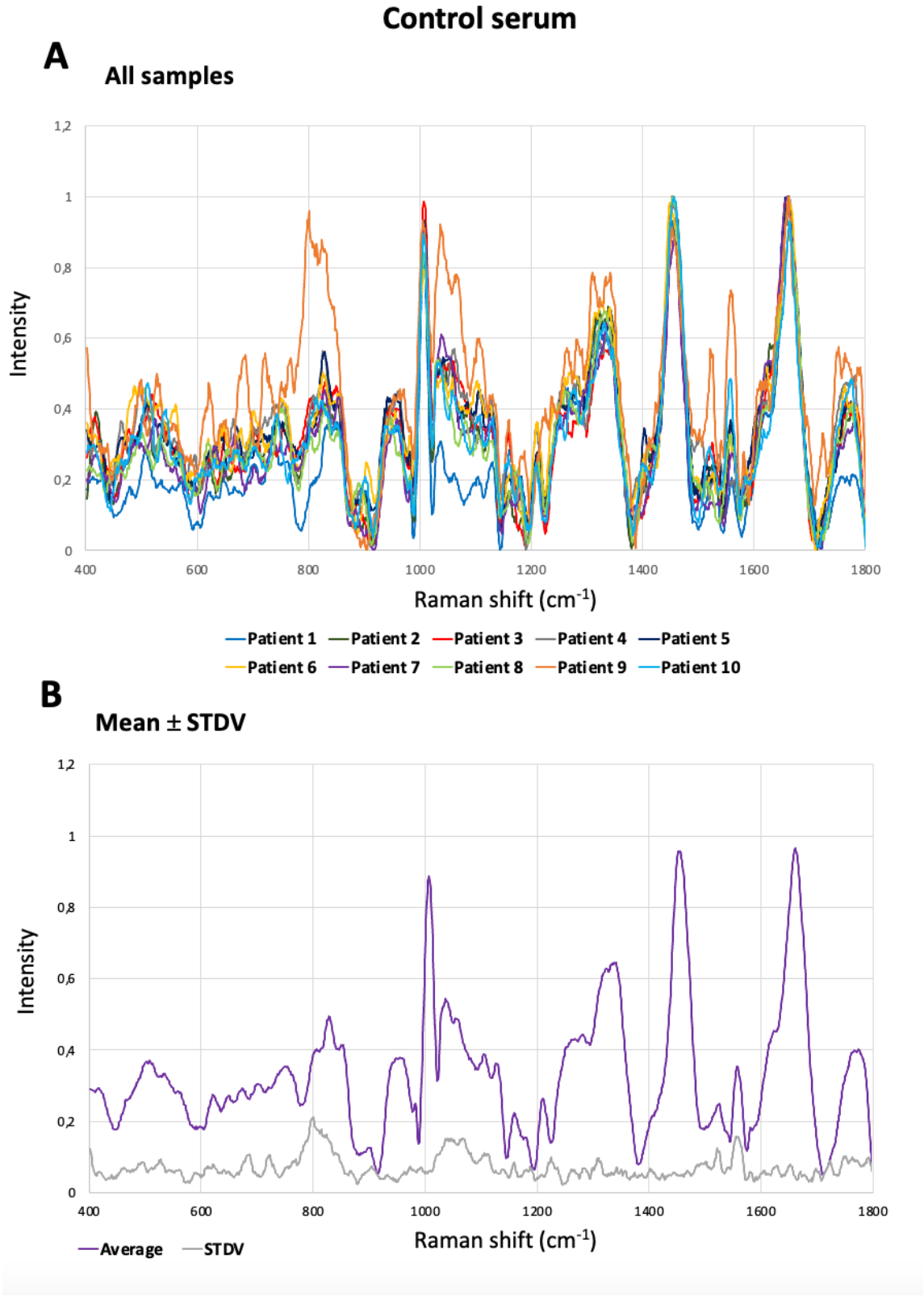


Figure 5.15 Mean serum spectra of non-glioma control patients. (A) All mean spectra for each patient (each sample). **(B)** Mean spectra (in purple) for controls with the standard deviation (in grey).

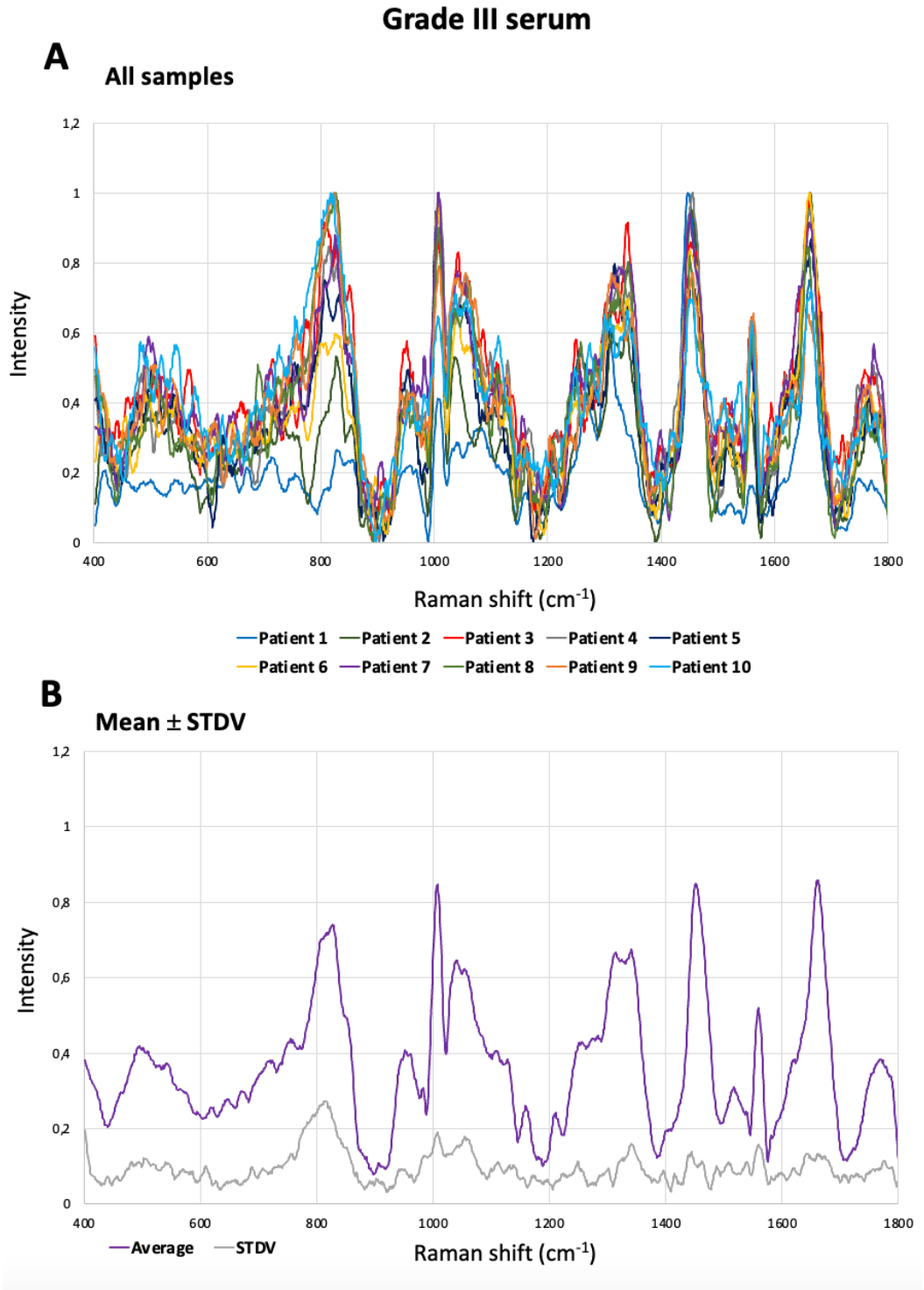


Figure 5.16 Mean serum spectra of grade III glioma patients. (A) All mean spectra for each patient (each sample). **(B)** Mean spectra (in purple) for grade III with the standard deviation (in grey).

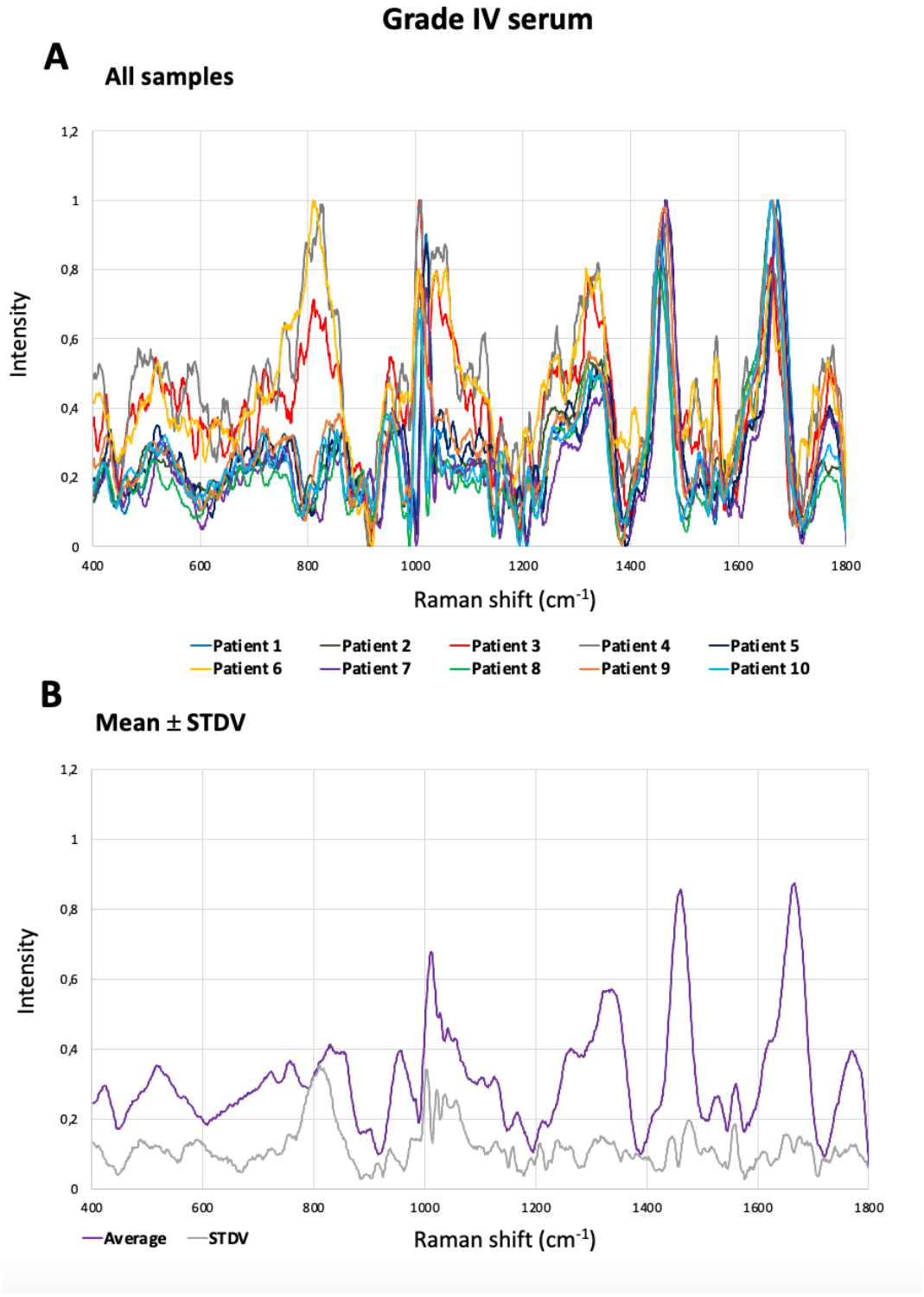


Figure 5.17 Mean serum spectra of grade IV patients. (A) All mean spectra for each patient (each sample). **(B)** Mean spectra (in purple) for grade IV with the standard deviation (in grey).

The PCA plots of the liquid serum samples using the two or three largest PCs are represented in **Figure 5.18**. The three PCs represented respectively 57.3%, 17.4%, and 5.1% of all the variance between all the samples (79.8% in total). Unlike the tissue samples, which showed very similar distribution sizes for the different grades, the serum samples showed higher distribution differences between the grades. Samples from control group were very close to each other and showed no big difference or heterogeneity, whereas patients from grade III showed more dispersion and so potentially, more heterogeneity in their composition (**Fig.5.18**).

The discrimination between the control and the grade III was very clearly made with a high distance between both groups and no grade overlapping. The small size of the control group distribution clearly suggests that distance between the glioma samples could have biologically significance since it is expected that malignant cancer samples would show more inter-patient heterogeneity compared with control samples. In the 3D PCA plot of the grade III and grade IV groups, the three first components represented respectively 67.7%, 11.1%, and 6% of all the variance between the samples (84.8% in total).

Table 5.5 lists all the variances explained for liquid serum. When comparing grade III and grade IV, grade IV samples also displayed a moderate increase in their distribution size in comparison with grade III samples that can indicate an increase in heterogeneity between grade III and grade IV. When looking retrospectively at the spectral comparison, the observed inter-patient heterogeneity at the 800 cm^{-1} peak appeared particularly high, especially for grade IV. Thus, corresponding patients' characteristics for samples displaying a high intensity at this peak were checked and no known specific clinical parameter was found that could account for this high difference.

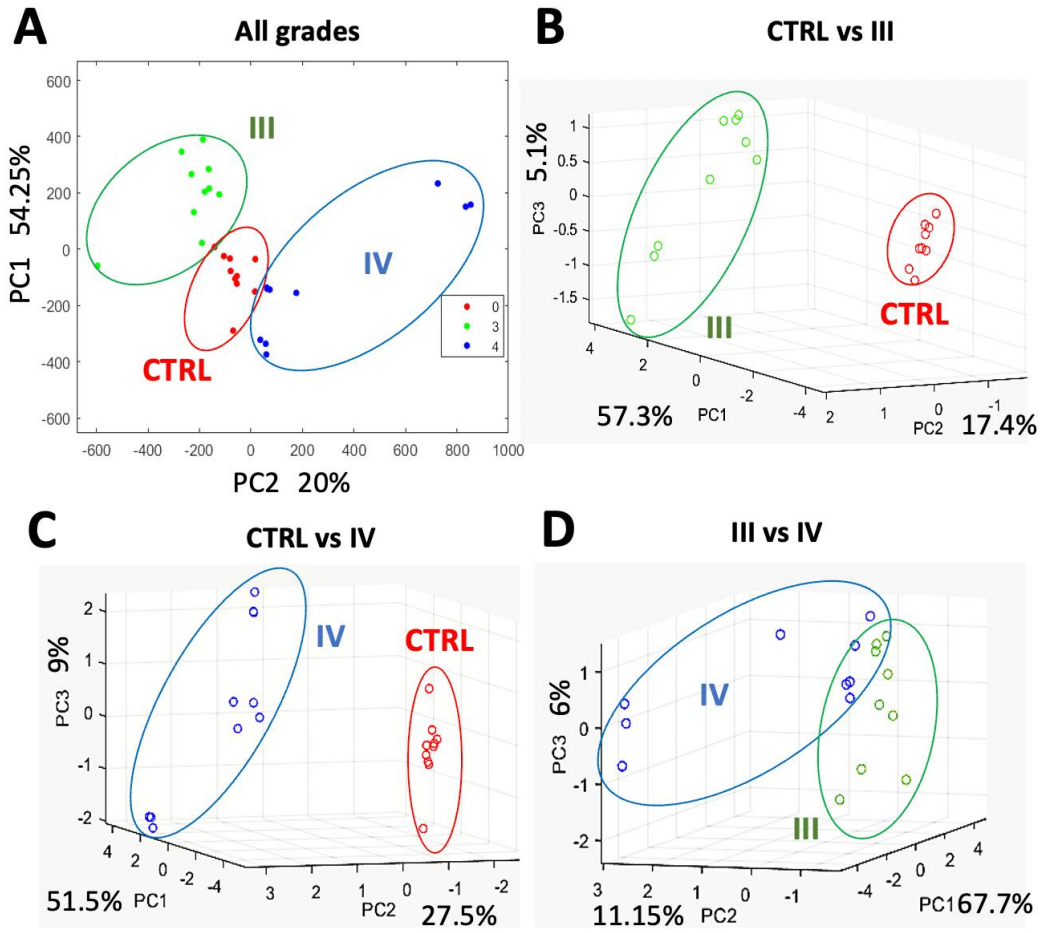


Figure 5.18 PCA plots of control, grade III, and grade IV liquid serum samples. **(A)** 2D PCA plot of all grades compared. **(B)** 3D PCA plots of controls vs grade III, **(C)** controls vs grade IV, and **(D)** grade III vs grade IV **(D)**. Percentage of the variance explained for each PC is indicated on the axes.

	% of variance explained			
Principal component	Fresh serum			
	CTRL vs III	CTRL vs IV	III vs IV	All grades
PC1	57.31	51.48	67.68	54.25
PC2	17.44	27.5	11.15	19.99
PC3	5.11	8.58	6.02	6.84
PC4	3.17	2.45	2.43	4.36
PC5	2.58	1.66	2.25	1.85
PC6	1.87	1.43	1.81	1.53
PC7	1.76	1.15	1.39	1.39
PC8	1.52	1.07	1.15	1.02
PC9	1.43	0.82	1.06	0.92
PC10	1.42	0.69	0.98	0.85
PC11	1.08	0.58	0.85	0.82
PC12	1.02	0.53	0.72	0.73
PC13	0.88	0.43	0.63	0.72
PC14	0.76	0.39	0.51	0.6
PC15	0.69	0.34	0.45	0.52
PC16	0.62	0.29	0.3	0.46
PC17	0.49	0.2	0.22	0.45
PC18	0.42	0.16	0.16	0.38

Table 5.5 Percentages of variance explained between liquid serum samples. The percentages of variance between the samples for each PC and each comparison are given.

Classification learning were then performed to generate the accuracy rates with the three largest PCs as described previously. Results are presented in **Table 5.6**. The accuracy rates were improved with the liquid serum in comparison with the tissue samples. The Bilayered neuronal network classification method gave the best discrimination: 90% between control and grade III, 85% between control and grade

IV, and importantly, 90% between grade III and grade IV glioma patients (with a 5-fold cross validation). The confusion matrices for the bilayered neuronal network are represented in **Figure 5.19**.

Classification accuracy (%)	Linear SVM	Linear discriminant	Cosine KNN	Logistic regression	Bilayered neural network	Narrow neural network
Fresh serum						
CTRL vs III	75%	75%	80%	85%	90%	80%
CTRL vs IV	60%	65%	70%	85%	85%	85%
III vs IV	60%	65%	65%	75%	90%	85%
All grades	56.7%	53.3%	60%	x	63.3%	66.7%

Table 5.6 Accuracy rates of liquid serum classification. Accuracy rates higher than 70% are highlighted in green and rates under 50% are highlighted in orange.

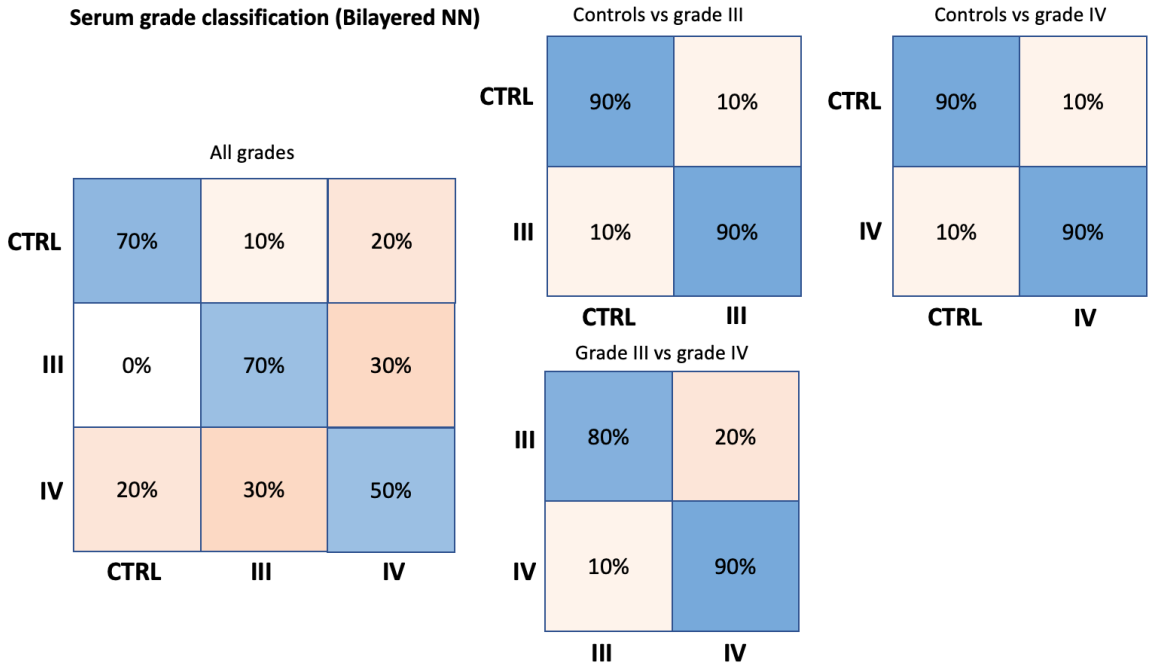
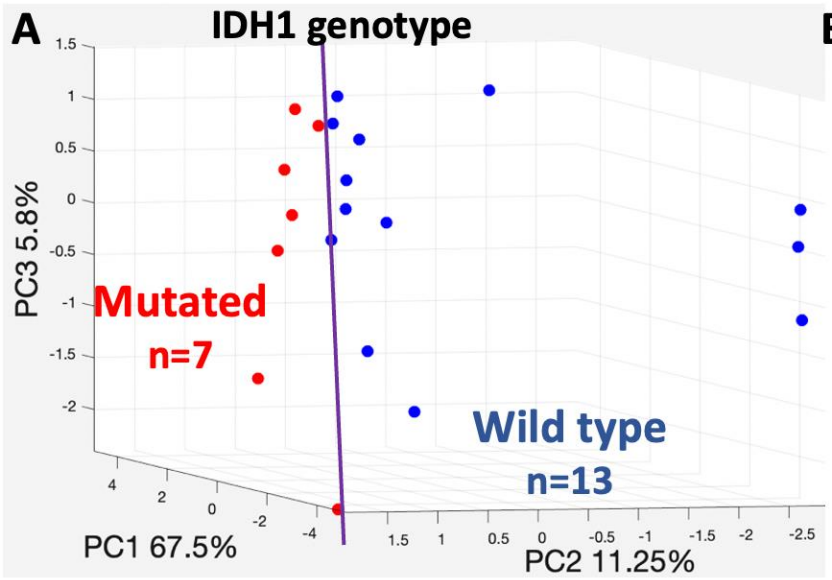


Figure 5.19 Confusion matrices of liquid serum glioma grades with Bilayered NN classification. Rates of misclassified samples are represented in beige and rates of correctly classified samples are in blue. The intensity of the colour is proportional to the rate. [Neural network (NN)].

Finally, the discrimination between the two IDH1 genotypes was tested as previously with the tissue on the 20 glioma patients contained in the serum sample set. The three first largest PCs explained 84.5% of the variance between the samples and the accuracy rate was overall lower than for the IDH1 discrimination in tissue (65% with Linear SVM). The clusterisation was visibly more challenging to make on the PCA plot. However, a trend of clusterisation could be observed (**Fig.5.20**) and the logistic regression gave an accuracy of 85%.



B

Principal component	Serum IDH-m vs IDH-WT
PC1	67.5438
PC2	11.2513
PC3	5.8362
PC4	2.5125
PC5	2.2862
PC6	1.7578
PC7	1.4391
PC8	1.1884
PC9	1.1016
PC10	0.9471
PC11	0.8704
PC12	0.7394
PC13	0.7129
PC14	0.5172
PC15	0.4603
PC16	0.3033
PC17	0.2285
PC18	0.1601

C

Classification accuracy (%)	Linear SVM	Linear discriminant	cosine KNN	Logistic regression	Bilayered neural network	Narrow neural network
IDH m vs WT serum	65%	60%	70%	85%	60%	55%

Figure 5.20 Discrimination between m-IDH1 and wt-IDH1 patients in serum. (A) 3D PCA plot of m-IDH1 and wt-IDH1 patients (grade III and IV). The purple line indicates the discrimination trend between the two group as determined subjectively by eye. **(B)** Percentages of variance explained by each PC. **(C)** Classification accuracy for different models of classification learning. Accuracy rate above 75% are highlighted in green.

5.3.6 Grade discrimination from dried blood serum samples

The question of whether it is more suitable to use dried (Ryzhikova *et al.* 2015) or fresh (liquid) serum (Nargis *et al.* 2021) samples for Raman cancer diagnostics have not been elucidated yet and both techniques are tested in the literature. It might be tempting to think that liquid serum is a better choice to represent the composition of the samples since this type of sample remains in its native organic form. However, dried samples could lead to better or worse discrimination accuracy rates. To further investigate this notion, the same protocol was tested as previously described with the only difference that the acquisitions were performed on dried serum. The results are presented in **Figure 5.21**. Some differences could be seen between the groups but overall, the differences were less obvious than from fresh serum (**Fig.5.21A**), even though the discrimination from the PCA was good when using the three largest PCs (explaining 89% of all the variance between the samples) (**Fig.5.21B**). Variance explained are represented in **Table 5.7**. Interestingly, the increased heterogeneity observed in glioma samples was strongly reduced when using dried samples. Overall, there were less significant difference between the peaks from the different groups and no significance between control and grade III samples (**Fig.5.21C**). Classification results are represented in **Table 5.8.** The classification accuracy was largely reduced when using dried serum instead of liquid serum. It could be concluded that fresh (liquid) serum might be more suitable for Raman cancer diagnostics. Moreover, the fresh serum method should be used if studying inter-patient heterogeneity.

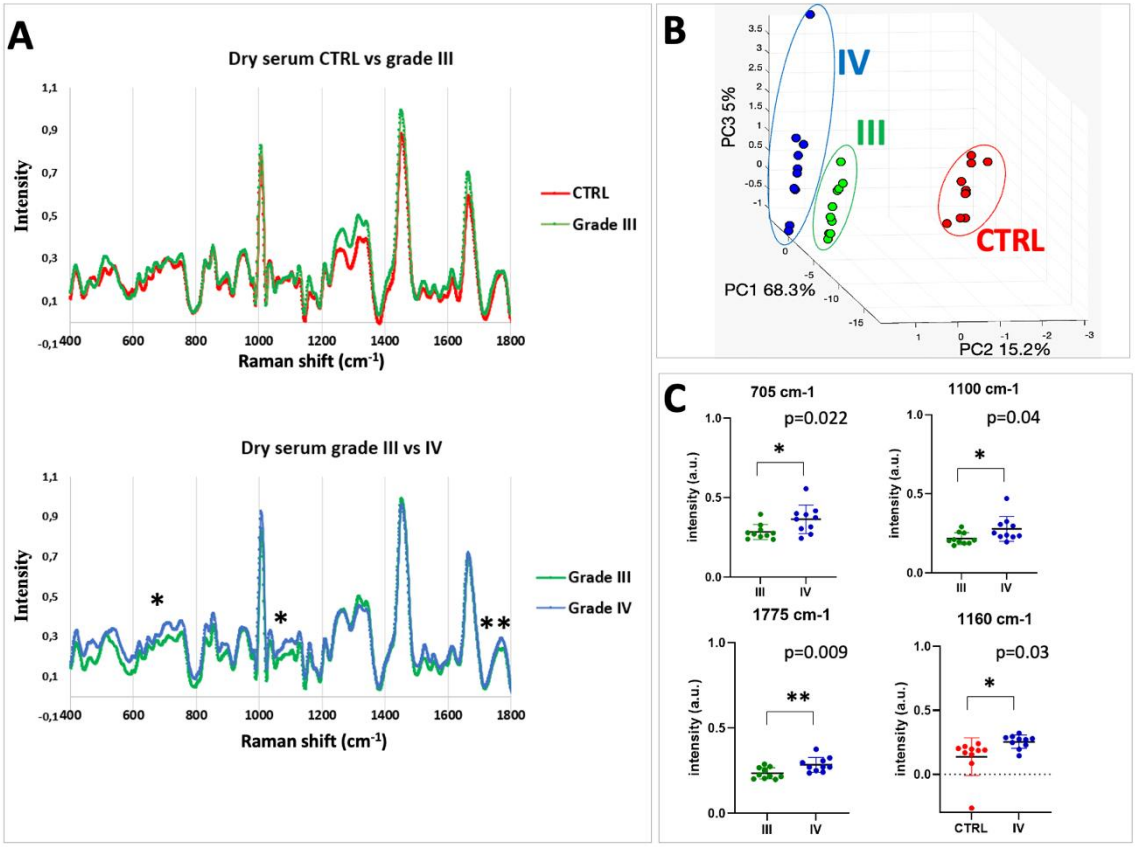


Figure 5.21 Grade discrimination from dry serum glioma patient samples. (A) Pair-wise comparison between the averaged spectra of control and grade III, and grade III and grade IV with the dry serum method. Asterisks indicate the peaks that were significantly different using t-test. **(B)** 3D PCA plot of the 30 dried glioma blood serum using the first three principal components (PC1, PC2, PC3). **(C)** Scatter plots of individual intensities (and mean ± standard deviation) at peaks showing significant difference using two-tailed unpaired t-test.

	% of variance explained			
Principal component	Dry serum			
	CTRL vs III	CTRL vs IV	III vs IV	All grades
PC1	78.95	71.62	50.13	68.27
PC2	11.85	14.83	14.22	15.2
PC3	2.34	4.62	7.08	5.53
PC4	1.94	2.23	6.56	2.48
PC5	1.05	1.84	5.01	1.81
PC6	0.74	0.98	3.81	1.12
PC7	0.49	0.57	2.47	0.79
PC8	0.47	0.53	1.91	0.66
PC9	0.39	0.5	1.51	0.55
PC10	0.31	0.42	1.32	0.41
PC11	0.25	0.37	1.25	0.38
PC12	0.22	0.32	0.94	0.33
PC13	0.21	0.24	0.91	0.31
PC14	0.18	0.21	0.84	0.26
PC15	0.16	0.16	0.78	0.24
PC16	0.12	0.15	0.52	0.2
PC17	0.1	0.14	0.34	0.19
PC18	0.08	0.12	0.3	0.17

Table 5.7 Percentage of variances explained between dried glioma serum samples. The percentages of variance between the samples for each PC and each comparison are given.

Classification accuracy (%)	Linear SVM	Linear discriminant	Cosine KNN	Logistic regression	Bilayered neural network	Narrow neural network
CTRL vs III	65%	70%	70%	65%	60%	50%
CTRL vs IV	40%	45%	35%	55%	55%	50%
III vs IV	68.4%	73.7%	58%	73.7%	58%	63.2%
All grades	40%	43.30%	40%	x	40%	40%

Table 5.8 Accuracy rates of dry serum glioma classification. Accuracy rates higher than 70% are highlighted in green and rates under 50% are highlighted in orange.

5.3.7 Analysis of the biomolecular changes in liquid serum

Pair-wise comparison of the representative normalised spectra of each group (averaged patients) was then made to identify the biomolecular changes between them as previously (**Fig.5.22**). Compared with the control group, grade III patients showed dominant intensities in wavenumber ranges that could mostly be assigned to glycans (**Table 5.9**). A reduction observed in grade III could be assigned to phospholipids and lipids. A similar trend was observed between grade III and grade IV with grade III samples showing increases compared to grade IV in the regions that have previously been assigned to glycans. These regions could also be assigned to cholesterol and lipids from the general literature database. The regions assigned to carotenoids were decreased in intensity in grade IV compared with grade III (**Table 5.9**). In some regions assigned to glycans, intensities in the control signature were slightly higher than grade IV intensities (data not shown) in the 800-850, 1000-1130, and 1210-1340 cm^{-1} regions.

Overall, peaks that could be assigned to glycosylation changes seemed to increase in intensity first in glioma (grade III) and then drop in grade IV, which would be in accordance with what happened in the tissue.

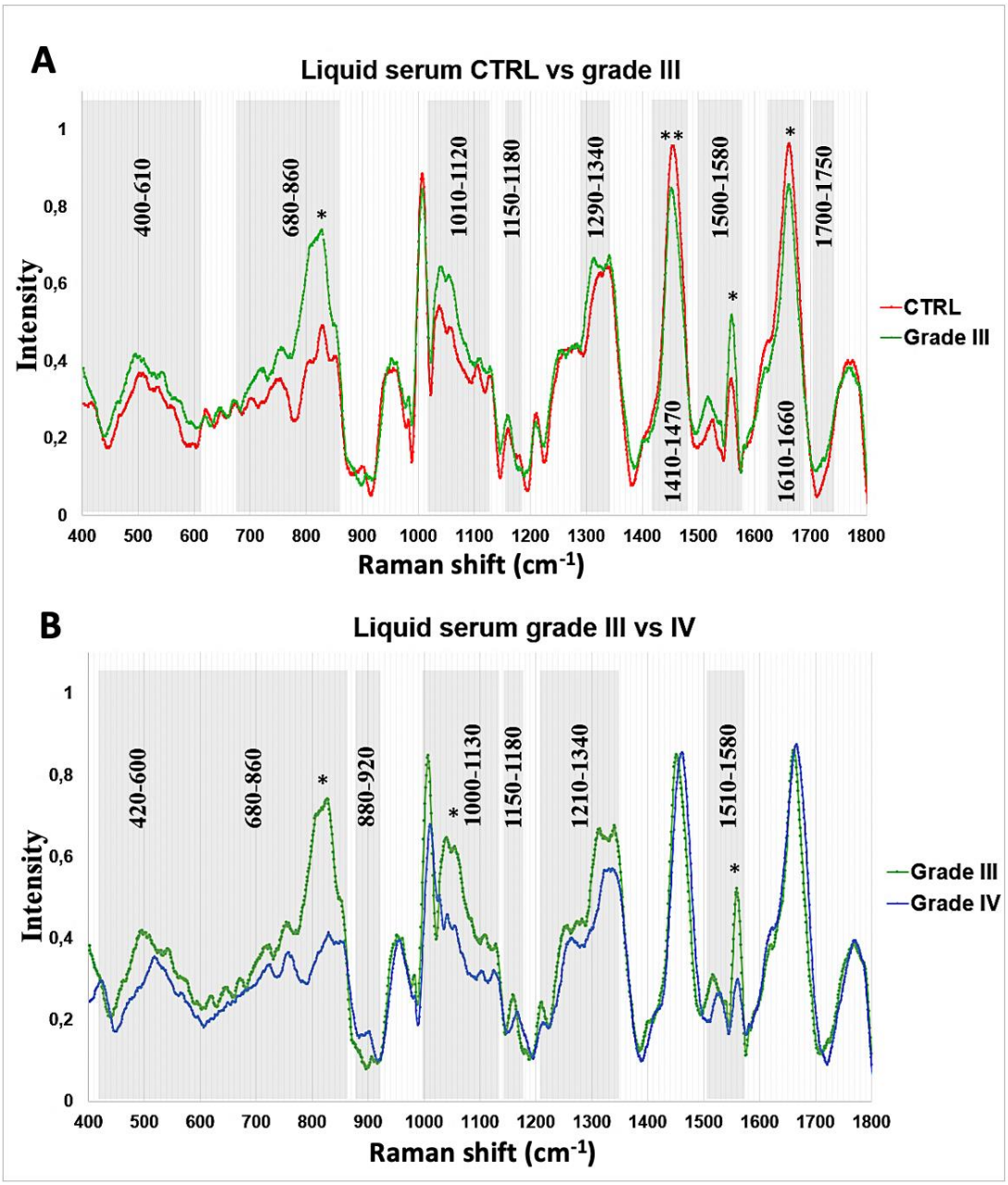


Figure 5.22 Spectral pair-wise comparison of liquid serum glioma grades. Pair-wise comparison between the averaged spectra of control and grade III **(A)**, and grade III and grade IV **(B)**. Asterisks indicate individual peaks that were significantly different using unpaired t-test.

Serum observations (cm^{-1})		Tentative assignments
III>CTRL	400-610	Galactose, mannose, fucose, glucose, galactosamine
	680-860	All glycans
	1010-1120	
	1290-1340	Glucose, NA, glucosamine, galactosamine
	1700-1750	Lipids (1732)
III<CTRL	1420-1470	Phospholipids (1441-1445)
	1610-1660	Lipids (1654)
III>IV	420-600	All glycans except NA
	680-860	Galactose, galactosamine, NA, fucose, glucose
	1000-1130	All glycans
	1210-1340	Glucose, galactose, glucosamine, galactosamine
	450-500	Cholesterols (446-476, 700-703)
	689-750	
	1010-1070	Lipids (1064-1068, 1220-1306, 1400, 1732)
	1290-1340	
	1380	
	1700-1750	
	1145-1160	Carotenoids (1157, 1521)
	1500-1520	
III<IV	880-920	Glucose (910-911), proteins (880-890)

Table 5.9 Summary of changes observed in fresh serum and their tentative assignments. Assignments were made from the in-house glycosylation database (in blue) and from the literature-based general database (in grey). [Neuraminic acid (NA)].

To further assess the significance of these changes, t-tests were again performed for the most discriminative wavenumbers. The peaks that showed significance are represented in **Figure 5.23** with their respective p-values and significance is indicated by asterisks on **Figure 5.22**. The tyrosine peak contained in the 700-850

cm^{-1} regions, broadly assigned to glycans, was significant between both controls and grade III samples, and grade III and grade IV samples. The difference between grade III and grade IV in the 1000-1130 region cm^{-1} was also significant. The 1455 cm^{-1} peak was the most significant for discrimination between controls and grade IV (p=0.0003) and controls and grade III (p=0.002). In total, three peaks could be used to significantly discriminate grade III and grade IV (830 cm^{-1}, p=0.01; 1050 cm^{-1}, p=0.05; and 1560 cm^{-1}, p=0.01).

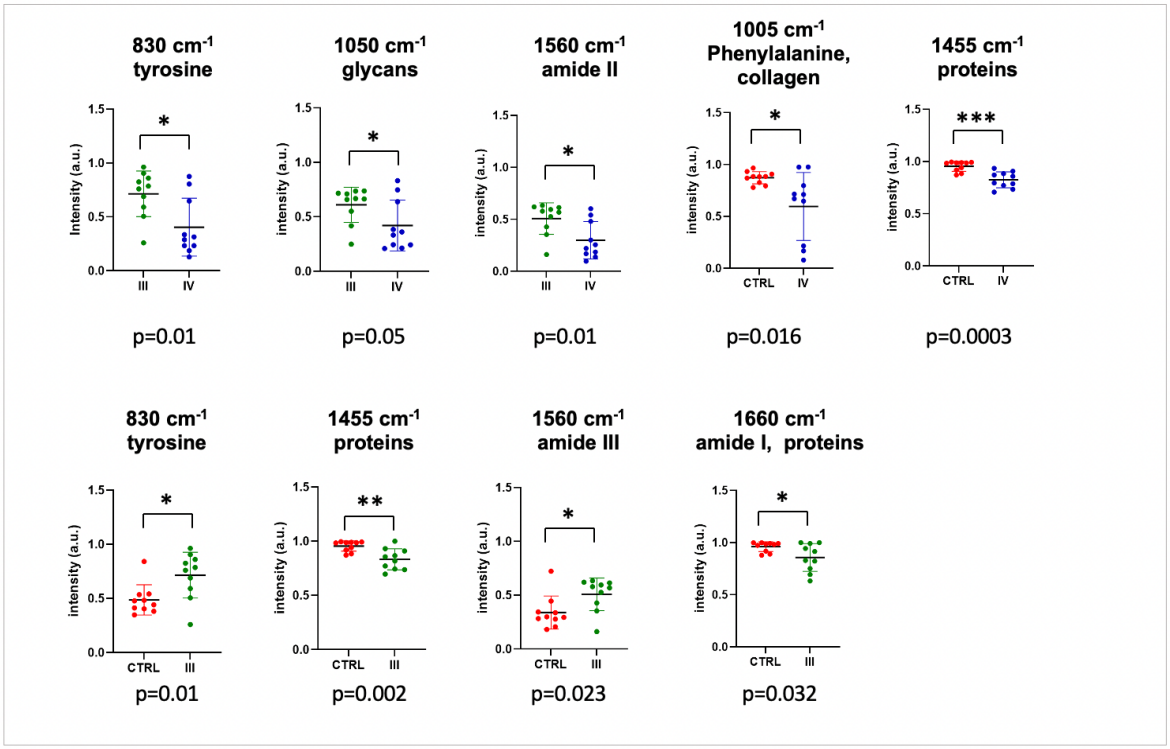


Figure 5.23 Pair-wise comparison of individual intensities between serum glioma grades. Scatter plots of intensity values (and mean ± standard deviation) at peaks showing significant difference using two-tailed unpaired t-test, with their corresponding assignments from the literature and from the glycosylation database. Control, grade III, and IV samples are represented in red, green, and blue, respectively.

5.4. Discussion

5.4.1 Classification of glioma grades from fixed tissue

In this chapter, RS combined with PCA and classification learning approaches was first used to evaluate the discrimination efficiency of histological grades II, III, and IV in glioma FFPE samples. RS, coupled with an optical microscope, allows for the accurate targeting of regions of interest through direct visualisation. The method used in this study was designed to represent an objective and convenient mean of routine brain cancer diagnostics. By using FFPE samples, it is ensured that the spectral interrogation can be performed at any time after the surgery. In addition, the use of labels or intensity enhancement techniques, that are sometimes used in combination with RS (Canetta 2021, Mamede *et al.* 2021), were not used in this study. This decision was made to test whether RS could be used as a single diagnostic tool with no demanding sample preparation or supplementary materials required. Finally, from a technical viewpoint, stainless-steel slides were preferred over the more traditional CaF_2 slides, not only for their low signal-to-noise ratio, but also for other advantageous characteristics such as their sturdiness and low price (Lewis *et al.* 2017).

This study showed that RS combined with classification learning could classify the patients following the histological (2016) WHO classification, with high accuracy (Louis *et al.* 2016). In addition, the three largest PCs were enough to observe this discrimination and the two largest PCs alone explained almost 70% of all the variances observed between the patients. This implies that the differences noted between the patients could be largely explained by the histological classification and the associated biological changes, justifying its use for diagnosis and biological composition analysis. Some spectral differences showed significant difference when using t-tests, which implies that this method could be used as an objective diagnostic tool in clinics and that only a set of selected peaks could be assessed to reduce the acquisition time.

5.4.2 Glycosylation changes between grade II and III tissue glioma

Between the grade II and the grade III tissue samples, which represent respectively low-grade and high-grade glioma, increase of intensity in the signature could largely

be assigned to glycans from the in-house glycosylation database as well as from the literature. This is in accordance with the highly disturbed glycan biosynthesis observed during cancer progression that has been shown to account for phenotypic aggressiveness (Peixoto *et al.* 2019). This study included glycans that are typically found on N and O-glycosylated proteins and whose addition results in a high increase of the total protein weight. A high number of glycosylation alterations is found in glioma with both a disturbed level of glycosylated proteins and an altered expression of glycosylation enzymes that results in an aberrant glycosylation pattern found on these proteins (Veillon *et al.* 2018, Tondepu *et al.* 2022). Some examples of glycosylation aberration not found in healthy tissue but detected in glioma include under-glycosylation of the ECM components such as brevican (Viapiano *et al.* 2005) and increase of some heparane sulphate proteoglycans (HSPGs) such as syndecan-1 (Veillon *et al.* 2018). In some cases, the link between the aberrant glycosylation and glioma progression was found. For example, the increased terminal sialylation on glycans, including integrins, was found to promote cell invasion and immunosuppression (Veillon *et al.* 2018, Tondepu *et al.* 2022), increased branching of N-glycans in GBM supports invasion, and overexpression of heavily glycosylated mucins supports invasion through activation of oncogenic signalling pathways (Li *et al.* 2014, Pang *et al.* 2022, Tondepu *et al.* 2022). Importantly, these altered glycosylation events were shown to contribute to the immune-suppressive nature of the glioma microenvironment. In a recent study, GBM cells were shown to overexpress the truncated glycan motif Tn antigen, which is then used to recruit the immune-suppressive tumour-associated macrophages (TAMs) (Dusoswa *et al.* 2020). Overall, the glycosylation pattern of glioma tissue is expected to be highly altered, which explains such differences observed between low-grade and high-grade glioma tissue in the present study, when looking at spectral ranges assigned to glycans. These glycan changes could be exploited in the future as biomarkers for glioma.

5.4.3 Other biomolecular changes between grade II and III tissue gliomas

In addition, changes of intensity between grade II and grade III patients could be assigned to an increase of proteins, haemoglobin, DNA, collagen, and lipids. Proteins, haemoglobin, and DNA are expected to increase during cancer progression due to higher cell density and vascular proliferation. The level of lipid is

more difficult to interpret from dewaxed fixed tissue since the dewaxing procedure has been shown to largely affect the lipid content present in the tissue section (Ali *et al.* 2013). However, since only pair-wise comparison between two grades was made, it is still reasonable to read the potential relative differences in the lipid content. Overall, the lipid composition has been shown highly dysregulated in brain tumour using RS in comparison with healthy cells (Köhler *et al.* 2009). GBM cells are characterised by an increased *de novo* lipid synthesis for energy and plasma membrane production that a high proliferation rate demand (Zhou *et al.* 2019, Depciuch *et al.* 2020, Bukva *et al.* 2021). In addition, an abnormal accumulation of lipid droplets in GBM tissue was shown to play a role in their proliferation (Taib *et al.* 2019). Overall, the lipid content is commonly shown to be higher in GBM (Banerjee *et al.* 2015), which agrees with this study, but this can also vary depending on what type of lipids or cells are examined. For example, the lipid content was shown to be reduced in GBM single apoptotic cells (Ricci *et al.* 2018). Finally, collagen rearrangement, for example, a switch in the level of the different collagen types, has been shown to occur during glioma progression and has been observed using RS in previous studies (Zhou *et al.* 2012, Payne *et al.* 2013, Pointer *et al.* 2017, Riva *et al.* 2021). Collagen is a rare component in the healthy brain microenvironment and significantly increases during glioma progression (Dong *et al.* 2010).

5.4.4 Glycosylation changes between grade III and IV tissue gliomas

The overall difference between grade III and grade IV grades was smaller, which is expected since these two grades are both classified as high-grade malignant gliomas and remain challenging to discriminate with the current diagnostic methods (Tian *et al.* 2019, Wang *et al.* 2019). However, peaks showing significant spectral differences could still be identified. The diagnosis of these two grades can affect the therapeutic decision as well as the prognosis, which justifies a search for more objective and accurate mean of classification. Between these two high grades, the spectral ranges assigned to glycosylation slightly decreased in this study; meaning, the glycosylation signature was overall upregulated at a higher degree (compared to grade II) in grade III than in grade IV, which was surprising at first sight, but was also observed at lower extent at the cellular level in the previous chapter. Glycosylation is a heavy modification that affects more than half of the proteins

found in the brain (Veillon *et al.* 2018). In addition, it is not possible with Raman spectroscopy to detect to which of the specific proteins each glycosylation can be assigned. It is thus complicated to predict to what extent the total level of glycosylation will be upregulated in each glioma grade.

Overall, an increase of the total glycans was observed in grade III, which persisted in grade IV but was slightly reduced. This might appear surprising at first, however, this finding seems to align with the previous finding of this thesis performed in glioma cells in addition to what have been observed before in glioma. *Furukawa et al.,* performed a serial oncogenic transformation in normal astrocytes, that eventually mimic the four grades of glioma *in vitro*, to study the level of N and O-glycans in each type of cells (Furukawa *et al.* 2015). An important number of glycans, such as core 1 O-glycans, showed an upregulation in grade III cells in comparison with grade II and IV, which showed a very similar level. In comparison to grade II, the level of sialylation (the addition of a neuraminic acid) was largely more elevated in grade III than in grade IV (Furukawa *et al.* 2015). In addition, as previously mentioned, different glycans have been found in their truncated version in grade IV glioma (Dusoswa *et al.* 2020), which implies that the total level of glycans could be downregulated in this type of cancer. Generally speaking, the total glycosylation pattern follows a complex dynamic during transformation and could be further exploited for glioma monitoring and diagnostics in future studies.

5.4.5 Other biomolecular changes between grade III and IV tissue gliomas

From the literature, the slight increase in grade IV compared with grade III could also be assigned to an increase in collagen, proteins, lipids, haemoglobin, and DNA, which was already observed between grade II and III samples. This confirms an overall augmentation of proteins, haemoglobin, and DNA during glioma progression, which represents a confirmation of what has been described in previous RS studies (Kast *et al.* 2014, Banerjee *et al.* 2015, Kopec *et al.* 2021). More specifically, this probably reflected an increase in the cell density and a high perfusion rate caused by microvascular proliferation, both histological characteristics of GBM (D'Alessio *et al.* 2019), which have been observed in previous studies using RS (Kast *et al.* 2014). In addition, there was a decrease of carotenoids in grade IV, which has been observed often in RS studies investigating cancer (Mehta *et al.* 2018, Kopec *et al.*

2021, Shrivastava *et al.* 2021) but also from studies measuring β-carotene levels in the serum from different brain tumours (Aggarwal *et al.* 2006). Carotenoids are only brought by food in the body and have antioxidant and anti-cancer properties. According to some studies (Kopec *et al.* 2021), the peak reduction assigned to carotenoids on RS spectra may be used as a biomarker in brain cancer to assess the aggressiveness of the tumour.

5.4.6 Discrimination between m-IDH1 and wt-IDH1 patients in tissue

RS was able to predict with high accuracy the IDH1 genotype of the patients, which is the marker used currently for cytogenetic classification (Louis *et al.* 2021). This result was expected since histological features and the IDH1 genotype are strongly correlated, as discussed earlier. Histological classification (WHO 2016) has been replaced by a cytogenetic and molecular classification (WHO 2021) during the course of this study. It was therefore important to verify if RS could be used in the context of the current gold standard method of central-nervous-system tumour diagnostics. This result is promising, especially since discrimination between two groups (IDH1 genotypes) is expected to be less challenging than discrimination between three groups (grade II, III, and IV). The results of this chapter suggest that clinically-available FFPE slides could be used for IDH1 genotyping. This diagnostic method represents several advantages in comparison with the immunostaining and sequencing examination used currently. First, RS does not require the purchase of different reagents or labels, or any long preparation of the samples, it can be then expected that this method would be less expensive but also faster than IHC labelling or sequencing. In addition, the results are expected to be more objective compared with IHC as they would be less dependent on the investigator's eyes or protocols used.

Moreover, the difference between the two IDH1 genotypes from the comparison of their respective averaged spectrum, was shown to be clearly condensed in the region identified by the glycosylation database as being very characteristics of the glycosylation changes, and no other important changes was observed within the rest of the signature. This is also promising since the assessment of only a selected portion of the spectrum would be potentially enough for diagnosis, which would lower the diagnostic time as well as the complexity of the spectral interpretation.

Finally, this result also shows that the utilisation of RS could lead to a better understanding of the biological changes between the two IDH1 genotypes. Indeed, this study suggests that m-IDH1 patients might have a different level of overall glycosylation than their wild-type counterparts. This could have biological and clinical importance and should be further confirmed and investigated.

5.4.7 Summary of the total biomolecular changes in glioma tissues

In conclusion from the tissue Raman spectral signatures, this approach could be used to classify glioma grades from fixed tissue with high accuracy and was even helpful for discrimination between grade III and IV glioma samples. Changes in the glycosylation patterns seemed to largely account for the successful discrimination between the different grades, with an increase of glycosylation specifically observed in grade III followed by a decrease in grade IV, which permits a good discrimination between the grades. **Figure 5.24** summarises the changes occurring between the grades in glioma at the cellular and tissue levels from the literature, that can be put in parallel with the findings of the present study, which further supports the fact that RS could permit the monitoring of biological changes during cancer progression.

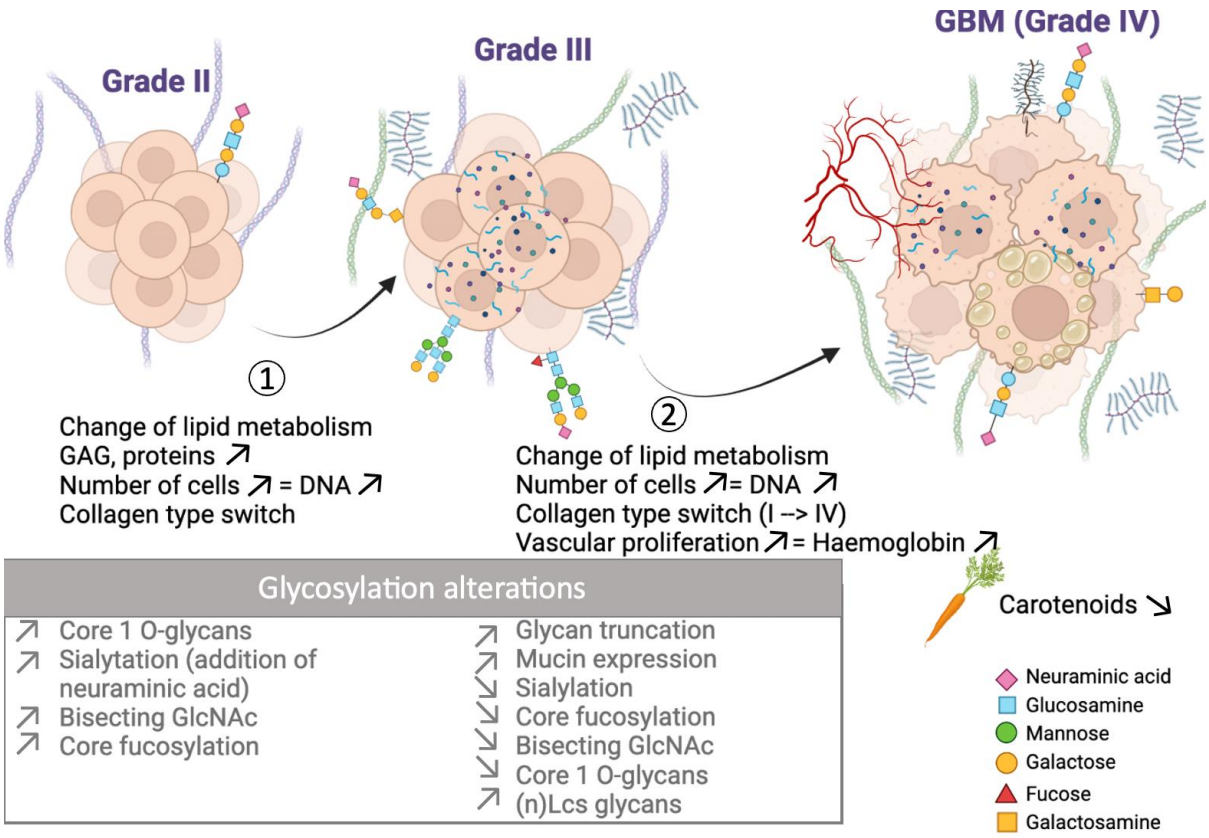


Figure 5.24 Summary of biomolecular changes identified in agreement with known changes in glioma. (1) Changes observed between grade II and grade III, and **(2)** between grade III and grade IV, according to the literature, can be put in parallel with the results of this study. A number of glycosylation patterns increase between grade II and grade III; and decrease between grade III and grade IV. [N-acetylglucosamine (GlcNAc)]. (Kast *et al.* 2014, Furukawa *et al.* 2015, D'Alessio *et al.* 2019, Taib *et al.* 2019, Dusoswa *et al.* 2020, Kopec *et al.* 2021). *Figure created with Biorender.com.*

One potential limitation of this approach was the collection length of single spectra performed manually on the tissue samples, which was assumed to be necessary to obtain a very clear discrimination between the three grades without any overlapping. To compensate for the high intra-patient heterogeneity present in tissue, several acquisitions on different spots across the tumour were interrogated in each patient. The total acquisition time per each patient took about one hour, which remains however potentially less time consuming than IHC analysis.

5.4.8 Classification of glioma grades from blood serum

Using liquid biopsies such as blood serum would represent an advantage compared to tissue biopsies since collecting this type of samples is minimally invasive and can be done multiple times to monitor disease progression. Serum has the advantage of not requiring any chemical processing or sample preparation before acquisition and thus, can be interrogated while being in the native form as found *in vivo*. Therefore, this type of sample conserves all the biological information that can be lost during sample processing. It should also be noted that serum, unlike solid biopsies that constitute only a portion of the tumour, can be more representative of the tumour pathophysiology.

Using the same methodology as previously, serum samples from benign non-glioma controls, grade III, and grade IV were interrogated to see primarily whether it would be useful for discrimination between the two high-grades glioma patient groups. A higher accuracy was observed between the groups with fresh (liquid) serum, but not with dried serum, in comparison with tissue. Using fresh serum, the three largest PCs were enough to explain 85% of all variances between the samples. Interestingly, high heterogeneity was observed within the high-grade glioma samples when using liquid serum, while the control group showed a much smaller distribution on the PCA plot. High inter-patient heterogeneity is expected among blood samples from cancer patients and has been described by previous RS studies, notably in nasopharyngeal cancer (Feng *et al.* 2010). This suggests that biological changes occurring during cancer progression can take different directions, which are purely patient-specific, and that importantly, are reflected in their blood serum. This supports the idea that glycosylation could be exploited for personalised medicine in liquid biopsies (Ferreira *et al.* 2021).

5.4.9 Glycosylation and other biomolecular changes in glioma blood serum

Overall, the changes between the groups were more obvious with more significant spectral differences in the serum and the same trend was observed between the tissue and the serum with an increased glycosylation level in high-grade glioma, which was largely sharper in grade III compared to grade IV and with a more significant difference in the serum than in the tissue. This suggests a promising use of the glycosylation pattern from liquid biopsies for cancer and specifically glioma

diagnostics and progression monitoring permitted by minimal invasiveness of blood collection. This also represents promise to understand the altered glycosylation processes occurring during cancer progression. The presence of individual glycoproteins in blood serum has already been shown useful for cancer diagnostics (Kirwan *et al.* 2015, Veillon *et al.* 2018). Glycoproteins such as mucins, have been shown to be in their truncated/hypoglycosylated forms (sialyl-Tn, Tn and T antigens) specifically in advanced cancer and have already been suggested as serum biomarkers (Singh *et al.* 2007, Nath *et al.* 2014, Radhakrishnan *et al.* 2014, Kudelka *et al.* 2015, Hanson *et al.* 2016, Munkley 2019, Giamougiannis *et al.* 2021).

Some other biomolecular changes could be also assigned to some extent to changes in the lipid content in the serum, which could be due to the dysregulated lipid biosynthesis observed during cancer progression as earlier discussed. More importantly, the level of carotenoids was reduced in the serum from grade IV glioma patients compared with the lower grades. The carotenoid level has been shown to be decreased in the serum of cancer patients in previous studies in cervical cancer (Shrivastava *et al.* 2020), in meningioma (Mehta *et al.* 2018), or in breast cancer (Eliassen *et al.* 2012) and therefore, it has been concluded that the high level of circulating carotenoids, which have antioxidant properties, could protect against cancer. Finally, a switch in intensity that could be assigned to proteins were also observed in serum from grade IV patients. This could correspond to a change in the secondary structure of proteins that can be explained by a high mutation rate a change of their regulation (Depciuch *et al.* 2020).

5.4.10 Discrimination between m-IDH1 and wt-IDH1 patients in serum

RS was able to discriminate with good accuracy the two IDH1 genotypes from fresh serum, although the number of patients in this case was limited (7 m-IDH1 and 12 wt-IDH1). Like for the tissue, this bolsters the use of RS in clinics, especially because in the case of fresh serum, the sample can be collected less invasively at multiple time points. A potential RS-driven IDH1 genotyping made possible from a small drop of blood serum is very attractive, since it would represent a fast, non-invasive, objective, and non-expensive diagnostic strategy for glioma. Therefore, it is important to further investigate this strategy on a higher number of glioma patients.

5.4.11 Summary of the total biomolecular changes in glioma blood serum

To conclude, RS on blood serum gave very good classification accuracy, even for discrimination between grade III and IV glioma, when using fresh serum. In addition to give better accuracy, this method presented several advantages compared to the diagnostic method using tissue. First, the acquisition time was reduced by five (about 20 minutes per patient), owing to the reduced number of acquisitions that were necessary to obtain the grade discrimination, which was made possible by very good intra-patient homogeneity of the sample observed with liquid biopsies. Even more than in tissue, changes in the glycosylation pattern seemed to account greatly to the discrimination between the groups.

Chapter 6: Discussion and future perspectives

6.1 Summary

6.1.1 General overview and discussion

MUC4 protein is a heavily glycosylated transmembrane protein implicated in carcinogenesis and invasion, partly through interaction with EGFR family members (Carraway *et al.* 2009). However, the role of MUC4 in the gelatinases' expression and more particularly MMP9 through interaction with EGFR remains to be defined. Moreover, given that MUC4 is not expressed in the normal brain, this mechanism has never been studied in brain cancer and warrants further investigations.

In this thesis, the role of three molecular players (MUC4, EGFR, and MMP9) in glioma was evaluated. First, the importance of these markers needed to be demonstrated in patients to justify their clinical relevance before further mechanistic investigations could be conducted. The clinical study in glioma patients demonstrated that MUC4 was at least as important as MMP9 in terms of grade prediction (diagnosis) and outcome prediction (prognosis), its expression was even more extended in the tissue and was present in the microvascular proliferation (MVP) structures of more than half of the GBM patients. MUC4 is not expressed in the normal brain, and therefore, may present high specificity as a potential biomarker for glioma. However, detecting MUC4 in liquid biopsy, which represents a less invasive mean of sample collection for the patients, showed only low protein levels, which is generally the case with transmembrane markers as they are not all secreted at high levels. MUC4 seems to be nevertheless a potent tissue diagnostic/prognostic biomarker, with important implication in gliomagenesis, especially if the clinical information can be combined with EGFR and MMP9 expression.

The expression of MUC4 in the MVP structures of GBM patients constitutes one of the most important findings of the present study. MUC4 has been shown to promote invasion and angiogenesis in carcinoma, which was associated in both cases with an increase of MMP9 expression (Ponnusamy *et al.* 2010, Tang *et al.* 2016). The present study indicates that it could also be the case in glioma. Although the

mechanism remains to be defined, it can be proposed that MUC4 expression in glioma cells might induce an angiogenic switch, which may be linked to MMP9 expression and therefore augmentation of VEGFR activation. MMP9 has been proposed to act mostly at early stage of angiogenesis (Hamano *et al.* 2003) and specifically expressed according to the present study in what seem to be nascent small MVP structures. MUC4, on the other hand, seems to act only (or also) during the latter stage of the process, since its expression was observed in larger and more numerous MVP structures.

Next, *in vitro* assays were performed to demonstrate the mechanistic interaction – physical and regulatory – between the different players of the pathway under investigation, as well as the effect of MUC4 on the cells. A plant-based drug, TMQ, that was shown to reduce MUC4 protein level in carcinomas (Torres *et al.* 2010), was also tested as an attempt to demonstrate that this pathway not only exists and is implicated in glioma, but also, can be therapeutically targeted. This study needs confirmation and further experiments but, overall, these assays gathered indicate that MUC4 indeed displayed anti-adhesive properties that might play a role in glioma cell dissemination. Glioma is a cancer showing particular diffuse characteristics, which aligns with high dissemination properties. MUC4 was undoubtedly interacting with EGFR in some of the cells, which is coherent with the fact that these two proteins were both extensively expressed in tissue samples from GBM patients according to the IHC results. TMQ was able to drastically reduce the presence of the protein, and consequently, this interaction. Finally, gene expression analysis supported the existence of a common regulatory pathway and the relevance of TMQ use for its downregulation. This plant-based treatment represents several advantages such as a good tolerance, and might be a good strategy for glioma patients, at least as an adjuvant or even a preventive drug in future clinical studies.

Using a broader approach on glioma cell lines, Raman spectroscopy (RS) was used to assess whether this tool - with recent applications in biological/biomedical research but not used in clinics yet - could represent an objective means of monitoring and diagnosis for glioma patients with a particular focus on parallel glycosylation detection. Glycosylation is highly affected in brain cancer (Veillon *et*

al. 2018) and although MUC4 overexpression in tissue seemed important, its aberrant glycosylation and expression would only represent one aspect of this broad change. Moreover, by looking at the literature, glycan changes seemed to be under-represented in Raman cancer studies and publicly-available Raman database. For these reasons, the spectroscopic approach was focused on the general pattern of glycosylation and a database containing the typical glycans found on mucins and other glycoproteins was made for reference. This approach could be applied successfully on live cells including the potential detection of small differences induced by a specific change in the glycosylation pattern. This last aspect will need to be confirmed with higher acquisition numbers and/or more optimisation. In particular, improvement of Raman technology and methodology, as well as improvement of the publicly-available database, will be needed for rigorous and systematic application of Raman spectroscopy in the biological and medical field.

Importantly, this methodology combined with classification learning approaches was useful to discriminate grade III and grade IV glioma from tissue and serum samples with an accuracy of 85% and 90%, respectively. The acquisition time was shorter when using the serum and the sample preparation was also easier, since the samples did not require deparaffinisation or any other preparation. Changes in the glycosylation have been identified and were shown to be largely responsible for the overall difference observed. In addition, a very small volume of the sample used as a liquid in its native state was sufficient. Therefore, the liquid biopsy approach seems to be more relevant for clinical applications, especially since it also represents a better option for the patients' wellbeing. Overall, RS hold great promise for objective diagnosis and monitoring with minimal pain.

6.1.2 Summary of methodological approaches

The general goal of this study was to investigate new mechanisms of glioma progression at the molecular level, to suggest new diagnostic/prognostic strategies and potentially, therapeutic approaches for future improvements of the GBM standard of care. More particularly, the study was articulated around mucins and glycosylation, whose involvement and relevance were considered to be still

overlooked in glioma. For this purpose, two methodological approaches were followed:

A) A specific approach: To characterise a suggested MUC4-mediated pathway in glioma, clinical and *in vitro* investigations were conducted as followed:

- EGFR, MUC4, and MMP9 expression was investigated using Immunohistochemistry (IHC) and immunofluorescence (IF) on 60 glioma patients of different histological grades. Staining was scored considering both intensity and extent of expression and the scores were compared between the grades (quantitative IHC approach). The localisation of each protein was also analysed in detailed (qualitative IHC approach).
- The Kaplan-Meier method was used to correlate the different levels of expression, in combination or separately, with the clinical outcome of the patients.
- ELISA assays were used to measure the level of the proteins (different MMP9 forms and MUC4) in the blood serum of the patients, to further assess the clinical significance of these markers in liquid biopsies as non-invasive approaches.
- Spheroids were generated with the hanging drop method from control and MUC4-overexpressing GBM cells. The shape and the number of spheroids could be used to evaluate the effect of MUC4 protein on the cells in a 3D context (spheroids). This was performed with and without the addition of TMQ to further understand how this drug could be relevant in MUC4 targeting therapies for GBM patients.
- Immunocytochemistry (ICC) with EGFR and MUC4 antibodies in GBM cells was also performed to visualise their localisation in the cells and to assess the effect of TMQ on the protein levels.
- RT-qPCRs in multiple contexts, notably, using *in vitro* cellular models grown in the presence of the putative ligands of the proposed pathway, were finally utilised to characterise the genetic regulatory mechanisms of this pathway.

B) A broader approach: To optimise an objective mean of glioma monitoring and discrimination, taking advantage of the highly altered glycosylation patterns, a Raman spectroscopy study was performed on *in vitro* cellular models and clinical samples:

- Raman spectroscopy (RS), a tool whose utilisation in biology is still under optimization, was used and optimised in glioma cell lines. The detection sensitivity of the RS approach was finally tested by comparing RS signals between control and MUC4-overexpressing cells.
- The same methodology was followed on different clinical samples (fixed tissue and serum). A high number of spectral acquisitions were collected, processed, and averaged on both glioma tissue and serum samples. Principal component analysis (PCA) and built-in classification machine learning tools were used to show whether classification of glioma grades with high accuracy was possible (according to their histology and even potentially IDH1 genotype).
- In parallel, biological changes identification was attempted. Since no objective RS database exists yet, peak assignments were made by comparison with an in-house glycosylation database, as well as with a general literature-based database as an attempt to avoid underestimating the importance of any molecule.

6.1.3 Summary of results

The results of the present study first supported different previous observations and were coherent when compared to the literature. First, EGFR overexpression was not specific to GBM and seemed to be totally absent from the visible microvascular proliferation (MVP) structures in GBM patients (Karpel-Massler *et al.* 2009, Rodriguez *et al.* 2012, Saadeh *et al.* 2018). In addition, MMP9 behaved as a diagnostic marker for glioma patients, expressed specifically in nascent MVP structures and haematological cells (Choe *et al.* 2002, Jiguet-Jiglaire *et al.* 2022). In addition, EGFR and MMP9 had common genetic regulation in GBM cells (Chang *et al.* 2015, Chen *et al.* 2017). Finally, Raman spectroscopy showed potential to objectively discriminate different cancer types and could detect characteristic cancer-related biomolecular changes between the grades, such as increased

protein, haemoglobin, lipid, DNA in grade IV, as well as a switch of collagen, and a decreased level of carotenoids (antioxidants) (Krafft *et al.* 2012, Auner *et al.* 2018, Kopec *et al.* 2021).

These observations were useful to confirm the successful conduct of this thesis and to verify potential suggested mechanisms. More importantly, this PhD thesis showed a set of novel findings, listed below in a point-by-point basis:

First approach (clinical and in vitro cellular model studies):

- Two types of EGFR expression could be observed in GBM-associated vessels: one expression in vessels that might result from neoplastic cells (EGFR-positive) following a vasculo-mimicry event, and one in vessels developing from existing normal vessels (EGFR-negative), which would solve a controversy since vasculo-mimicry has been proposed in glioma but EGFR-expressing vessels were not commonly reported in GBM tissue.
- The number of MMP9-expressing haematological cells were observed in low-grade and were increased in higher glioma grades.
- MUC4 was overexpressed during gliomagenesis and could represent a useful glioma diagnostic biomarker. The diagnostic and prognostic efficiency of MUC4 was at least as good as the one of MMP9. MUC4 protein expression in glioma tissue was more extended than the one of MMP9 with presence in the entire tumour. The expression seemed to appear from low-grade, earlier than MMP9, and then was elevated in the higher grades. The difference between the grades was even more significant with MUC4 than with MMP9. Finally, MUC4 was strongly associated with wt-IDH1 genotype, which is itself associated with aggressive glioma.
- MUC4 was expressed in the MVP structures of GBM patients and could play an important role in GBM angiogenesis. In more than half of GBM patients, MUC4 was expressed in the glioma-associated vessels. This expression was unambiguous and was observed in more than half of the patients, and more dramatic and systematic than MVP MMP9 expression, which seemed to be more restricted and confined to small nascent MVP structures.

- MUC4 was expressed with EGFR and with MMP9 (separately) in the same GBM MVP structures. This suggests that these three proteins could interact in a common pathway in the GBM vessels, with a role in their formation.
- EGFR tissue expression alone was not useful at predicting the clinical outcome of glioma patients. The difference between the two groups of expression was not significant but there was no differential trend in OS and PFS. On the other hand, MUC4 and MMP9 high expression seemed to present a trend of worsen outcome although it was not significant. When combining MMP9 with MUC4, with or without EGFR, the trend of worsen prognosis was significant.
- The level of MMP9 protein, associated with its inhibitor TIMP1 and alone, in the serum was significantly elevated in grade IV than in grade III patients, while MUC4 secreted protein levels were only minimally elevated.
- The transient overexpression of MUC4 in GBM cells using a spheroid (3D model) led to smaller and more numerous spheroids either by impeding the fusion or supporting the dissociation. In both cases, MUC4 seemed to trigger GBM cell dissemination in 3D context.
- TMQ completely impeded spheroid formation in control (un-transfected) GBM cells and to a lesser extent, in transfected MUC4-overexpressing cells.
- In GBM cells, EGFR expression dominated MUC4 expression. After overexpression of MUC4, EGFR and MUC4 co-localised in the cells, more particularly, both proteins seemed to be restricted to one pole of the membrane. In parallel, EGFR expression was abrogated at the mRNA level.
- TMQ was able to downregulate MMP9 at the gene expression level. This downregulation was more efficient than the one induced by the standard chemotherapy drug TMZ.

Second approach (Raman spectroscopy and machine learning):

- Raman spectroscopy (RS) was able to discriminate with high accuracy between different glioma cell lines, with only one acquisition per cell in

monolayer (2D) and in spheroid (3D) models, between grade III and two different types of grade IV cell lines.

- The discrimination between normal GBM cells and MUC4-overexpressing GBM cells, as expected, was more challenging with RS, but was still possible within the first five PCs. This suggests that the changes induced by the MUC4 transient overexpression could be captured by the RS.
- RS was able to discriminate with high accuracy between the glioma histological grades and importantly, was even able to discriminate grade III and grade IV glioma with an accuracy of 85%. RS was also able to discriminate between the two IDH1 genotypes (wild-type and mutated) with an accuracy of 80%.
- In serum samples, RS could discriminate between grade III and grade IV with an accuracy of 90%, which was even higher than the accuracy obtained from the tissue. IDH1 discrimination from the serum samples needs to be repeated on a higher number of samples but looked very promising (a trend of discrimination was observed and an accuracy of 85% was reached with logistic regression).
- Some Raman spectral peaks could be used for objective diagnosis in future clinical applications as they displayed significant changes between the grades in both tissue and serum.
- Overall, changes in glycosylation patterns seemed to account for the most important biomolecular changes observed between the grades in both tissue and serum.

6.2 Limitations and future perspectives of the current study

The present study was successful at reaching the main objectives stated in the beginning of this thesis and to facilitate the understanding of the importance of mucins and glycosylation changes in glioma. However, some limitations and further questions could be raised and should be discussed. In the clinical study, IHC and survival analysis were performed on 60 glioma patients. This number was reasonable to conduct a quantitative analysis. However, the fact that these samples were categorised led to a reduction of their number in their respective group. It would

be more ideal to have a higher number of samples in each group, especially in the GBM group, as the proteins of interest displayed an interesting pattern of expression that could be quantitatively and qualitatively further analysed in a higher number of samples. The IHC method remains dependent to the investigator's eyes and so using more objective technique such as RNA and exome sequencing will be necessary from patient samples. Moreover, using laser micro-dissection to isolate the microvasculature regions positive for MUC4 would be an efficient way to perform transcriptomic analysis specifically in these structures to verify that MUC4 gene regulation is altered.

In addition, the survival analysis should consider more parameters. Since the potential biomarkers in this study were also analysed in combination, the groups compared were of relatively modest size. In addition, the present study could be underpowered, which would explain why the trend showing worse survival in patients with high MUC4 and MMP9 individual expression was not significant. We can predict that a study comprising more patients might show significant worsen survival for high MUC4 tissue expression alone. Therefore, a higher number of patients should be involved in the future in a large clinical cohort, especially since MUC4 might be an important oncogene in glioma. This protein should be analysed in more details considering its association with more clinical features such as the age and the time and extent of tumour resection.

The serum analysis was successful at showing a significant increase of MMP9 level in GBM compared with grade III even though the number of patients was limited. This should be repeated on a higher number of patients, chosen, and classified according to the new brain cancer classification. The control group used in this study contained patients with benign non-glioma samples. MMPs have been shown to be involved in a wide variety of diseases (Rybakowski 2009) and it cannot be certain that the level of MMP9 did not have clinical relevance in these patients, which would explain why the inter-patient heterogeneity was high in this group, while it was not in the grade III glioma group. In addition, the grade III glioma patients contained in the serum study were in majority oligodendrogliomas (and not astrocytomas). Finally, this study gave a good first indication of the relevance of MUC4 as a liquid biopsy biomarker. The results were not surprising, since transmembrane proteins

are only rarely secreted by definition. However, it has been shown that these types of proteins, including mucins, had an elevated secretion in cancer (Kufe 2009, Filippou *et al.* 2018). A slight elevation of a protein compared to the average concentration could still give prognostic and diagnostic indication if significant. The trend observed in this study was not significant in the serum, but it may be assumed that measuring this marker in the cerebrospinal fluid would be more relevant, since this type of fluid is in more direct contact with the central nervous system in comparison to the blood. Investigating the incidence of MUC4 elevation in this type of fluid, combined with the one of MMP9, would be necessary.

Of important note, this study was designed in 2019, before the new 2021 WHO classification of central-nervous-system tumours. Future investigations of MMP9 and MUC4 should focus on the comparison between the new glioma groups (i.e., oligodendroglioma, m-IDH1 grade 2,3, and 4 astrocytomas, and wt-IDH1 (new GBM)). An effort was made to collect the IDH1 genotype of all patients to update the analysis. Moreover, the new classification is not expected to significantly change the diagnosis of GBM (histological GBMs were already wt-IDH1 in the vast majority of the cases). Therefore, it is reasonable to say that this limitation was at least partly overcome, but the investigation of oligodendroglioma as a specific separate entity should be done in the future. In addition, *in vitro* studies can be suggested such as assays designed to characterise the role of MUC4 in angiogenesis and invasion using patient-derived cell lines and *in vivo* studies using mice models. In particular it would be important to test if MUC4 expression in glioma is responsible for anti-EGFR antibody therapy resistance in this type of cancer. More importantly, MUC4 specific overexpression in the GBM-related microvasculature could be responsible for bevacizumab (anti-VEGFR anti-angiogenic antibody therapy) resistance observed in GBM patients. Importantly, this GBM-specific bevacizumab resistance remains to be elucidated. Experiments of MUC4 knockout and/or inhibition with TMQ in mice should be conducted with concomitant bevacizumab administration to answer that question. Combination of anti-MUC4 and anti-VEGFR treatments should be tested and could represent a very interesting therapeutic strategy since glioma is characterized by a highly disturbed vascular system and high angiogenic rate. The implication of MUC4 protein in angiogenesis should also be tested after

knockout or inhibition, the level of CD31-positive cells and haemoglobin within the tissue should be measured.

Once the role of MUC4 protein is more characterized in glioma cells, it would be crucial to identify the different glycosylation variants of MUC4 that could be more or less related to cancer progression. Variations in the glycosylation pattern could directly play a role in cancer progression and also be used as cancer bio-markers. After isolation of MUC4 protein and/or MUC4-positive angiogenic structures, the glycosylation pattern can be explored with techniques such glycoblotting, O-glycomics, mass spectrometry, and Raman spectroscopy. Moreover, the effect of TMQ on MUC4 glycosylation could be tested with that approach.

To understand the molecular mechanisms behind the clinical results, a set of *in vitro* assays was conducted. This study gave indications on the involvement of the three proteins in a common pathway that might be responsible for gliomagenesis, as well as the relevance of TMQ. However, it is clear that *in vitro* investigations will require more time and efforts and that these experiments need to be repeated with more studied conditions. Numerous additional *in vitro* assays can be suggested including toxicity assays with TMQ in combination with standard therapies as well as gene silencing of MUC4 followed by invasion assay (e.g., assessment of the formation of spheroid protrusions within a matrix). In addition, a stable MUC4 transfection assay needs to be optimised to control the level of MUC4 expression in the cellular models. Overall, it will be useful to measure the effect of MUC4 and the different drugs on the overall cellular oncogenic behaviour (invasion, proliferation, survival, chemoresistance). Co-immunoprecipitation assay would be necessary to prove the MUC4-EGFR interaction. Using different MUC4 parts and variants as bait proteins would allow the identification of critical domains for this interaction. Moreover, once the critical MUC4 domain is identified, small molecules inhibitors of protein-protein interaction could be used to prevent this interaction. This could represent a therapeutic strategy as well as an *in vitro* approach to study the impact of this interaction in glioma cells. To delineate the suggested pathway and regulation between its constituents at protein and mRNA level, additional RT-qPCRs, and western blots (or ELISA, which constitutes a more sensitive technique) will be needed, and notably, the intermediate transcription factors implicated in the

pathway remain to be identified. The level of MUC4 expression need to be measured in different patients (patient derived-cells). One assay should consist in transfecting MUC4 in glioma cells and measuring the impact on the cellular morphology, appearance of lamellipodia, actin rearrangement, motility ability, and activation of downstream invasion/angiogenesis actors (MMP9, MMP2, VEGF, vimentin…).
Finally, it will be interesting to measure the level of MUC4 protein in extracellular vesicles secreted from glioma cells.

Within the spectroscopic approach context, the most important limitation was acquisition time. To be rigorous, RS comparison needs to be performed preferentially between spectra acquired within a relatively short time-period. This is justified by the overall high sensitivity of the Raman technique, which is influenced by different parameters, even the temperature in the room (Fujioka, 1929). Given the amount of time it takes to collect manual acquisitions at different spots in one sample, the routine collection of numerous samples is expected to be somewhat time-consuming for investigators. Moreover, a complete and publicly-available database for the interpretation of the signal is necessary to overcome the current situation that relies on a degree of subjectivity. This limitation was partly overcome in this study using an in-house database in addition to a more general literature-based database. Therefore, improvement of RS analysis will lie in the improvement of the overall methodology and optimisation of its applicability in the medical field. In respect of the present analysis, the discrimination of IDH1 genotypes from a small drop of liquid serum deserves a special attention given the potential help it could provide in the future.

6.3 Working model and conclusion

In conclusion, this present PhD thesis provided a set of evidence to support a working model separated into two approaches that can be presented and further investigated in parallel:

1) The MUC4/EGFR/MMP9 is involved in gliomagenesis. This axis could constitute a diagnostic/prognostic biomarker panel as well as a potential therapeutic target axis.
2) The resulting aberrant glycan pattern observed in glioma could potentially be captured by Raman spectroscopy from tissue samples collected after surgery, but also at different time points from blood serum and cells. This pattern can be captured broadly, but also more specifically after future methodological optimisation.

Figure 6.1 below represents the working model and the concluding remarks.

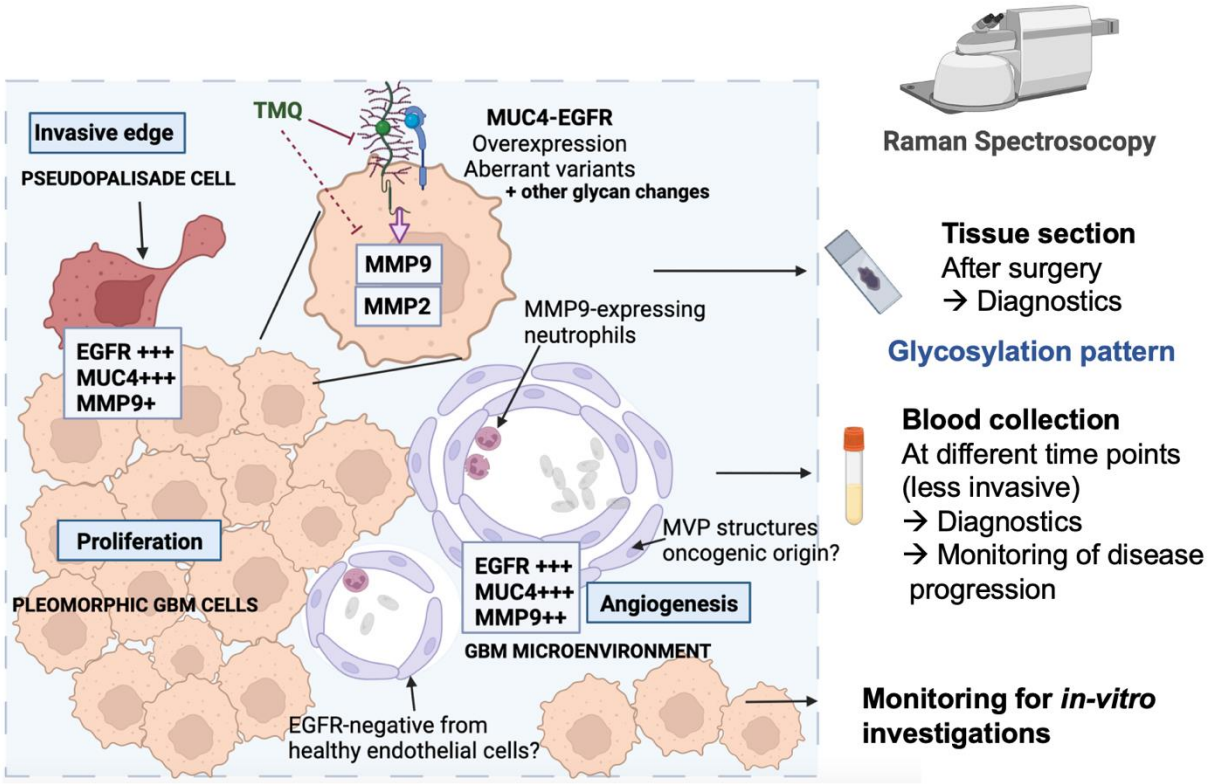
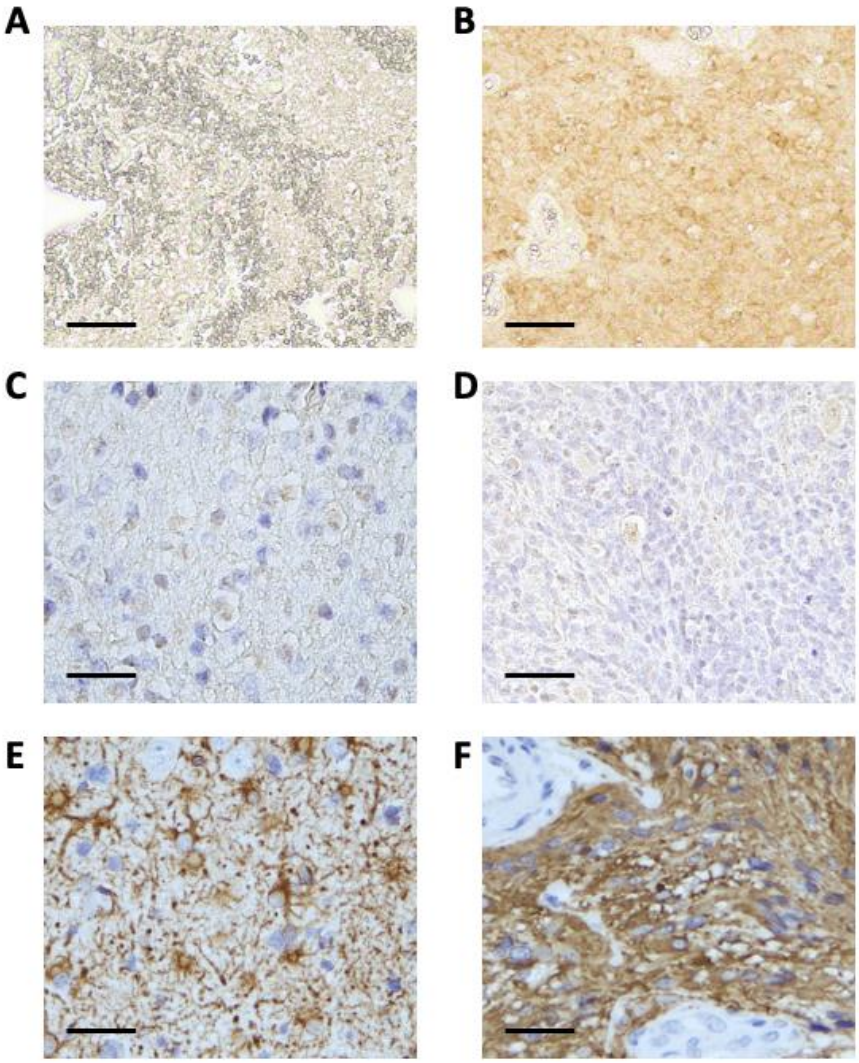


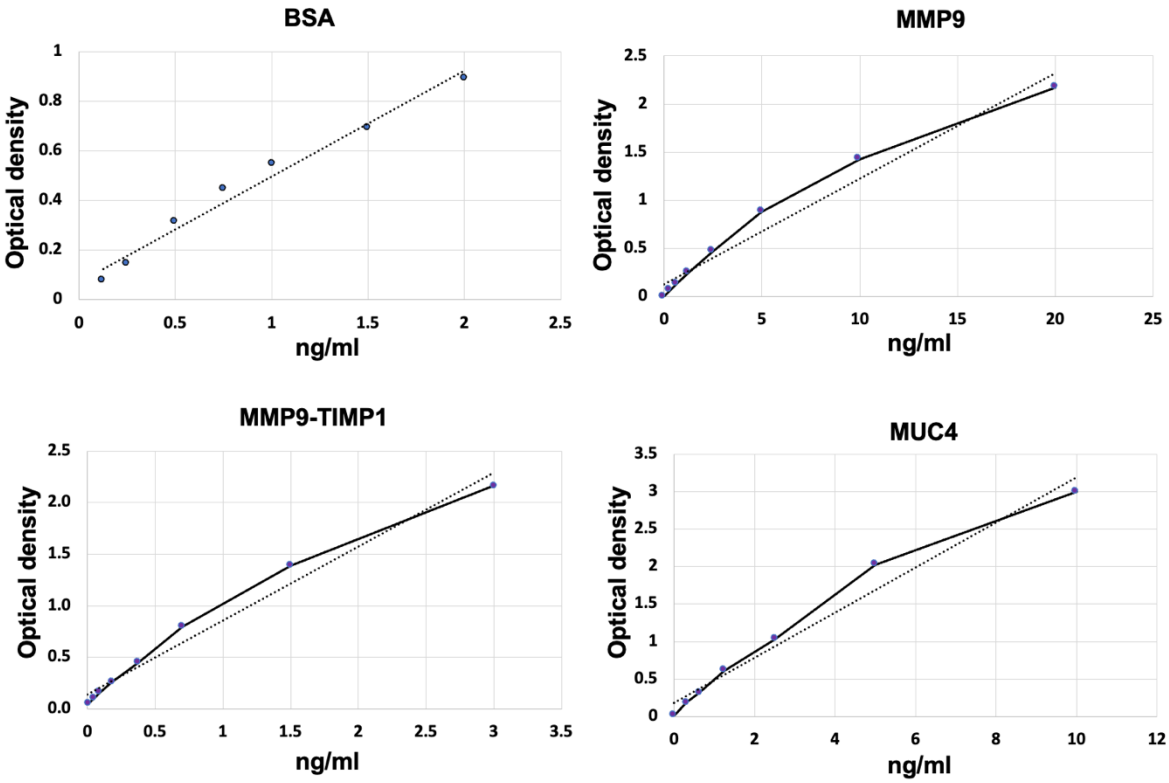
Figure 6.1 Working model of the present study and concluding remarks. MUC4 protein is overexpressed in GBM. Its expression can be observed in glioma neomorphic cells, as well as in MVP structures. This supports a new role of MUC4 in angiogenesis. In some of these structures, the MUC4 expression is associated with EGFR and MMP9 expression. Thus, these particular vessels could have oncogenic origin, while other vessels, dominant in number, might be born from endothelial cells (EGFR-negative), and still express potentially MUC4, which is abundantly expressed in vessels and might have a direct role in classical angiogenic event. In glioma cells, MUC4 and EGFR are overexpressed and interact in the presence of their respective ligands (Gal-3, NRG1, and EGF). This activates the downstream signalling pathway of MMP9, and to a lesser extent, MMP2, whose involvement in gliomagenesis have been well described. This oncogenic pathway can be blocked or prevented by the plant-based drug thymoquinone (TMQ) through MUC4 protein degradation. This would also permit anti-EGFR treatments to reach EGFR by rendering it more accessible. These molecular players could be used as diagnostic and prognostic biomarkers. On the other hand, broad and specific glycosylation dynamic can be captured by Raman spectroscopy in the tissue to guide diagnosis and at different time points in the serum for disease monitoring and glioma grade discrimination. Finally, this approach can be

applied directly on GBM cells for *in-vitro* investigations. *Figure created with Biorender.com.*

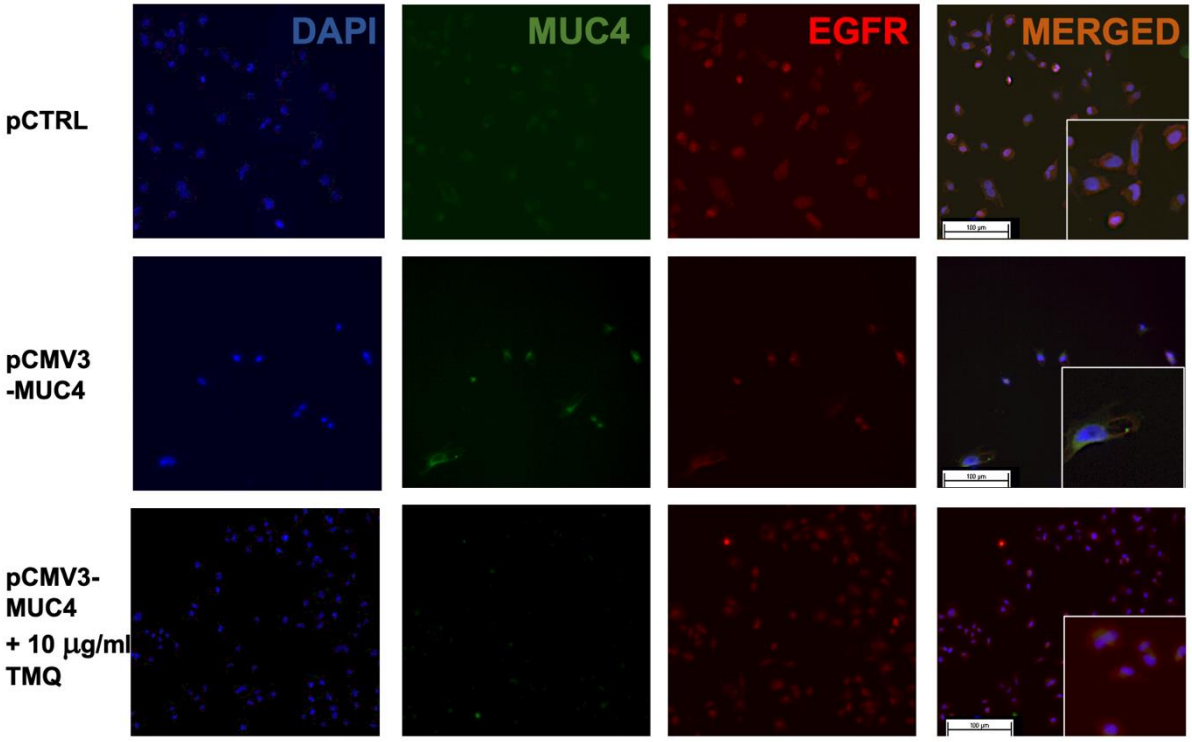
Appendix



Appendix 1. Immunohistochemistry controls. A) Negative control for IDH1 genotyping (with no haematoxylin staining). The antibody recognizing the most common IDH1 mutation was used in an IDH-wild type sample. **B)** Positive control for IDH1 genotyping. The IDH1 antibody was used in an IDH-mutated sample and gave positive staining. **C)** Negative control for IHC protocol validation in grade III glioma with primary antibody omitted. **D)** Negative control for IHC validation in grade IV glioma. **E)** Positive control using GFAP (expressed by any glioma and brain tissue sample) in grade III glioma and **F)** in grade IV glioma. Scale bar = 100 μm.



Appendix 2. Standard curves for Bradford and ELISA assays.



Appendix 3. Individual staining of the ICC assay for MUC4 and EGFR.

References

Abramczyk, H. and A. Imiela (2018). "The biochemical, nanomechanical and chemometric signatures of brain cancer." Spectrochim Acta A Mol Biomol Spectrosc **188**: 8-19.

Acuff, H. B., K. J. Carter, B. Fingleton, D. L. Gorden and L. M. Matrisian (2006). "Matrix metalloproteinase-9 from bone marrow-derived cells contributes to survival but not growth of tumor cells in the lung microenvironment." Cancer Res **66**(1): 259-266.

Aggarwal, S., M. Subberwal, S. Kumar and M. Sharma (2006). "Brain tumor and role of beta-carotene, a-tocopherol, superoxide dismutase and glutathione peroxidase." J Cancer Res Ther **2**(1): 24-27.

Akiyama, Y., S. Jung, B. Salhia, S. Lee, S. Hubbard, M. Taylor, T. Mainprize, K. Akaishi, W. van Furth and J. T. Rutka (2001). "Hyaluronate receptors mediating glioma cell migration and proliferation." J Neurooncol **53**(2): 115-127.

Ali, S. M., F. Bonnier, A. Tfayli, H. Lambkin, K. Flynn, V. McDonagh, C. Healy, T. Clive Lee, F. M. Lyng and H. J. Byrne (2013). "Raman spectroscopic analysis of human skin tissue sections ex-vivo: evaluation of the effects of tissue processing and dewaxing." J Biomed Opt **18**(6): 061202.

Alzial, G., O. Renoult, F. Paris, C. Gratas, A. Clavreul and C. Pecqueur (2022). "Wild-type isocitrate dehydrogenase under the spotlight in glioblastoma." Oncogene **41**(5): 613-621.

Anthony, S. M. and J. A. Timlin (2017). "Removing Cosmic Spikes Using a Hyperspectral Upper-Bound Spectrum Method." Appl Spectrosc **71**(3): 507-519.

Arboleda, P. H. and G. R. Loppnow (2000). "Raman spectroscopy as a discovery tool in carbohydrate chemistry." Anal Chem **72**(9): 2093-2098.

Arcella, A., F. Limanaqi, R. Ferese, F. Biagioni, M. A. Oliva, M. Storto, M. Fanelli, S. Gambardella and F. Fornai (2020). "Dissecting Molecular Features of Gliomas: Genetic Loci and Validated Biomarkers." International journal of molecular sciences **21**(2): 685.

Ardi, V. C., T. A. Kupriyanova, E. I. Deryugina and J. P. Quigley (2007). "Human neutrophils uniquely release TIMP-free MMP-9 to provide a potent catalytic stimulator of angiogenesis." Proc Natl Acad Sci U S A **104**(51): 20262-20267.

Aubertin, K., V. Q. Trinh, M. Jermyn, P. Baksic, A. A. Grosset, J. Desroches, K. St-Arnaud, M. Birlea, M. C. Vladoiu, M. Latour, R. Albadine, F. Saad, F. Leblond and D. Trudel (2018). "Mesoscopic characterization of prostate cancer using Raman spectroscopy: potential for diagnostics and therapeutics." BJU Int **122**(2): 326-336.

Auner, G. W., S. K. Koya, C. Huang, B. Broadbent, M. Trexler, Z. Auner, A. Elias, K. C. Mehne and M. A. Brusatori (2018). "Applications of Raman spectroscopy in cancer diagnosis." Cancer Metastasis Rev **37**(4): 691-717.

Babu, R. and D. C. Adamson (2012). "Rindopepimut: an evidence-based review of its therapeutic potential in the treatment of EGFRvIII-positive glioblastoma." Core Evid **7**: 93-103.

Banerjee, H. N., A. Banerji, A. N. Banerjee, E. Riddick, J. Petis, S. Evans, M. Patel, C. Parson, V. Smith, E. Gwebu and S. Voisin (2015). "Deciphering the Finger Prints of Brain Cancer Glioblastoma Multiforme from Four Different Patients by Using Near Infrared Raman Spectroscopy." J Cancer Sci Ther **7**(2): 44-47.

Barthel, L., M. Hadamitzky, P. Dammann, M. Schedlowski, U. Sure, B. K. Thakur and S. Hetze (2022). "Glioma: molecular signature and crossroads with tumor microenvironment." Cancer and Metastasis Reviews **41**(1): 53-75.

Beiriger, J., A. Habib, N. Jovanovich, C. V. Kodavali, L. Edwards, N. Amankulor and P. O. Zinn (2022). "The Subventricular Zone in Glioblastoma: Genesis, Maintenance, and Modeling." Front Oncol **12**: 790976.

Bekes, E. M., B. Schweighofer, T. A. Kupriyanova, E. Zajac, V. C. Ardi, J. P. Quigley and E. I. Deryugina (2011). "Tumor-recruited neutrophils and neutrophil TIMP-free MMP-9 regulate coordinately the levels of tumor angiogenesis and efficiency of malignant cell intravasation." Am J Pathol **179**(3): 1455-1470.

Bellail, A. C., S. B. Hunter, D. J. Brat, C. Tan and E. G. Van Meir (2004). "Microregional extracellular matrix heterogeneity in brain modulates glioma cell invasion." Int J Biochem Cell Biol **36**(6): 1046-1069.
Bissett, D., K. J. O'Byrne, J. von Pawel, U. Gatzemeier, A. Price, M. Nicolson, R. Mercier, E. Mazabel, C. Penning, M. H. Zhang, M. A. Collier and F. A. Shepherd (2005). "Phase III study of matrix metalloproteinase inhibitor prinomastat in non-small-cell lung cancer." J Clin Oncol **23**(4): 842-849.
Bitler, B. G., I. Menzl, C. L. Huerta, B. Sands, W. Knowlton, A. Chang and J. A. Schroeder (2009). "Intracellular MUC1 peptides inhibit cancer progression." Clin Cancer Res **15**(1): 100-109.
Borgia, J. A., S. Basu, L. P. Faber, A. W. Kim, J. S. Coon, K. A. Kaiser-Walters, C. Fhied, S. Thomas, O. Rouhi, W. H. Warren, P. Bonomi and M. J. Liptay (2009). "Establishment of a Multi-Analyte Serum Biomarker Panel to Identify Lymph Node Metastases in Non-small Cell Lung Cancer." Journal of Thoracic Oncology **4**(3): 338-347.
Bradley, L. M., M. F. Douglass, D. Chatterjee, S. Akira and B. J. Baaten (2012). "Matrix metalloprotease 9 mediates neutrophil migration into the airways in response to influenza virus-induced toll-like receptor signaling." PLoS Pathog **8**(4): e1002641.
Bukva, M., G. Dobra, J. Gomez-Perez, K. Koos, M. Harmati, E. Gyukity-Sebestyen, T. Biro, A. Jenei, S. Kormondi, P. Horvath, Z. Konya, A. Klekner and K. Buzas (2021). "Raman Spectral Signatures of Serum-Derived Extracellular Vesicle-Enriched Isolates May Support the Diagnosis of CNS Tumors." Cancers (Basel) **13**(6).
Burel-Vandenbos, F., L. Turchi, M. Benchetrit, E. Fontas, Z. Pedeutour, V. Rigau, F. Almairac, D. Ambrosetti, J.-F. Michiels and T. Virolle (2013). "Cells with intense EGFR staining and a high nuclear to cytoplasmic ratio are specific for infiltrative glioma: A useful marker in neuropathological practice." Neuro-oncology **15**.
Canetta, E. (2021). "Current and Future Advancements of Raman Spectroscopy Techniques in Cancer Nanomedicine." Int J Mol Sci **22**(23).
Carraway, K. L., G. Theodoropoulos, G. A. Kozloski and C. A. Carothers Carraway (2009). "Muc4/MUC4 functions and regulation in cancer." Future Oncol **5**(10): 1631-1640.
Chang, L., D. Zhao, H. B. Liu, Q. S. Wang, P. Zhang, C. L. Li, W. Z. Du, H. J. Wang, X. Liu, Z. R. Zhang and C. L. Jiang (2015). "Activation of sonic hedgehog signaling enhances cell migration and invasion by induction of matrix metalloproteinase-2 and -9 via the phosphoinositide-3 kinase/AKT signaling pathway in glioblastoma." Mol Med Rep **12**(5): 6702-6710.
Chaturvedi, D., S. A. Balaji, V. K. Bn, F. Ariese, S. Umapathy and A. Rangarajan (2016). "Different Phases of Breast Cancer Cells: Raman Study of Immortalized, Transformed, and Invasive Cells." Biosensors (Basel) **6**(4).
Chaturvedi, P., A. P. Singh and S. K. Batra (2008). "Structure, evolution, and biology of the MUC4 mucin." Faseb j **22**(4): 966-981.
Chaturvedi, P., A. P. Singh, S. Chakraborty, S. C. Chauhan, S. Bafna, J. L. Meza, P. K. Singh, M. A. Hollingsworth, P. P. Mehta and S. K. Batra (2008). "MUC4 mucin interacts with and stabilizes the HER2 oncoprotein in human pancreatic cancer cells." Cancer Res **68**(7): 2065-2070.
Chaturvedi, P., A. P. Singh, N. Moniaux, S. Senapati, S. Chakraborty, J. L. Meza and S. K. Batra (2007). "MUC4 mucin potentiates pancreatic tumor cell proliferation, survival, and invasive properties and interferes with its interaction to extracellular matrix proteins." Mol Cancer Res **5**(4): 309-320.
Chau, A. H., J. T. Motz, J. A. Gardecki, S. Waxman, B. E. Bouma and G. J. Tearney (2008). "Fingerprint and high-wavenumber Raman spectroscopy in a human-swine coronary xenograft in vivo." J Biomed Opt **13**(4): 040501.
Chen, M., N. McReynolds, E. C. Campbell, M. Mazilu, J. Barbosa, K. Dholakia and S. J. Powis (2015). "The use of wavelength modulated Raman spectroscopy in label-free identification of T lymphocyte subsets, natural killer cells and dendritic cells." PLoS One **10**(5): e0125158.
Chen, R., M. Smith-Cohn, A. L. Cohen and H. Colman (2017). "Glioma Subclassifications and Their Clinical Significance." Neurotherapeutics **14**(2): 284-297.
Chen, X. C., X. T. Wei, J. H. Guan, H. Shu and D. Chen (2017). "EGF stimulates glioblastoma metastasis by induction of matrix metalloproteinase-9 in an EGFR-dependent mechanism." Oncotarget **8**(39): 65969-65982.

Cheng, M. and L. Liu (2020). "MUC15 promotes growth and invasion of glioma cells by activating Raf/MEK/ERK pathway." Clin Exp Pharmacol Physiol **47**(6): 1041-1048.
Chenxi Zhang, Y. H., Bo Sun, Wenli Zhang, Shujun Liu, Jiajia Liu, Hong Lv, Guojun Zhang, Xixiong Kang (2020). "Label-free serum detection based on Raman spectroscopy for the diagnosis and classification of glioma." Journal of Raman Spectroscopy **51**(10): 1977-1985.
Chetty, C., S. K. Vanamala, C. S. Gondi, D. H. Dinh, M. Gujrati and J. S. Rao (2012). "MMP-9 induces CD44 cleavage and CD44 mediated cell migration in glioblastoma xenograft cells." Cell Signal **24**(2): 549-559.
Cho, S., M. Chae, B. Shin, H. Kim and A. Kim (2008). "Akt- and MAPK-mediated activation and secretion of MMP-9 into stroma in breast cancer cells upon heregulin treatment." Molecular medicine reports **1**: 83-88.
Choe, G., J. K. Park, L. Jouben-Steele, T. J. Kremen, L. M. Liau, H. V. Vinters, T. F. Cloughesy and P. S. Mischel (2002). "Active matrix metalloproteinase 9 expression is associated with primary glioblastoma subtype." Clin Cancer Res **8**(9): 2894-2901.
Chugh, S., V. S. Gnanapragassam, M. Jain, S. Rachagani, M. P. Ponnusamy and S. K. Batra (2015). "Pathobiological Implications of Mucin Glycans in Cancer: Sweet Poison and Novel Targets." Biochim Biophys Acta **1856**(2): 211-225.
Cohen, A. L., S. L. Holmen and H. Colman (2013). "IDH1 and IDH2 mutations in gliomas." Current neurology and neuroscience reports **13**(5): 345-345.
Collins, V. P., D. T. Jones and C. Giannini (2015). "Pilocytic astrocytoma: pathology, molecular mechanisms and markers." Acta Neuropathol **129**(6): 775-788.
Corsetti, S., T. Rabl, D. McGloin and G. Nabi (2018). "Raman spectroscopy for accurately characterizing biomolecular changes in androgen-independent prostate cancer cells." J Biophotonics **11**(3).
Cribaro, G. P., E. Saavedra-López, L. Romarate, I. Mitxitorena, L. R. Díaz, P. V. Casanova, M. Roig-Martínez, J. M. Gallego, A. Perez-Vallés and C. Barcia (2021). "Three-dimensional vascular microenvironment landscape in human glioblastoma." Acta Neuropathol Commun **9**(1): 24.
Crow, P., N. Stone, C. A. Kendall, J. S. Uff, J. A. Farmer, H. Barr and M. P. Wright (2003). "The use of Raman spectroscopy to identify and grade prostatic adenocarcinoma in vitro." Br J Cancer **89**(1): 106-108.
D'Alessio, A., G. Proietti, G. Sica and B. M. Scicchitano (2019). "Pathological and Molecular Features of Glioblastoma and Its Peritumoral Tissue." Cancers **11**(4): 469.
D'Amico, R. S., Z. K. Englander, P. Canoll and J. N. Bruce (2017). "Extent of Resection in Glioma-A Review of the Cutting Edge." World Neurosurg **103**: 538-549.
Damayanti, N. P., Y. Fang, M. R. Parikh, A. P. Craig, J. Kirshner and J. Irudayaraj (2013). "Differentiation of cancer cells in two-dimensional and three-dimensional breast cancer models by Raman spectroscopy." J Biomed Opt **18**(11): 117008.
Dang, L., D. W. White, S. Gross, B. D. Bennett, M. A. Bittinger, E. M. Driggers, V. R. Fantin, H. G. Jang, S. Jin, M. C. Keenan, K. M. Marks, R. M. Prins, P. S. Ward, K. E. Yen, L. M. Liau, J. D. Rabinowitz, L. C. Cantley, C. B. Thompson, M. G. Vander Heiden and S. M. Su (2009). "Cancer-associated IDH1 mutations produce 2-hydroxyglutarate." Nature **462**(7274): 739-744.
Daniel C. Harris, M. D. B. (1990). Symmetry and spectroscopy: an introduction to vibrational and electronic spectroscopy, Dover publications.
Davis, M. E. (2016). "Glioblastoma: Overview of Disease and Treatment." Clin J Oncol Nurs **20**(5 Suppl): S2-8.
De Pascalis, I., L. Morgante, S. Pacioni, Q. G. D'Alessandris, S. Giannetti, M. Martini, L. Ricci-Vitiani, M. Malinverno, E. Dejana, L. M. Larocca and R. Pallini (2018). "Endothelial trans-differentiation in glioblastoma recurring after radiotherapy." Mod Pathol **31**(9): 1361-1366.
Delclaux, C., C. Delacourt, M. P. D'Ortho, V. Boyer, C. Lafuma and A. Harf (1996). "Role of gelatinase B and elastase in human polymorphonuclear neutrophil migration across basement membrane." Am J Respir Cell Mol Biol **14**(3): 288-295.

Depciuch, J., B. Tołpa, P. Witek, K. Szmuc, E. Kaznowska, M. Osuchowski, P. Król and J. Cebulski (2020). "Raman and FTIR spectroscopy in determining the chemical changes in healthy brain tissues and glioblastoma tumor tissues." Spectrochim Acta A Mol Biomol Spectrosc **225**: 117526.
Dhanisha, S. S., C. Guruvayoorappan, S. Drishya and P. Abeesh (2018). "Mucins: Structural diversity, biosynthesis, its role in pathogenesis and as possible therapeutic targets." Crit Rev Oncol Hematol **122**: 98-122.
Dong, H., L. Luo, S. Hong, H. Siu, Y. Xiao, L. Jin, R. Chen and M. Xiong (2010). "Integrated analysis of mutations, miRNA and mRNA expression in glioblastoma." BMC Syst Biol **4**: 163.
Dong, X., A. Noorbakhsh, B. R. Hirshman, T. Zhou, J. A. Tang, D. C. Chang, B. S. Carter and C. C. Chen (2016). "Survival trends of grade I, II, and III astrocytoma patients and associated clinical practice patterns between 1999 and 2010: A SEER-based analysis." Neurooncol Pract **3**(1): 29-38.
Du, R., K. V. Lu, C. Petritsch, P. Liu, R. Ganss, E. Passegue, H. Song, S. Vandenberg, R. S. Johnson, Z. Werb and G. Bergers (2008). "HIF1alpha induces the recruitment of bone marrow-derived vascular modulatory cells to regulate tumor angiogenesis and invasion." Cancer Cell **13**(3): 206-220.
Duhem-Tonnelle, V., I. Bieche, S. Vacher, A. Loyens, C. A. Maurage, F. Collier, M. Baroncini, S. Blond, V. Prevot and A. Sharif (2010). "Differential distribution of erbB receptors in human glioblastoma multiforme: expression of erbB3 in CD133-positive putative cancer stem cells." J Neuropathol Exp Neurol **69**(6): 606-622.
Dusoswa, S. A., J. Verhoeff, E. Abels, S. P. Méndez-Huergo, D. O. Croci, L. H. Kuijper, E. de Miguel, V. Wouters, M. G. Best, E. Rodriguez, L. A. M. Cornelissen, S. J. van Vliet, P. Wesseling, X. O. Breakefield, D. P. Noske, T. Würdinger, M. L. D. Broekman, G. A. Rabinovich, Y. van Kooyk and J. J. Garcia-Vallejo (2020). "Glioblastomas exploit truncated O-linked glycans for local and distant immune modulation via the macrophage galactose-type lectin." Proc Natl Acad Sci U S A **117**(7): 3693-3703.
Eliassen, A. H., S. J. Hendrickson, L. A. Brinton, J. E. Buring, H. Campos, Q. Dai, J. F. Dorgan, A. A. Franke, Y. T. Gao, M. T. Goodman, G. Hallmans, K. J. Helzlsouer, J. Hoffman-Bolton, K. Hultén, H. D. Sesso, A. L. Sowell, R. M. Tamimi, P. Toniolo, L. R. Wilkens, A. Winkvist, A. Zeleniuch-Jacquotte, W. Zheng and S. E. Hankinson (2012). "Circulating carotenoids and risk of breast cancer: pooled analysis of eight prospective studies." J Natl Cancer Inst **104**(24): 1905-1916.
Erskine, D., J. P. Taylor, A. Thomas, D. Collerton, I. McKeith, A. Khundakar, J. Attems and C. Morris (2019). "Pathological Changes to the Subcortical Visual System and its Relationship to Visual Hallucinations in Dementia with Lewy Bodies." Neurosci Bull **35**(2): 295-300.
Faoláin, E. O., M. B. Hunter, J. M. Byrne, P. Kelehan, H. A. Lambkin, H. J. Byrne and F. M. Lyng (2005). "Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents." J Histochem Cytochem **53**(1): 121-129.
Felder, M., A. Kapur, J. Gonzalez-Bosquet, S. Horibata, J. Heintz, R. Albrecht, L. Fass, J. Kaur, K. Hu, H. Shojaei, R. J. Whelan and M. S. Patankar (2014). "MUC16 (CA125): tumor biomarker to cancer therapy, a work in progress." Mol Cancer **13**: 129.
Feng, S., R. Chen, J. Lin, J. Pan, G. Chen, Y. Li, M. Cheng, Z. Huang, J. Chen and H. Zeng (2010). "Nasopharyngeal cancer detection based on blood plasma surface-enhanced Raman spectroscopy and multivariate analysis." Biosens Bioelectron **25**(11): 2414-2419.
Ferreira, J. A., M. Relvas-Santos, A. Peixoto, M. N. S. A and L. Lara Santos (2021). "Glycoproteogenomics: Setting the Course for Next-generation Cancer Neoantigen Discovery for Cancer Vaccines." Genomics Proteomics Bioinformatics **19**(1): 25-43.
Filippou, P. S., A. H. Ren, D. Korbakis, L. Dimitrakopoulos, A. Soosaipillai, V. Barak, S. Frenkel, J. Pe'er, M. Lotem, S. Merims, R. Molina, I. Blasutig, D. P. Bogdanos and E. P. Diamandis (2018). "Exploring the potential of mucin 13 (MUC13) as a biomarker for carcinomas and other diseases." Clin Chem Lab Med **56**(11): 1945-1953.
Forsyth, P. A., H. Wong, T. D. Laing, N. B. Rewcastle, D. G. Morris, H. Muzik, K. J. Leco, R. N. Johnston, P. M. Brasher, G. Sutherland and D. R. Edwards (1999). "Gelatinase-A (MMP-2), gelatinase-B (MMP-9) and membrane type matrix metalloproteinase-1 (MT1-MMP) are involved in different aspects of the pathophysiology of malignant gliomas." Br J Cancer **79**(11-12): 1828-1835.

Foty, R. (2011). "A simple hanging drop cell culture protocol for generation of 3D spheroids." J Vis Exp(51).
Friedl, P. and K. Wolf (2008). "Tube travel: the role of proteases in individual and collective cancer cell invasion." Cancer Res **68**(18): 7247-7249.
Friedmann-Morvinski, D., E. A. Bushong, E. Ke, Y. Soda, T. Marumoto, O. Singer, M. H. Ellisman and I. M. Verma (2012). "Dedifferentiation of neurons and astrocytes by oncogenes can induce gliomas in mice." Science **338**(6110): 1080-1084.
Fujioka, Y. (1929). "Influence of Temperature on Raman Lines." Nature **124**(3114): 11-11.
Funes, M., J. K. Miller, C. Lai, K. L. Carraway, 3rd and C. Sweeney (2006). "The mucin Muc4 potentiates neuregulin signaling by increasing the cell-surface populations of ErbB2 and ErbB3." J Biol Chem **281**(28): 19310-19319.
Furukawa, J., M. Tsuda, K. Okada, T. Kimura, J. Piao, S. Tanaka and Y. Shinohara (2015). "Comprehensive Glycomics of a Multistep Human Brain Tumor Model Reveals Specific Glycosylation Patterns Related to Malignancy." PLoS One **10**(7): e0128300.
Gajjar, K., L. D. Heppenstall, W. Pang, K. M. Ashton, J. Trevisan, Patel, II, V. Llabjani, H. F. Stringfellow, P. L. Martin-Hirsch, T. Dawson and F. L. Martin (2012). "Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis." Anal Methods **5**: 89-102.
Galli, R., M. Meinhardt, E. Koch, G. Schackert, G. Steiner, M. Kirsch and O. Uckermann (2019). "Rapid Label-Free Analysis of Brain Tumor Biopsies by Near Infrared Raman and Fluorescence Spectroscopy-A Study of 209 Patients." Frontiers in oncology **9**: 1165-1165.
Gialeli, C., A. D. Theocharis and N. K. Karamanos (2011). "Roles of matrix metalloproteinases in cancer progression and their pharmacological targeting." Febs j **278**(1): 16-27.
Giamougiannis, P., P. L. Martin-Hirsch and F. L. Martin (2021). "The evolving role of MUC16 (CA125) in the transformation of ovarian cells and the progression of neoplasia." Carcinogenesis **42**(3): 327-343.
Goyal, S. N., C. P. Prajapati, P. R. Gore, C. R. Patil, U. B. Mahajan, C. Sharma, S. P. Talla and S. K. Ojha (2017). "Therapeutic Potential and Pharmaceutical Development of Thymoquinone: A Multitargeted Molecule of Natural Origin." Frontiers in Pharmacology **8**.
Gravendeel, L. A. M. and P. J. French (2011). Molecular Subtypes of Gliomas. Tumors of the Central Nervous System, Volume 2: Gliomas: Glioblastoma (Part 2). M. A. Hayat. Dordrecht, Springer Netherlands**:** 25-29.
Grimm, S. A. and M. C. Chamberlain (2016). "Anaplastic astrocytoma." CNS Oncol **5**(3): 145-157.
Gritsenko, P. G., O. Ilina and P. Friedl (2012). "Interstitial guidance of cancer invasion." J Pathol **226**(2): 185-199.
Groves, M. D., V. K. Puduvalli, C. A. Conrad, M. R. Gilbert, W. K. Yung, K. Jaeckle, V. Liu, K. R. Hess, K. D. Aldape and V. A. Levin (2006). "Phase II trial of temozolomide plus marimastat for recurrent anaplastic gliomas: a relationship among efficacy, joint toxicity and anticonvulsant status." J Neurooncol **80**(1): 83-90.
Haar, C. P., P. Hebbar, G. C. t. Wallace, A. Das, W. A. Vandergrift, 3rd, J. A. Smith, P. Giglio, S. J. Patel, S. K. Ray and N. L. Banik (2012). "Drug resistance in glioblastoma: a mini review." Neurochem Res **37**(6): 1192-1200.
Hagemann, C., J. Anacker, R. I. Ernestus and G. H. Vince (2012). "A complete compilation of matrix metalloproteinase expression in human malignant gliomas." World J Clin Oncol **3**(5): 67-79.
Hainsworth, J. D., T. Ervin, E. Friedman, V. Priego, P. B. Murphy, B. L. Clark and R. E. Lamar (2010). "Concurrent radiotherapy and temozolomide followed by temozolomide and sorafenib in the first-line treatment of patients with glioblastoma multiforme." Cancer **116**(15): 3663-3669.
Hainsworth, J. D., K. C. Shih, G. C. Shepard, G. W. Tillinghast, B. T. Brinker and D. R. Spigel (2012). "Phase II study of concurrent radiation therapy, temozolomide, and bevacizumab followed by bevacizumab/everolimus as first-line treatment for patients with glioblastoma." Clin Adv Hematol Oncol **10**(4): 240-246.

Hamada, T., T. Wakamatsu, M. Miyahara, S. Nagata, M. Nomura, Y. Kamikawa, N. Yamada, S. K. Batra, S. Yonezawa and K. Sugihara (2012). "MUC4: a novel prognostic factor of oral squamous cell carcinoma." Int J Cancer **130**(8): 1768-1776.
Hamano, Y., M. Zeisberg, H. Sugimoto, J. C. Lively, Y. Maeshima, C. Yang, R. O. Hynes, Z. Werb, A. Sudhakar and R. Kalluri (2003). "Physiological levels of tumstatin, a fragment of collagen IV alpha3 chain, are generated by MMP-9 proteolysis and suppress angiogenesis via alphaV beta3 integrin." Cancer Cell **3**(6): 589-601.
Hammoud, M. K., H. K. Yosef, T. Lechtonen, K. Aljakouch, M. Schuler, W. Alsaidi, I. Daho, A. Maghnouj, S. Hahn, S. F. El-Mashtoly and K. Gerwert (2018). "Raman micro-spectroscopy monitors acquired resistance to targeted cancer therapy at the cellular level." Sci Rep **8**(1): 15278.
Hanson, R. L. and M. A. Hollingsworth (2016). "Functional Consequences of Differential O-glycosylation of MUC1, MUC4, and MUC16 (Downstream Effects on Signaling)." Biomolecules **6**(3).
Heissig, B., C. Nishida, Y. Tashiro, Y. Sato, M. Ishihara, M. Ohki, I. Gritli, J. Rosenkvist and K. Hattori (2010). "Role of neutrophil-derived matrix metalloproteinase-9 in tissue regeneration." Histol Histopathol **25**(6): 765-770.
Hovey, E. J., K. M. Field, M. A. Rosenthal, E. H. Barnes, L. Cher, A. K. Nowak, H. Wheeler, K. Sawkins, A. Livingstone, P. Phal, C. Goh and J. Simes (2017). "Continuing or ceasing bevacizumab beyond progression in recurrent glioblastoma: an exploratory randomized phase II trial." Neurooncol Pract **4**(3): 171-181.
Huang, H. (2018). "Matrix Metalloproteinase-9 (MMP-9) as a Cancer Biomarker and MMP-9 Biosensors: Recent Advances." Sensors (Basel) **18**(10).
Im, K., S. Mareninov, M. F. P. Diaz and W. H. Yong (2019). "An Introduction to Performing Immunofluorescence Staining." Methods Mol Biol **1897**: 299-311.
Imran, M., A. Rauf, I. A. Khan, M. Shahbaz, T. B. Qaisrani, S. Fatmawati, T. Abu-Izneid, A. Imran, K. U. Rahman and T. A. Gondal (2018). "Thymoquinone: A novel strategy to combat cancer: A review." Biomed Pharmacother **106**: 390-402.
Iturrioz-Rodríguez, N., D. De Pasquale, P. Fiaschi and G. Ciofani (2022). "Discrimination of glioma patient-derived cells from healthy astrocytes by exploiting Raman spectroscopy." Spectrochim Acta A Mol Biomol Spectrosc **269**: 120773.
Iwasaki, K., A. Araki, C. M. Krishna, R. Maruyama, T. Yamamoto and H. Noothalapati (2021). "Identification of Molecular Basis for Objective Discrimination of Breast Cancer Cells (MCF-7) from Normal Human Mammary Epithelial Cells by Raman Microspectroscopy and Multivariate Curve Resolution Analysis." Int J Mol Sci **22**(2).
Jahan, R., M. A. Macha, S. Rachagani, S. Das, L. M. Smith, S. Kaur and S. K. Batra (2018). "Axed MUC4 (MUC4/X) aggravates pancreatic malignant phenotype by activating integrin-beta1/FAK/ERK pathway." Biochim Biophys Acta Mol Basis Dis **1864**(8): 2538-2549.
Jeon, H. Y., J. K. Kim, S. W. Ham, S. Y. Oh, J. Kim, J. B. Park, J. Y. Lee, S. C. Kim and H. Kim (2016). "Irradiation induces glioblastoma cell senescence and senescence-associated secretory phenotype." Tumour Biol **37**(5): 5857-5867.
Jermyn, M., J. Desroches, J. Mercier, K. St-Arnaud, M. C. Guiot, F. Leblond and K. Petrecca (2016). "Raman spectroscopy detects distant invasive brain cancer cells centimeters beyond MRI capability in humans." Biomed Opt Express **7**(12): 5129-5137.
Jeuken, J., A. Sijben, C. Alenda, J. Rijntjes, M. Dekkers, S. Boots-Sprenger, R. McLendon and P. Wesseling (2009). "Robust detection of EGFR copy number changes and EGFR variant III: technical aspects and relevance for glioma diagnostics." Brain Pathol **19**(4): 661-671.
Jiguet-Jiglaire, C., S. Boissonneau, E. Denicolai, V. Hein, R. Lasseur, J. Garcia, S. Romain, R. Appay, T. Graillon, W. Mason, A. F. Carpentier, A. A. Brandes, L. Ouafik, W. Wick, A. Baaziz, J. P. Gigan, R. J. Argüello, D. Figarella-Branger, O. Chinot and E. Tabouret (2022). "Plasmatic MMP9 released from tumor-infiltrating neutrophils is predictive for bevacizumab efficacy in glioblastoma patients: an AVAglio ancillary study." Acta Neuropathol Commun **10**(1): 1.
Johansson, N., M. Ahonen and V. M. Kahari (2000). "Matrix metalloproteinases in tumor invasion." Cell Mol Life Sci **57**(1): 5-15.

John R. Ferraro, K. N. a. C. W. B. (2003). Introductory Raman Spectroscopy.
Jolliffe, I. and J. Cadima (2016). "Principal component analysis: A review and recent developments." Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences **374**: 20150202.
Jonckheere, N., N. Skrypek, F. Frenois and I. Van Seuningen (2013). "Membrane-bound mucin modular domains: from structure to function." Biochimie **95**(6): 1077-1086.
Jonckheere, N., N. Skrypek and I. Van Seuningen (2014). "Mucins and tumor resistance to chemotherapeutic drugs." Biochim Biophys Acta **1846**(1): 142-151.
Joshi, S. and D. Yu (2017). Chapter 8 - Immunofluorescence. Basic Science Methods for Clinical Researchers. M. Jalali, F. Y. L. Saldanha and M. Jalali. Boston, Academic Press**:** 135-150.
Kalli, M. and T. Stylianopoulos (2018). "Defining the Role of Solid Stress and Matrix Stiffness in Cancer Cell Proliferation and Metastasis." Front Oncol **8**: 55.
Kang, C. Z. Y. H. B. S. W. Z. S. L. J. L. H. L. G. Z. X. (2020). "Label-free serum detection based on Raman spectroscopy for the diagnosis and classification of glioma." Journal of Raman Spectroscopy **51**(10): 1977-1985.
Karpel-Massler, G., U. Schmidt, A. Unterberg and M. E. Halatsch (2009). "Therapeutic inhibition of the epidermal growth factor receptor in high-grade gliomas: where do we stand?" Mol Cancer Res **7**(7): 1000-1012.
Kast, R. E., G. W. Auner, M. L. Rosenblum, T. Mikkelsen, S. M. Yurgelevic, A. Raghunathan, L. M. Poisson and S. N. Kalkanis (2014). "Raman molecular imaging of brain frozen tissue sections." J Neurooncol **120**(1): 55-62.
Kerr, L. T., H. J. Byrne and B. M. Hennelly (2015). "Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen." Analytical Methods **7**(12): 5041-5052.
Kim, H., S. Zheng, S. S. Amini, S. M. Virk, T. Mikkelsen, D. J. Brat, J. Grimsby, C. Sougnez, F. Muller, J. Hu, A. E. Sloan, M. L. Cohen, E. G. Van Meir, L. Scarpace, P. W. Laird, J. N. Weinstein, E. S. Lander, S. Gabriel, G. Getz, M. Meyerson, L. Chin, J. S. Barnholtz-Sloan and R. G. Verhaak (2015). "Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution." Genome Res **25**(3): 316-327.
Kirwan, A., M. Utratna, M. E. O'Dwyer, L. Joshi and M. Kilcoyne (2015). "Glycosylation-Based Serum Biomarkers for Cancer Diagnostics and Prognostics." BioMed research international **2015**: 490531-490531.
Köhler, M., S. Machill, R. Salzer and C. Krafft (2009). "Characterization of lipid extracts from brain tissue and tumors using Raman spectroscopy and mass spectrometry." Anal Bioanal Chem **393**(5): 1513-1520.
Kolli-Bouhafs, K., A. Boukhari, A. Abusnina, E. Velot, J. P. Gies, C. Lugnier and P. Rondé (2012). "Thymoquinone reduces migration and invasion of human glioblastoma cells associated with FAK, MMP-2 and MMP-9 down-regulation." Invest New Drugs **30**(6): 2121-2131.
Komatsu, M., C. A. Carraway, N. L. Fregien and K. L. Carraway (1997). "Reversible disruption of cell-matrix and cell-cell interactions by overexpression of sialomucin complex." J Biol Chem **272**(52): 33245-33254.
Komatsu, M., S. Jepson, M. E. Arango, C. A. Carothers Carraway and K. L. Carraway (2001). "Muc4/sialomucin complex, an intramembrane modulator of ErbB2/HER2/Neu, potentiates primary tumor growth and suppresses apoptosis in a xenotransplanted tumor." Oncogene **20**(4): 461-470.
Kopec, M., M. Błaszczyk, M. Radek and H. Abramczyk (2021). "Raman imaging and statistical methods for analysis various type of human brain tumors and breast cancers." Spectrochim Acta A Mol Biomol Spectrosc **262**: 120091.
Krafft, C., B. Dietzek, M. Schmitt and J. Popp (2012). "Raman and coherent anti-Stokes Raman scattering microspectroscopy for biomedical applications." Journal of Biomedical Optics **17**(4).
Krafft, C., S. B. Sobottka, G. Schackert and R. Salzer (2005). "Near infrared Raman spectroscopic mapping of native brain tissue and intracranial tumors." Analyst **130**(7): 1070-1077.

Kudelka, M. R., T. Ju, J. Heimburg-Molinaro and R. D. Cummings (2015). "Simple sugars to complex disease--mucin-type O-glycans in cancer." Advances in cancer research **126**: 53-135.
Kufe, D. W. (2009). "Mucins in cancer: function, prognosis and therapy." Nat Rev Cancer **9**(12): 874-885.
Larjavaara, S., R. Mantyla, T. Salminen, H. Haapasalo, J. Raitanen, J. Jaaskelainen and A. Auvinen (2007). "Incidence of gliomas by anatomic location." Neuro Oncol **9**(3): 319-325.
Larkin, P. (2011). "Chapter 1 - Introduction: Infrared and Raman Spectroscopy. In Infrared and Raman Spectroscopy." Ed. Elsevier: Oxford,.
Lee, H. K., M. S. Cho and T. H. Kim (2012). "Prognostic significance of muc4 expression in gallbladder carcinoma." World J Surg Oncol **10**: 224.
Lewis, A. T., R. Gaifulina, M. Isabelle, J. Dorney, M. L. Woods, G. R. Lloyd, K. Lau, M. Rodriguez-Justo, C. Kendall, N. Stone and G. M. Thomas (2017). "Mirrored stainless steel substrate provides improved signal for Raman spectroscopy of tissue and cells." J Raman Spectrosc **48**(1): 119-125.
Li, P., H. Wang, M. Hou, D. Li and H. Bai (2017). "Upstream stimulating factor1 (USF1) enhances the proliferation of glioblastoma stem cells mainly by activating the transcription of mucin13 (MUC13)." Pharmazie **72**(2): 98-102.
Li, Q., B. Chen, J. Cai, Y. Sun, G. Wang, Y. Li, R. Li, Y. Feng, B. Han, J. Li, Y. Tian, L. Yi and C. Jiang (2016). "Comparative Analysis of Matrix Metalloproteinase Family Members Reveals That MMP9 Predicts Survival and Response to Temozolomide in Patients with Primary Glioblastoma." PLoS One **11**(3): e0151815.
Li, W., C. Wu, Y. Yao, B. Dong, Z. Wei, X. Lv, J. Zhang and Y. Xu (2014). "MUC4 modulates human glioblastoma cell proliferation and invasion by upregulating EGFR expression." Neurosci Lett **566**: 82-87.
Lin, D., J. Pan, H. Huang, G. Chen, S. Qiu, H. Shi, W. Chen, Y. Yu, S. Feng and R. Chen (2014). "Label-free blood plasma test based on surface-enhanced Raman scattering for tumor stages detection in nasopharyngeal cancer." Sci Rep **4**: 4751.
Liu, M.-F., Y.-Y. Hu, T. Jin, K. Xu, S.-H. Wang, G.-Z. Du, B.-L. Wu, L.-Y. Li, L.-Y. Xu, E.-M. Li and H.-X. Xu (2015). "Matrix Metalloproteinase-9/Neutrophil Gelatinase-Associated Lipocalin Complex Activity in Human Glioma Samples Predicts Tumor Presence and Clinical Prognosis." Disease markers **2015**: 138974-138974.
Liu, Z., C. Lu, H. Hu, Z. Cai, Q. Liang, W. Sun, L. Jiang and G. Hu (2019). "LINC00909 promotes tumor progression in human glioma through regulation of miR-194/MUC1-C axis." Biomed Pharmacother **116**: 108965.
Livak, K. J. and T. D. Schmittgen (2001). "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method." Methods **25**(4): 402-408.
Louis, D. N., A. Perry, G. Reifenberger, A. von Deimling, D. Figarella-Branger, W. K. Cavenee, H. Ohgaki, O. D. Wiestler, P. Kleihues and D. W. Ellison (2016). "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary." Acta Neuropathol **131**(6): 803-820.
Louis, D. N., A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. K. Ng, S. M. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling and D. W. Ellison (2021). "The 2021 WHO Classification of Tumors of the Central Nervous System: a summary." Neuro Oncol **23**(8): 1231-1251.
Louis, D. N., A. Perry, P. Wesseling, D. J. Brat, I. A. Cree, D. Figarella-Branger, C. Hawkins, H. K. Ng, S.
Ma, H., X. X. Han and B. Zhao (2020). "Enhanced Raman spectroscopic analysis of protein post-translational modifications." TrAC Trends in Analytical Chemistry **131**: 116019.
Macha, M. A., S. R. Krishn, R. Jahan, K. Banerjee, S. K. Batra and M. Jain (2015). "Emerging potential of natural products for targeting mucins for therapy against inflammation and cancer." Cancer Treat Rev **41**(3): 277-288.
Mahesparan, R., T. A. Read, M. Lund-Johansen, K. O. Skaftnesmo, R. Bjerkvig and O. Engebraaten (2003). "Expression of extracellular matrix components in a highly infiltrative in vivo glioma model." Acta Neuropathol **105**(1): 49-57.

Majumder, A., S. Ray and A. Banerji (2019). "Epidermal growth factor receptor-mediated regulation of matrix metalloproteinase-2 and matrix metalloproteinase-9 in MCF-7 breast cancer cells." Mol Cell Biochem **452**(1-2): 111-121.
Mamede, A. P., I. P. Santos, A. L. M. Batista de Carvalho, P. Figueiredo, M. C. Silva, M. V. Tavares, M. P. M. Marques and L. A. E. Batista de Carvalho (2021). "A New Look into Cancer-A Review on the Contribution of Vibrational Spectroscopy on Early Diagnosis and Surgery Guidance." Cancers (Basel) **13**(21).
Mammoto, T., A. Jiang, E. Jiang, D. Panigrahy, M. W. Kieran and A. Mammoto (2013). "Role of collagen matrix in tumor angiogenesis and glioblastoma multiforme progression." Am J Pathol **183**(4): 1293-1305.
Mandel, A., P. Larsson, M. Sarwar, J. Semenas, A. S. Syed Khaja and J. L. Persson (2018). "The interplay between AR, EGF receptor and MMP-9 signaling pathways in invasive prostate cancer." Molecular Medicine **24**(1): 34.
Manne, A., A. Esnakula, L. Abushahin and A. Tsung (2021). "Understanding the Clinical Impact of MUC5AC Expression on Pancreatic Ductal Adenocarcinoma." Cancers (Basel) **13**(12).
Marcus, H. J., K. L. Carpenter, S. J. Price and P. J. Hutchinson (2010). "In vivo assessment of high-grade glioma biochemistry using microdialysis: a study of energy-related molecules, growth factors and cytokines." J Neurooncol **97**(1): 11-23.
Marro, M., C. Nieva, A. de Juan and A. Sierra (2018). "Unravelling the Metabolic Progression of Breast Cancer Cells to Bone Metastasis by Coupling Raman Spectroscopy and a Novel Use of Mcr-Als Algorithm." Anal Chem **90**(9): 5594-5602.
McNamara, M. G., S. Sahebjam and W. P. Mason (2013). "Emerging biomarkers in glioblastoma." Cancers (Basel) **5**(3): 1103-1119.
Medeiros, N. I., R. C. Fares, E. P. Franco, G. R. Sousa, R. T. Mattos, A. T. Chaves, M. D. Nunes, W. O. Dutra, R. Correa-Oliveira, M. O. Rocha and J. A. Gomes (2017). "Differential Expression of Matrix Metalloproteinases 2, 9 and Cytokines by Neutrophils and Monocytes in the Clinical Forms of Chagas Disease." PLoS Negl Trop Dis **11**(1): e0005284.
Mehta, K., A. Atak, A. Sahu, S. Srivastava and M. K. Chilakapati (2018). "An early investigative serum Raman spectroscopy study of meningioma." The Analyst **143**.
Melhem, J. M., J. Detsky, M. J. Lim-Fat and J. R. Perry (2022). "Updates in IDH-Wildtype Glioblastoma." Neurotherapeutics.
Menashi, S., M. Serova, L. Ma, S. Vignot, S. Mourah and F. Calvo (2003). "Regulation of Extracellular Matrix Metalloproteinase Inducer and Matrix Metalloproteinase Expression by Amphiregulin in Transformed Human Breast Epithelial Cells." Cancer Research **63**(22): 7575-7580.
Mercapide, J., R. Lopez De Cicco, D. E. Bassi, J. S. Castresana, G. Thomas and A. J. P. Klein-Szanto (2002). "Inhibition of furin-mediated processing results in suppression of astrocytoma cell growth and invasiveness." Clin Cancer Res **8**(6): 1740-1746.
Mercogliano, M. F., M. De Martino, L. Venturutti, M. A. Rivas, C. J. Proietti, G. Inurrigarro, I. Frahm, D. H. Allemand, E. G. Deza, S. Ares, F. G. Gercovich, P. Guzman, J. C. Roa, P. V. Elizalde and R. Schillaci (2017). "TNFalpha-Induced Mucin 4 Expression Elicits Trastuzumab Resistance in HER2-Positive Breast Cancer." Clin Cancer Res **23**(3): 636-648.
Mirabdaly, S., D. Elieh Ali Komi, Y. Shakiba, A. Moini and A. Kiani (2020). "Effects of temozolomide on U87MG glioblastoma cell expression of CXCR4, MMP2, MMP9, VEGF, anti-proliferatory cytotoxic and apoptotic properties." Mol Biol Rep **47**(2): 1187-1197.
Momeny, M., J. M. Saunus, F. Marturana, A. E. McCart Reed, D. Black, G. Sala, S. Iacobelli, J. D. Holland, D. Yu, L. Da Silva, P. T. Simpson, K. K. Khanna, G. Chenevix-Trench and S. R. Lakhani (2015). "Heregulin-HER3-HER2 signaling promotes matrix metalloproteinase-dependent blood-brain-barrier transendothelial migration of human breast cancer cell lines." Oncotarget **6**(6): 3932-3946.
Mukherjee, D. and A. Quinones-Hinojosa (2011). Impact of Extent of Resection on Outcomes in Patients with High-Grade Gliomas. Tumors of the Central Nervous System, Volume 2: Gliomas: Glioblastoma (Part 2). M. A. Hayat. Dordrecht, Springer Netherlands**:** 173-179.

Mukhopadhyay, P., I. Lakshmanan, M. P. Ponnusamy, S. Chakraborty, M. Jain, P. Pai, L. M. Smith, S. M. Lele and S. K. Batra (2013). "MUC4 overexpression augments cell migration and metastasis through EGFR family proteins in triple negative breast cancer cells." PLoS One **8**(2): e54455.
Munaut, C., A. Noel, O. Hougrand, J. M. Foidart, J. Boniver and M. Deprez (2003). "Vascular endothelial growth factor expression correlates with matrix metalloproteinases MT1-MMP, MMP-2 and MMP-9 in human glioblastomas." Int J Cancer **106**(6): 848-855.
Munkley, J. (2019). "The glycosylation landscape of pancreatic cancer." Oncol Lett **17**(3): 2569-2575.
Nakano, I. (2014). "Proneural-mesenchymal transformation of glioma stem cells: do therapies cause evolution of target in glioblastoma?" Future Oncol **10**(9): 1527-1530.
Nargis, H., H. Nawaz, H. Bhatti, K. Jilani and M. Saleem (2021). "Comparison of surface enhanced Raman spectroscopy and Raman spectroscopy for the detection of breast cancer based on serum samples." Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **246**: 119034.
Nath, S. and P. Mukherjee (2014). "MUC1: a multifaceted oncoprotein with a key role in cancer progression." Trends Mol Med **20**(6): 332-342.
Nishikawa, R., X. D. Ji, R. C. Harmon, C. S. Lazar, G. N. Gill, W. K. Cavenee and H. J. Huang (1994). "A mutant epidermal growth factor receptor common in human glioma confers enhanced tumorigenicity." Proc Natl Acad Sci U S A **91**(16): 7727-7731.
Nozawa, H., C. Chiu and D. Hanahan (2006). "Infiltrating neutrophils mediate the initial angiogenic switch in a mouse model of multistage carcinogenesis." Proc Natl Acad Sci U S A **103**(33): 12493-12498.
Ohgaki, H., P. Dessen, B. Jourde, S. Horstmann, T. Nishikawa, P. L. Di Patre, C. Burkhard, D. Schuler, N. M. Probst-Hensch, P. C. Maiorka, N. Baeza, P. Pisani, Y. Yonekawa, M. G. Yasargil, U. M. Lutolf and P. Kleihues (2004). "Genetic pathways to glioblastoma: a population-based study." Cancer Res **64**(19): 6892-6899.
Ohgaki, H. and P. Kleihues (2013). "The definition of primary and secondary glioblastoma." Clin Cancer Res **19**(4): 764-772.
Pallud, J. and E. Mandonnet (2011). Quantitative Approach of the Natural Course of Diffuse Low-Grade Gliomas. Tumors of the Central Nervous System, Volume 2: Gliomas: Glioblastoma (Part 2). M. A. Hayat. Dordrecht, Springer Netherlands**:** 163-172.
Pang, Y., Y. Zhang, H. Y. Zhang, W. H. Wang, G. Jin, J. W. Liu and Z. J. Zhu (2022). "MUC13 promotes lung cancer development and progression by activating ERK signaling." Oncol Lett **23**(1): 37.
Park, M. J., M. S. Kim, I. C. Park, H. S. Kang, H. Yoo, S. H. Park, C. H. Rhee, S. I. Hong and S. H. Lee (2002). "PTEN suppresses hyaluronic acid-induced matrix metalloproteinase-9 expression in U87MG glioblastoma cells through focal adhesion kinase dephosphorylation." Cancer Res **62**(21): 6318-6322.
Parlatan, U., M. T. Inanc, B. Y. Ozgor, E. Oral, E. Bastu, M. B. Unlu and G. Basar (2019). "Raman spectroscopy as a non-invasive diagnostic technique for endometriosis." Scientific Reports **9**(1): 19795.
Parsons, D. W., S. Jones, X. Zhang, J. C.-H. Lin, R. J. Leary, P. Angenendt, P. Mankoo, H. Carter, I. M. Siu, G. L. Gallia, A. Olivi, R. McLendon, B. A. Rasheed, S. Keir, T. Nikolskaya, Y. Nikolsky, D. A. Busam, H. Tekleab, L. A. Diaz, Jr., J. Hartigan, D. R. Smith, R. L. Strausberg, S. K. N. Marie, S. M. O. Shinjo, H. Yan, G. J. Riggins, D. D. Bigner, R. Karchin, N. Papadopoulos, G. Parmigiani, B. Vogelstein, V. E. Velculescu and K. W. Kinzler (2008). "An integrated genomic analysis of human glioblastoma multiforme." Science (New York, N.Y.) **321**(5897): 1807-1812.
Pasquier, B., D. Pasquier, A. N'Golet, M. H. Panh and P. Couderc (1980). "Extraneural metastases of astrocytomas and glioblastomas: clinicopathological study of two cases and review of literature." Cancer **45**(1): 112-125.
Paw, I., R. C. Carpenter, K. Watabe, W. Debinski and H. W. Lo (2015). "Mechanisms regulating glioma invasion." Cancer Lett **362**(1): 1-7.
Payne, L. S. and P. H. Huang (2013). "The pathobiology of collagens in glioma." Molecular cancer research : MCR **11**(10): 1129-1140.
Peddinti, R., R. Zeine, D. Luca, R. Seshadri, A. Chlenski, K. Cole, B. Pawel, H. R. Salwen, J. M. Maris and S. L. Cohn (2007). "Prominent Microvascular Proliferation in Clinically Aggressive Neuroblastoma." Clinical Cancer Research **13**(12): 3499-3506.

Peixoto, A., M. Relvas-Santos, R. Azevedo, L. L. Santos and J. A. Ferreira (2019). "Protein Glycosylation and Tumor Microenvironment Alterations Driving Cancer Hallmarks." Frontiers in Oncology **9**(380).
Phillips, H. S., S. Kharbanda, R. Chen, W. F. Forrest, R. H. Soriano, T. D. Wu, A. Misra, J. M. Nigro, H. Colman, L. Soroceanu, P. M. Williams, Z. Modrusan, B. G. Feuerstein and K. Aldape (2006). "Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis." Cancer Cell **9**(3): 157-173.
Pino, V., V. P. Ramsauer, P. Salas, C. A. Carothers Carraway and K. L. Carraway (2006). "Membrane mucin Muc4 induces density-dependent changes in ERK activation in mammary epithelial and tumor cells: role in reversal of contact inhibition." J Biol Chem **281**(39): 29411-29420.
Piyush, T., A. R. Chacko, P. Sindrewicz, J. Hilkens, J. M. Rhodes and L. G. Yu (2017). "Interaction of galectin-3 with MUC1 on cell surface promotes EGFR dimerization and activation in human epithelial cancer cells." Cell Death Differ **24**(11): 1937-1947.
Pochampalli, M. R., R. M. el Bejjani and J. A. Schroeder (2007). "MUC1 is a novel regulator of ErbB1 receptor trafficking." Oncogene **26**(12): 1693-1701.
Pointer, K. B., P. A. Clark, A. B. Schroeder, M. S. Salamat, K. W. Eliceiri and J. S. Kuo (2017). "Association of collagen architecture with glioblastoma patient survival." Journal of neurosurgery **126**(6): 1812-1821.
Ponnusamy, M. P., I. Lakshmanan, M. Jain, S. Das, S. Chakraborty, P. Dey and S. K. Batra (2010). "MUC4 mucin-induced epithelial to mesenchymal transition: a novel mechanism for metastasis of human ovarian cancer cells." Oncogene **29**(42): 5741-5754.
Ponnusamy, M. P., A. P. Singh, M. Jain, S. Chakraborty, N. Moniaux and S. K. Batra (2008). "MUC4 activates HER2 signalling and enhances the motility of human ovarian cancer cells." Br J Cancer **99**(3): 520-526.
Price, A., Q. Shi, D. Morris, M. E. Wilcox, P. M. Brasher, N. B. Rewcastle, D. Shalinsky, H. Zou, K. Appelt, R. N. Johnston, V. W. Yong, D. Edwards and P. Forsyth (1999). "Marked inhibition of tumor growth in a malignant glioma tumor model by a novel synthetic matrix metalloproteinase inhibitor AG3340." Clin Cancer Res **5**(4): 845-854.
Pullen NA, P. A., Perry MM, Jaworski DM, Loveson KF, Arthur DJ (2018). "Current insights into matrix metalloproteinases and glioma progression: transcending the degradation boundary." Metalloproteinases In Medicine **5:13-30**.
Quesnel, A., A. Broughton, G. S. Karagiannis and P. S. Filippou (2022). "Message in the bottle: regulation of the tumor microenvironment via exosome-driven proteolysis." Cancer Metastasis Rev.
Quesnel, A., G. S. Karagiannis and P. S. Filippou (2020). "Extracellular proteolysis in glioblastoma progression and therapeutics." Biochim Biophys Acta Rev Cancer **1874**(2): 188428.
Rachagani, S., M. A. Macha, M. P. Ponnusamy, D. Haridas, S. Kaur, M. Jain and S. K. Batra (2012). "MUC4 potentiates invasion and metastasis of pancreatic cancer cells through stabilization of fibroblast growth factor receptor 1." Carcinogenesis **33**(10): 1953-1964.
Radhakrishnan, P., S. Dabelsteen, F. B. Madsen, C. Francavilla, K. L. Kopp, C. Steentoft, S. Y. Vakhrushev, J. V. Olsen, L. Hansen, E. P. Bennett, A. Woetmann, G. Yin, L. Chen, H. Song, M. Bak, R. A. Hlady, S. L. Peters, R. Opavsky, C. Thode, K. Qvortrup, K. T.-B. G. Schjoldager, H. Clausen, M. A. Hollingsworth and H. H. Wandall (2014). "Immature truncated O-glycophenotype of cancer directly induces oncogenic features." Proceedings of the National Academy of Sciences **111**(39): E4066-E4075.
Rak, J., C. Milsom, L. May, P. Klement and J. Yu (2006). "Tissue factor in cancer and angiogenesis: the molecular link between genetic tumor progression, tumor neovascularization, and cancer coagulopathy." Semin Thromb Hemost **32**(1): 54-70.
Rao, J. S., M. Yamamoto, S. Mohaman, Z. L. Gokaslan, G. N. Fuller, W. G. Stetler-Stevenson, V. H. Rao, L. A. Liotta, G. L. Nicolson and R. E. Sawaya (1996). "Expression and localization of 92 kDa type IV collagenase/gelatinase B (MMP-9) in human gliomas." Clin Exp Metastasis **14**(1): 12-18.
Rathore, S., H. Akbari, M. Rozycki, K. G. Abdullah, M. P. Nasrallah, Z. A. Binder, R. V. Davuluri, R. A. Lustig, N. Dahmane, M. Bilello, D. M. O'Rourke and C. Davatzikos (2018). "Radiomic MRI signature reveals three distinct subtypes of glioblastoma with different clinical and molecular characteristics, offering prognostic value beyond IDH1." Sci Rep **8**(1): 5087.

Reily, C., T. J. Stewart, M. B. Renfrow and J. Novak (2019). "Glycosylation in health and disease." Nat Rev Nephrol **15**(6): 346-366.
Rempe, R. G., A. M. S. Hartz and B. Bauer (2016). "Matrix metalloproteinases in the brain and blood-brain barrier: Versatile breakers and makers." J Cereb Blood Flow Metab **36**(9): 1481-1507.
Ricci, M., F. Ragonese, B. Gironi, M. Paolantoni, A. Morresi, L. Latterini, B. Fioretti and P. Sassi (2018). "Glioblastoma single-cell microRaman analysis under stress treatments." Scientific Reports **8**.
Ricci, S., E. Guadagno, D. Bruzzese, M. Del Basso De Caro, C. Peca, F. G. Sgulò, F. Maiuri and A. Di Carlo (2017). "Evaluation of matrix metalloproteinase type IV-collagenases in serum of patients with tumors of the central nervous system." J Neurooncol **131**(2): 223-232.
Riemenschneider, M. J., J. W. Jeuken, P. Wesseling and G. Reifenberger (2010). "Molecular diagnostics of gliomas: state of the art." Acta Neuropathol **120**(5): 567-584.
Riva, M., T. Sciortino, R. Secoli, E. D'Amico, S. Moccia, B. Fernandes, M. Conti Nibali, L. Gay, M. Rossi, E. De Momi and L. Bello (2021). "Glioma biopsies Classification Using Raman Spectroscopy and Machine Learning Models on Fresh Tissue Samples." Cancers (Basel) **13**(5).
Rodriguez, F. J., B. A. Orr, K. L. Ligon and C. G. Eberhart (2012). "Neoplastic cells are a rare component in human glioblastoma microvasculature." Oncotarget **3**(1): 98-106.
Rodríguez-Camacho, A., J. G. Flores-Vázquez, J. Moscardini-Martelli, J. A. Torres-Ríos, A. Olmos-Guzmán, C. S. Ortiz-Arce, D. R. Cid-Sánchez, S. R. Pérez, M. D. S. Macías-González, L. C. Hernández-Sánchez, J. C. Heredia-Gutiérrez, G. A. Contreras-Palafox, J. J. E. Suárez-Campos, M. Celis-López, G. A. Gutiérrez-Aceves and S. Moreno-Jiménez (2022). "Glioblastoma Treatment: State-of-the-Art and Future Perspectives." Int J Mol Sci **23**(13).
Roman, M., T. P. Wrobel, A. Panek, C. Paluszkiewicz and W. M. Kwiatek (2020). "Lipid droplets in prostate cancer cells and effect of irradiation studied by Raman microspectroscopy." Biochim Biophys Acta Mol Cell Biol Lipids **1865**(9): 158753.
Rong, Y., D. L. Durden, E. G. Van Meir and D. J. Brat (2006). "'Pseudopalisading' necrosis in glioblastoma: a familiar morphologic feature that links vascular pathology, hypoxia, and angiogenesis." J Neuropathol Exp Neurol **65**(6): 529-539.
Rowson-Hodel, A. R., J. H. Wald, J. Hatakeyama, W. K. O'Neal, J. R. Stonebraker, K. VanderVorst, M. J. Saldana, A. D. Borowsky, C. Sweeney and K. L. Carraway, 3rd (2018). "Membrane Mucin Muc4 promotes blood cell association with tumor cells and mediates efficient metastasis in a mouse model of breast cancer." Oncogene **37**(2): 197-207.
Rybakowski, J. K. (2009). "Matrix Metalloproteinase-9 (MMP9)-A Mediating Enzyme in Cardiovascular Disease, Cancer, and Neuropsychiatric Disorders." Cardiovasc Psychiatry Neurol **2009**: 904836.
Ryzhikova, E., O. Kazakov, L. Halamkova, D. Celmins, P. Malone, E. Molho, E. A. Zimmerman and I. K. Lednev (2015). "Raman spectroscopy of blood serum for Alzheimer's disease diagnostics: specificity relative to other types of dementia." J Biophotonics **8**(7): 584-596.
Saadeh, F. S., R. Mahfouz and H. I. Assi (2018). "EGFR as a clinical marker in glioblastomas and other gliomas." Int J Biol Markers **33**(1): 22-32.
Sabari, J., D. Lax, D. Connors, I. Brotman, E. Mindrebo, C. Butler, I. Entersz, D. Jia and R. A. Foty (2011). "Fibronectin matrix assembly suppresses dispersal of glioblastoma cells." PLoS One **6**(9): e24810.
Sahin, U., G. Weskamp, K. Kelly, H.-M. Zhou, S. Higashiyama, J. Peschon, D. Hartmann, P. Saftig and C. P. Blobel (2004). "Distinct roles for ADAM10 and ADAM17 in ectodomain shedding of six EGFR ligands." The Journal of cell biology **164**(5): 769-779.
Sartori, E., R. Langer, E. Vassella, E. Hewer, P. Schucht, I. Zlobec and S. Berezowska (2017). "Low co-expression of epidermal growth factor receptor and its chaperone heat shock protein 90 is associated with worse prognosis in primary glioblastoma, IDH-wild-type." Oncol Rep **38**(4): 2394-2400.
Sawaya, R., M. Yamamoto, O. J. Ramo, M. L. Shi, A. Rayford and J. S. Rao (1995). "Plasminogen activator inhibitor-1 in brain tumors: relation to malignancy and necrosis." Neurosurgery **36**(2): 375-380; discussion 380-371.
Seifert, M., G. Schackert, A. Temme, E. Schröck, A. Deutsch and B. Klink (2020). "Molecular Characterization of Astrocytoma Progression Towards Secondary Glioblastomas Utilizing Patient-Matched Tumor Pairs." Cancers (Basel) **12**(6).

Seker-Polat, F., N. Pinarbasi Degirmenci, I. Solaroglu and T. Bagci-Onder (2022). "Tumor Cell Infiltration into the Brain in Glioblastoma: From Mechanisms to Clinical Perspectives." Cancers (Basel) **14**(2).
Senapati, S., P. Chaturvedi, W. G. Chaney, S. Chakraborty, V. S. Gnanapragassam, A. R. Sasson and S. K. Batra (2011). "Novel INTeraction of MUC4 and galectin: potential pathobiological implications for metastasis in lethal pancreatic cancer." Clin Cancer Res **17**(2): 267-274.
Senapati, S., V. S. Gnanapragassam, N. Moniaux, N. Momi and S. K. Batra (2012). "Role of MUC4-NIDO domain in the MUC4-mediated metastasis of pancreatic cancer cells." Oncogene **31**(28): 3346-3356.
Seshacharyulu, P., M. P. Ponnusamy, S. Rachagani, I. Lakshmanan, D. Haridas, Y. Yan, A. K. Ganti and S. K. Batra (2015). "Targeting EGF-receptor(s) - STAT1 axis attenuates tumor growth and metastasis through downregulation of MUC4 mucin in human pancreatic cancer." Oncotarget **6**(7): 5164-5181.
Shergalis, A., A. Bankhead, 3rd, U. Luesakul, N. Muangsin and N. Neamati (2018). "Current Challenges and Opportunities in Treating Glioblastoma." Pharmacol Rev **70**(3): 412-445.
Shrivastava, A., L. Aggarwal, M. K. Chilakapati, S. Pradhan, S. Mishra, S. Choudhary, C. Patel, S. Singla, A. Chaurasiya and R. Singh (2020). "Diagnostic and Prognostic Application of Raman Spectroscopy in Carcinoma Cervix: A Biomolecular Approach." Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy **250**: 119356.
Silsirivanit, A. (2019). "Glycosylation markers in cancer." Adv Clin Chem **89**: 189-213.
Simon, T., E. Jackson and G. Giamas (2020). "Breaking through the glioblastoma micro-environment via extracellular vesicles." Oncogene.
Singh, R. and D. Bandyopadhyay (2007). "MUC1: a target molecule for cancer therapy." Cancer Biol Ther **6**(4): 481-486.
Skog, J., T. Wurdinger, S. van Rijn, D. Meijer, L. Gainche, M. Sena-Esteves, W. T. Curry, R. S. Carter, A. M. Krichevsky and X. O. Breakefield (2008). "Glioblastoma microvesicles transport RNA and protein that promote tumor growth and provide diagnostic biomarkers." (1465-7392 (Print)).
Soda, Y., T. Marumoto, D. Friedmann-Morvinski, M. Soda, F. Liu, H. Michiue, S. Pastorino, M. Yang, R. M. Hoffman, S. Kesari and I. M. Verma (2011). "Transdifferentiation of glioblastoma cells into vascular endothelial cells." Proc Natl Acad Sci U S A **108**(11): 4274-4280.
Șovrea, A. S., B. Boșca, C. S. Melincovici, A. M. Constantin, A. Crintea, M. Mărginean, E. Dronca, M. E. Jianu, R. Suflețel, D. Gonciar, M. Bungărdean and C. B. Crivii (2022). "Multiple Faces of the Glioblastoma Microenvironment." Int J Mol Sci **23**(2).
Srinivasan, G. (2010). Vibrational spectroscopic imaging for biomedical applications.
Steed, T. C., J. M. Treiber, K. Patel, V. Ramakrishnan, A. Merk, A. R. Smith, B. S. Carter, A. M. Dale, L. M. L. Chow and C. C. Chen (2016). "Differential localization of glioblastoma subtype: implications on glioblastoma pathogenesis." Oncotarget **7**(18): 24899-24907.
Stoup, N., M. Liberelle, C. Schulz, S. Cavdarli, R. Vasseur, R. Magnez, F. Lahdaoui, N. Skrypek, F. Peretti, F. Frénois, X. Thuru, P. Melnyk, N. Renault, N. Jonckheere, N. Lebègue and I. Van Seuningen (2021). "The EGF Domains of MUC4 Oncomucin Mediate HER2 Binding Affinity and Promote Pancreatic Cancer Cell Tumorigenesis." Cancers (Basel) **13**(22).
Stupp, R., M. E. Hegi, B. Neyns, R. Goldbrunner, U. Schlegel, P. M. Clement, G. G. Grabenbauer, A. F. Ochsenbein, M. Simon, P. Y. Dietrich, T. Pietsch, C. Hicking, J. C. Tonn, A. C. Diserens, A. Pica, M. Hermisson, S. Krueger, M. Picard and M. Weller (2010). "Phase I/IIa study of cilengitide and temozolomide with concomitant radiotherapy followed by cilengitide and temozolomide maintenance therapy in patients with newly diagnosed glioblastoma." J Clin Oncol **28**(16): 2712-2718.
Stupp, R., W. P. Mason, M. J. van den Bent, M. Weller, B. Fisher, M. J. Taphoorn, K. Belanger, A. A. Brandes, C. Marosi, U. Bogdahn, J. Curschmann, R. C. Janzer, S. K. Ludwin, T. Gorlia, A. Allgeier, D. Lacombe, J. G. Cairncross, E. Eisenhauer and R. O. Mirimanoff (2005). "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma." N Engl J Med **352**(10): 987-996.
Subhash C Chauhan, A. P. S., Fernanda Ruiz, Sonny L Johansson, Maneesh Jain, and L. M. Smith (2006). "Aberrant expression of MUC4 in ovarian carcinoma: diagnostic significance alone and in combination with MUC1 and MUC16 (CA125)." Modern Pathology **19, 1386–1394**.

Sun, C., Q. Wang, H. Zhou, S. Yu, A. R. Simard, C. Kang, Y. Li, Y. Kong, T. An, Y. Wen, F. Shi and J. Hao (2013). "Antisense MMP-9 RNA inhibits malignant glioma cell growth in vitro and in vivo." Neurosci Bull **29**(1): 83-93.
Surmacki, J. M., I. Quiros-Gonzalez and S. E. Bohndiek (2022). "Evaluation of Label-Free Confocal Raman Microspectroscopy for Monitoring Oxidative Stress In Vitro in Live Human Cancer Cells." Antioxidants (Basel) **11**(3).
Taal, W., J. E. Bromberg and M. J. van den Bent (2015). "Chemotherapy in glioma." CNS Oncol **4**(3): 179-192.
Tabarés, G., C. M. Radcliffe, S. Barrabés, M. Ramírez, R. N. Aleixandre, W. Hoesel, R. A. Dwek, P. M. Rudd, R. Peracaula and R. de Llorens (2006). "Different glycan structures in prostate-specific antigen from prostate cancer sera in relation to seminal plasma PSA." Glycobiology **16**(2): 132-145.
Taib, B., A. Aboussalah, M. Moniruzzaman, S. Chen, N. Haughey, S. Kim and R. Ahima (2019). "Lipid accumulation and oxidation in glioblastoma multiforme." Scientific Reports **9**.
Tang, F. Y., E. P. Chiang and Y. C. Sun (2008). "Resveratrol inhibits heregulin-beta1-mediated matrix metalloproteinase-9 expression and cell invasion in human breast cancer cells." J Nutr Biochem **19**(5): 287-294.
Tang, J., Y. Zhu, K. Xie, X. Zhang, X. Zhi, W. Wang, Z. Li, Q. Zhang, L. Wang, J. Wang and Z. Xu (2016). "The role of the AMOP domain in MUC4/Y-promoted tumour angiogenesis and metastasis in pancreatic cancer." J Exp Clin Cancer Res **35**(1): 91.
Tang, Y., P. Kesavan, M. T. Nakada and L. Yan (2004). "Tumor-stroma interaction: positive feedback regulation of extracellular matrix metalloproteinase inducer (EMMPRIN) expression and matrix metalloproteinase-dependent generation of soluble EMMPRIN." Mol Cancer Res **2**(2): 73-80.
Tian, Z., C. Chen, Y. Fan, X. Ou, J. Wang, X. Ma and J. Xu (2019). "Glioblastoma and Anaplastic Astrocytoma: Differentiation Using MRI Texture Analysis." Frontiers in Oncology **9**.
Tondepu, C. and L. Karumbaiah (2022). "Glycomaterials to Investigate the Functional Role of Aberrant Glycosylation in Glioblastoma." Adv Healthc Mater **11**(4): e2101956.
Torres, M. P., S. Chakraborty, J. Souchek and S. K. Batra (2012). "Mucin-based targeted pancreatic cancer therapy." Curr Pharm Des **18**(17): 2472-2481.
Torres, M. P., M. P. Ponnusamy, S. Chakraborty, L. M. Smith, S. Das, H. A. Arafat and S. K. Batra (2010). "Effects of thymoquinone in the expression of mucin 4 in pancreatic cancer cells: implications for the development of novel cancer therapies." Mol Cancer Ther **9**(5): 1419-1431.
Tove, L. L., H. Andreas Hanssøn, S. Stein and H. Sverre (2012). "Prognostic value of histological features in diffuse astrocytomas WHO grade II." Int J Clin Exp Pathol **5**(2): 152-158.
Tsutsumida, H., M. Goto, S. Kitajima, I. Kubota, Y. Hirotsu, J. Wakimoto, S. K. Batra, K. Imai and S. Yonezawa (2007). "MUC4 expression correlates with poor prognosis in small-sized lung adenocarcinoma." Lung Cancer **55**(2): 195-203.
Turcan, S., D. Rohle, A. Goenka, L. A. Walsh, F. Fang, E. Yilmaz, C. Campos, A. W. Fabius, C. Lu, P. S. Ward, C. B. Thompson, A. Kaufman, O. Guryanova, R. Levine, A. Heguy, A. Viale, L. G. Morris, J. T. Huse, I. K. Mellinghoff and T. A. Chan (2012). "IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype." Nature **483**(7390): 479-483.
Turcotte, P., S. Alex and D. Vocelle (1992). "Studies of the role of water on some imine–acid complexes by means of Raman spectroscopy." Canadian Journal of Chemistry-revue Canadienne De Chimie - CAN J CHEM **70**: 2751-2757.
Ulasov, I., B. Thaci, P. Sarvaiya, R. Yi, D. Guo, B. Auffinger, P. Pytel, L. Zhang, C. K. Kim, A. Borovjagin, M. Dey, Y. Han, A. Y. Baryshnikov and M. S. Lesniak (2013). "Inhibition of MMP14 potentiates the therapeutic effect of temozolomide and radiation in gliomas." Cancer Med **2**(4): 457-467.
Uyar, R. (2022). "Glioblastoma microenvironment: The stromal interactions." Pathol Res Pract **232**: 153813.
van Bodegraven, E. J., J. V. van Asperen, P. A. J. Robe and E. M. Hol (2019). "Importance of GFAP isoform-specific analyses in astrocytoma." Glia **67**(8): 1417-1433.
van Putten, J. P. M. and K. Strijbis (2017). "Transmembrane Mucins: Signaling Receptors at the Intersection of Inflammation and Cancer." J Innate Immun **9**(3): 281-299.

Veillon, L., C. Fakih, H. Abou-El-Hassan, F. Kobeissy and Y. Mechref (2018). "Glycosylation Changes in Brain Cancer." ACS Chem Neurosci **9**(1): 51-72.
Verhaak, R. G., K. A. Hoadley, E. Purdom, V. Wang, Y. Qi, M. D. Wilkerson, C. R. Miller, L. Ding, T. Golub, J. P. Mesirov, G. Alexe, M. Lawrence, M. O'Kelly, P. Tamayo, B. A. Weir, S. Gabriel, W. Winckler, S. Gupta, L. Jakkula, H. S. Feiler, J. G. Hodgson, C. D. James, J. N. Sarkaria, C. Brennan, A. Kahn, P. T. Spellman, R. K. Wilson, T. P. Speed, J. W. Gray, M. Meyerson, G. Getz, C. M. Perou and D. N. Hayes (2010). "Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1." Cancer Cell **17**(1): 98-110.
Viapiano, M. S., W. L. Bi, J. Piepmeier, S. Hockfield and R. T. Matthews (2005). "Novel tumor-specific isoforms of BEHAB/brevican identified in human malignant gliomas." Cancer Res **65**(15): 6726-6733.
Visse, R. and H. Nagase (2003). "Matrix metalloproteinases and tissue inhibitors of metalloproteinases: structure, function, and biochemistry." Circ Res **92**(8): 827-839.
von Achenbach, C., M. Weller and E. Szabo (2018). "Epidermal growth factor receptor and ligand family expression and activity in glioblastoma." J Neurochem **147**(1): 99-109.
Wager, M., L. Karayan-Tapon and C.-J. Larsen (2011). Gliomagenesis: Advantages and Limitations of Biomarkers. Tumors of the Central Nervous System, Volume 2: Gliomas: Glioblastoma (Part 2). M. A. Hayat. Dordrecht, Springer Netherlands**:** 11-24.
Wang, X., D. Wang, Z. Yao, B. Xin, B. Wang, C. Lan, Y. Qin, S. Xu, D. He and Y. Liu (2019). "Machine Learning Models for Multiparametric Glioma Grading With Quantitative Result Interpretations." Frontiers in Neuroscience **12**(1046).
Waqar, M., D. M. Trifiletti, C. McBain, J. O'Connor, D. J. Coope, L. Akkari, A. Quinones-Hinojosa and G. R. Borst (2022). "Early Therapeutic Interventions for Newly Diagnosed Glioblastoma: Rationale and Review of the Literature." Curr Oncol Rep **24**(3): 311-324.
Watkins, S., S. Robel, I. F. Kimbrough, S. M. Robert, G. Ellis-Davies and H. Sontheimer (2014). "Disruption of astrocyte-vascular coupling and the blood-brain barrier by invading glioma cells." Nat Commun **5**: 4196.
Weller, M., M. van den Bent, M. Preusser, E. Le Rhun, J. C. Tonn, G. Minniti, M. Bendszus, C. Balana, O. Chinot, L. Dirven, P. French, M. E. Hegi, A. S. Jakola, M. Platten, P. Roth, R. Rudà, S. Short, M. Smits, M. J. B. Taphoorn, A. von Deimling, M. Westphal, R. Soffietti, G. Reifenberger and W. Wick (2021). "EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood." Nature Reviews Clinical Oncology **18**(3): 170-186.
Wesseling, P. (2018). "WHO 2016 Classification of gliomas." Neuropathol Appl Neurobiol **44**(2): 139-150.
Westphal, M., C. L. Maire and K. Lamszus (2017). "EGFR as a Target for Glioblastoma Treatment: An Unfulfilled Promise." CNS Drugs **31**(9): 723-735.
Wick, W., M. Weller, M. Weiler, T. Batchelor, A. W. Yung and M. Platten (2011). "Pathway inhibition: emerging molecular targets for treating glioblastoma." Neuro Oncol **13**(6): 566-579.
Wiencke, J. K., S. Zheng, N. Jelluma, T. Tihan, S. Vandenberg, T. Tamguney, R. Baumber, R. Parsons, K. R. Lamborn, M. S. Berger, M. R. Wrensch, D. A. Haas-Kogan and D. Stokoe (2007). "Methylation of the PTEN promoter defines low-grade gliomas and secondary glioblastoma." Neuro Oncol **9**(3): 271-279.
Wiercigroch, E., E. Szafraniec, K. Czamara, M. Z. Pacia, K. Majzner, K. Kochan, A. Kaczor, M. Baranska and K. Malek (2017). "Raman and infrared spectroscopy of carbohydrates: A review." Spectrochim Acta A Mol Biomol Spectrosc **185**: 317-335.
Wood, M. D., A. M. Halfpenny and S. R. Moore (2019). "Applications of molecular neuro-oncology - a review of diffuse glioma integrated diagnosis and emerging molecular entities." Diagn Pathol **14**(1): 29.
Workman, H. C., J. K. Miller, E. Q. Ingalla, R. P. Kaur, D. I. Yamamoto, L. A. Beckett, L. J. Young, R. D. Cardiff, A. D. Borowsky, K. L. Carraway, C. Sweeney and K. L. Carraway, 3rd (2009). "The membrane mucin MUC4 is elevated in breast tumor lymph node metastases relative to matched primary tumors and confers aggressive properties to breast cancer cells." Breast Cancer Res **11**(5): R70.

Xia, P., A. H. Choi, Z. Deng, Y. Yang, J. Zhao, Y. Wang, P. R. Hardwidge and G. Zhu (2017). "Cell membrane-anchored MUC4 promotes tumorigenicity in epithelial carcinomas." Oncotarget **8**(8): 14147-14157.
Xue, Q., L. Cao, X. Y. Chen, J. Zhao, L. Gao, S. Z. Li and Z. Fei (2017). "High expression of MMP9 in glioma affects cell proliferation and is associated with patient survival rates." Oncol Lett **13**(3): 1325-1330.
Yamamoto, M., R. Sawaya, S. Mohanam, A. K. Bindal, J. M. Bruner, K. Oka, V. H. Rao, M. Tomonaga, G. L. Nicolson and J. S. Rao (1994). "Expression and localization of urokinase-type plasminogen activator in human astrocytomas in vivo." Cancer Res **54**(14): 3656-3661.
Yang, W. B. and C. Y. Li (2014). "Correlations of MUC15 overexpression with clinicopathological features and prognosis of glioma." J Huazhong Univ Sci Technolog Med Sci **34**(2): 254-259.
Yao, J., S. Xiong, K. Klos, N. Nguyen, R. Grijalva, P. Li and D. Yu (2001). "Multiple signaling pathways involved in activation of matrix metalloproteinase-9 (MMP-9) by heregulin-beta1 in human breast cancer cells." Oncogene **20**(56): 8066-8074.
Yarnell, E. and K. Abascal (2011). "Nigella sativa : Holy Herb of the Middle East." Alternative and Complementary Therapies **17**: 99-105.
Yurkovetsky, Z. R., F. Y. Linkov, E. M. D and A. E. Lokshin (2006). "Multiple biomarker panels for early detection of ovarian cancer." Future Oncol **2**(6): 733-741.
Zhang, C., Y. Han, B. Sun, W. Zhang, S. Liu, J. Liu, H. Lv, G. Zhang and X. Kang (2020). "Label-free serum detection based on Raman spectroscopy for the diagnosis and classification of glioma." Journal of Raman Spectroscopy **51**(10): 1977-1985.
Zhang, X., F. Yu, J. Li, D. Song, H. Li, K. Wang, Q. He and S. Wang (2019). "Investigation on the Cancer Invasion and Metastasis of Skin Squamous Cell Carcinoma by Raman Spectroscopy." Molecules **24**(11).
Zhao, J., G. Li, Z. Zhao, J. Wang, G. Gao and S. He (2012). "Matrix Metalloproteinase-9 Expression is Increased in Astrocytic Glioma and Associated with Prognosis of Patients." Japanese Journal of Clinical Oncology **42**(11): 1060-1065.
Zhao, Y., C. E. Lyons, Jr., A. Xiao, D. J. Templeton, Q. A. Sang, K. Brew and I. M. Hussaini (2008). "Urokinase directly activates matrix metalloproteinases-9: a potential role in glioblastoma invasion." Biochem Biophys Res Commun **369**(4): 1215-1220.
Zhi, X., J. Tao, K. Xie, Y. Zhu, Z. Li, J. Tang, W. Wang, H. Xu, J. Zhang and Z. Xu (2014). "MUC4-induced nuclear translocation of β-catenin: a novel mechanism for growth, metastasis and angiogenesis in pancreatic cancer." Cancer Lett **346**(1): 104-113.
Zhou, W. and D. R. Wahl (2019). "Metabolic Abnormalities in Glioblastoma and Metabolic Strategies to Overcome Treatment Resistance." Cancers (Basel) **11**(9).
Zhou, W., X. Yu, S. Sun, X. Zhang, W. Yang, J. Zhang, X. Zhang and Z. Jiang (2019). "Increased expression of MMP-2 and MMP-9 indicates poor prognosis in glioma recurrence." Biomed Pharmacother **118**: 109369.
Zhou, Y., C. H. Liu, Y. Sun, Y. Pu, S. Boydston-White, Y. Liu and R. R. Alfano (2012). "Human brain cancer studied by resonance Raman spectroscopy." J Biomed Opt **17**(11): 116021.
Zhou, Y., C. H. Liu, B. Wu, X. Yu, G. Cheng, K. Zhu, K. Wang, C. Zhang, M. Zhao, R. Zong, L. Zhang, L. Shi and R. R. Alfano (2019). "Optical biopsy identification and grading of gliomas using label-free visible resonance Raman spectroscopy." J Biomed Opt **24**(9): 1-12.