Magnetic Resonance Imaging and Spectroscopic Markers for the Detection and Metabolic Characterisation of IDH-Mutant Gliomas

Thomas W. Leather BSc. MRes

For the degree of Doctor of Philosophy

University of Liverpool

September 2020

Centre for Pre-clinical Imaging

Department of Molecular and Clinical Cancer Medicine

Institute of Systems, Molecular and Integrative Biology

University of Liverpool

Crown Street, Liverpool, L69 3BX, UK

Acknowledgments

Table of Contents

Abbreviations

2-HG – 2-hydroxygluarate
AA – anaplastic astrocytoma
Ala – alanine
ALAT – alanine aminotransferase
ANTS – advanced normalisation tools
ASL – arterial spin labelling
Asp – aspartate
AUC – area under curve
Cho – choline
CNS – central nervous system
COSY – correlation spectroscopy
CPMG – Carr-Purcell-Meiboom-Gill
CRLB – Cramer-Rao Lower Bounds
CT – computed tomography
DWI – diffusion weighted imaging
EGLN – egl-9 family hypoxia inducible factor
EORTC – European organisation for research and treatment of cancer
FDR – false discovery rate
FDR – false discovery rate
FID – free induction decay
FLAIR – fluid attenuated inversion recovery
FOV – field of view
fSM – functional spectroscopic mapping
GA – genetic algorithm
GABA – gamma-aminobutyric acid
GBM - glioblastoma
Glc – glucose
GLCM – grey level cooccurrence matrix
Gln – glutamine
GLRLM – grey level run length matrix
Glu – glutamate
Gly – glycine
GLZLM - grey level zone length matrix
Gpe – phase encoding gradient
GRE – gradient echo
Gro – read out gradient
GSH – glutathione
Gss – slice selection gradient
HGG – high grade glioma
HMBC - heteronuclear multiple bond correlation
HSQC – heteronuclear single quantum coherence
IBSI – image biomarker standardisation initiative
IDH – Isocitrate dehydrogenase

INEPT – Insensitive nuclei enhanced by polarization transfer

ISN – international society for neuropathologists

KNN – K-nearest neighbour

Lac – lactate

LC-MS – liquid chromatography-mass spectrometry

LGG – low grade glioma

MGMT - O-6-methylguanine DNA methyltransferase

mM – millimolar

MRS – magnetic resonance spectroscopy

MTT – mean time to malignant transformation

Mut – mutant

NAA – N-acetyl aspartate

NAAG - N-acetlyaspartylglutamate

NMR – nuclear magnetic resonance

NMR – nuclear magnetic resonance

OS – overall survival

PCA – principle component analysis

PCho – phosphocholine

PCs – principle components

PE – phosphotidylethanolamine

PET – positron emitting tomography

PFS – progression free survival

PLS-DA – partial least squares discriminant analysis

PRESS – point-resolved spectroscopy

RF – radiofrequency

ROC – receiver operator curve

ROI – region of interest

Scyllo-Ins – scyllo-inositol

SE – spin echo

Semi-LASER – semi-localisation by an adiabatic selective refocusing

SNR – signal-noise ratio

SVM – support vector machine

Tau – taurine

tCr – total choline

TE – time to echo

TET – Ten Eleven Translocation

TOCSY – total correlation spectroscopy

TR – repetition time

TSP - 3-(trimethylsilyl) 3,3,3,3-tetradeutero-propionic acid

VIP – variable importance in projection

VOI – volume of interest

WHO – world health organisation

WT – wild type

αKG - α-ketoglutarate

Abstract

Background: The discovery of mutations in the isocitrate dehydrogenase (IDH) 1/2 gene induced a paradigm shift in the determination of diagnosis, prognosis and treatment of glioma. In 2016 the World Health Organisation (WHO) incorporated mutations in IDH1/2 into the diagnostic criteria for intracranial tumours as they have been demonstrated in 70-80% of WHO grade II-III gliomas, as well as in grade IV secondary glioblastoma (GBM)[1].

Methods: This thesis explores several methods to further investigate and understand the non-invasive detection of IDH mutation as well as the pathophysiology of the mutation. Firstly, the extraction of quantitative features from MR images of glioma patients, and the subsequent application of machine learning strategies in order to carry out non-invasive classification according the IDH mutational status. Second was use a metabolomics analytical approach to 1 and 2 dimensional NMR spectroscopy to understand the metabolic implications of the presence of the mutation. Finally, the exploration of the utility of a novel magnetic resonance spectroscopy (MRS) sequence for the non-invasive detection of the IDH mutation induced oncometaboite 2-hydroxygluerate (2-HG).

Results: The use of multiway ANOVA allowed for selection of imaging features that were independent of external variables to be used in the classification of images into IDH Mut and WT. The precision metrics indicated that the investigation warrants further investigation to ensure the validity of classification. A Metabolomics approach to multi-dimensional NMR spectroscopy allowed for the identification of a metabolic phenotype associated with IDH mutation. 2-HG, Myo-Inositol, Alanine, Choline, 3-hydroxybutyrate and Methyl succinate were all found to be significantly altered between IDH Mut and WT tumours, which may aid in understanding the malignant behaviour of these tumours, as well as providing a potential spectroscopic fingerprint for diagnosis. Finally, the utility of a novel semi-LASER MRS sequence for the detection of 2-HG non-invasively failed to yield reliable and accurate results.

Conclusion: Although the radiomics investigation was unable to reliably classify images into IDH Mut and WT based on image texture features alone, the NMR metabolomics experiment revealed an IDH Mut metabolic phenotype that may provide both biological and diagnostic information. Taken together, the future of personalized tumour healthcare and diagnostics is likely to incur a multifaceted approach, bringing together information from a variety of sources to provide detailed information regarding the status of the tumour.

1. Gliomas

1.1 Background

Gliomas represent the most common form of intracranial tumours in adults, accounting for 81% of malignant brain tumours[2]. They may arise anywhere within the central nervous system but are most often observed within the brain. These tumours originate from either astrocytic, oligodendrocytic or a mixture of the two lineages, and have been classified by the World Health Organisation (WHO) into both a grading system, and histological and molecular subgroups[1]. Gliomas are most commonly categorised into grades I-IV based on their malignant behaviour.

Brian tumours are typically synonymous with high morbidity and mortality due to a vast range of factors including their problematic localisation and often locally invasive and aggressive nature. This heterogenous set of tumours are known to arise from neuroglial progenitors, and hence are classified into their subcategories based on morphological likeness to the neuroglial cell types found in the healthy human brain. Other descriptive factors that are often used in pathological categorisation include location (specific cortices, brain stem lesions, optic nerve), distinctive differentiation patterns (for example pilomyxoid and pilocytic) and anaplastic features[3]. When considering the WHO grading system, features of anaplasia are of particular importance when assigning malignancy grade to an individual tumour, whereby WHO grade I displays the least malignant behaviour, and WHO grade IV exhibits the most aggressive, infiltrative malignant conduct.

The past decade has seen extensive progress being made in understanding the molecular mechanisms that drive the pathogenesis of gliomas. In May 2014, the international society for neuropathologists (ISN) called a meeting in Haarlem, Netherlands, in order to determine the best method of incorporating molecular information into tumour diagnostics. The resultant "ISN Haarlem guidelines", whereby a layered categorisation system was proposed with histological classification (layer 2), WHO grading system (layer 3) and molecular information (layer 4) listed below an integrated diagnosis (layer 1), so as to reduce the inherent interobserver variability that most often accompanies traditional pathological diagnosis, and to increase the accuracy of prognostic prediction[4]. These developments have increased our awareness of molecular and genetic events that contribute towards and drive malignancy, so much so that in 2016, the WHO incorporated a list of molecular markers into the diagnostic criteria for tumours of the central nervous system (table 1.1). This new integrated approach to pathological diagnosis, combining both traditional morphological characteristics and genetic,

epigenetic and transcriptional aberrations has paved the way for not only increased diagnostic classification, but also facilitation of a more personalised approach to targeted therapies in the treatment and management of the disease[4]. The utility of such a strategy has been demonstrated in Sahm et al., whereby the disputed group of oligoastrocytomas may be completely resolved into discrete groups of either oligodendroglioma or astrocytoma based on the status of three particular molecular makers; mutations in the isocitrate dehydrogenase 1 or 2(IDH1/2) gene, loss or retention of expression of nuclear alpha-thalassemia/mental retardation syndrome X-linked (ATRX) and co-deletion of chromosome arms 1p and 19q.

Marker	Diagnostic/Prognostic Significance
IDH1/2 Mutation	Usually first genetic event, prognostic marker for survival and treatment response
1p/19q codeletion	Marker for oligodendroglioma and predictive value in treatment response of anaplastic oligodendroglioma
MGMT Methylation	Predictive of improved outcome in anaplastic oligodendroglioma, astrocytoma and GBM. May be predictive of response to alkylating agents such as temozolomide
ATRX loss	If lost, correlates with reduced median survival and increased genetic instability
PTEN	has been shown to corelate with poor prognosis in grade III glioma and GBM
EGFR amplification	Indicative of increased tumour invasiveness in GBM and anaplastic oligodendroglioma and astrocytoma

Table 1.1 List of the molecular markers that have been identified as important in the diagnosis and prognosis of glioma by the WHO and incorporated into the WHO diagnostic criteria 2016

1.2 Epidemiology

Primary brain tumours are the 9th most common form of tumour in the UK which accounts for 3% of all newly diagnosed tumours. The incidence rate in 2017 for the UK was reported at 18.8 cases per 100,000 population[5]. These tumours have a higher incidence rate amongst older members of the population where in 2017 23% of new cases were in patients aged 75 and over[5]. Gliomas, in particular grade IV GBM have a dismal mortality, whereby the average survival time post-diagnosis is 12-18 months, with only 25% of GBM patients surviving more than 12 months, and only 5% of patients surviving over 5 years[5].

1.3 Pathophysiology

Gliomas, much like other tumours, may be considered as a genetic disease of a single cell originating from a specific lineage. The architecture of these genetic alterations gives rise to the clinical manifestation of the tumour, and initial genetic and molecular events shape the evolution of populations of malignant cells within each lesion. These distinctive features of brain tumours are increasingly involved in the categorisation and diagnosis of glioma.

The most common subtypes of glioma to be observed in the young adult population are diffuse WHO grade II and III tumours. Such lesions may be distinguished by their distinct infiltrative nature within the brain parenchyma and an innate propensity to both recur and progress towards an increasingly malignant phenotype. The aforementioned inclusion of molecular markers into the diagnostic criteria has previously expedited the differentiation of mixed oligoastrocytomas into distinct classes of astrocytoma and oligodendroglioma, averting the requirement for the mixed subtype[6]. Both the astrocytic and oligodendrocytic genotypes may carry mutations in the IDH1/2 gene[7,8]. The role of IDH1/2 mutations in the pathogenesis of glioma will be discussed in further detail in following sections, however this mutation is likely to represent the primary tumour-initiating event of diffuse gliomas[9]. Malignant evolution of diffuse, lower grade gliomas into grade IV glioblastoma (GBM), is most often restricted to those tumours defined as an astrocytic subtype, and occurs due to the homozygous deletion of the CDKN2A-CDKN2B locus, among other genetic modifications[10]. Anaplastic astrocytoma (AA) is an invasive, malignant primary tumour that is more common amongst younger primary CNS tumour patients (median age 41 years)[11]. These tumours have been demonstrated to share common pathological lineage to that of secondary GBM whereby the most pronounced difference between AA and GBM is that AA are lacking in endothelial proliferation and necrosis. IHD mutation, much like in secondary GBM, plays an important role in the pathological and malignant progression of AA[11,12].

WHO grade IV GBMs are the most commonly observed glioma and display the most aggressive, malignant phenotype. These tumours primarily present with demographics >50 years of age, nevertheless GBM may also occur (to a lesser extent) in children and young adults. Interestingly, there is a growing body of evidence to suggest that GBM in young demographics are molecularly and genetically distinct from their counterparts observed in adults[13]. Recent findings have shown there to be six subgroups of GBM characterised by distinct molecular alterations and methylation profiles[13].

1.4 IDH Mutation in Glioma

The field of neuro-oncology has undergone a paradigm shift following the identification of mutations in the IDH 1/2 genes in glioma and gliomagenesis. Such mutations have been suggested to play an important role in the development and progression of intracranial gliomas, often marking the first genetic transformation to occur, and imparting a range of downstream effects on a myriad of cellular processes[14]. Mutations in IDH1/2 genes have been identified in 70–80% of WHO grade II-III gliomas, and these mutations are particularly prevalent in grade IV secondary glioblastomas (GBM) arising as a progression from lower grade tumours, suggesting such mutations may act as a driving force behind malignant progression[1,7]. Mutations in IDH1/2 confer a gain-of-function neomorphic enzymatic activity, resulting in the aberrant production and subsequent accumulation of 2-hydroxygluarate (2-HG), which has been suggested to be an oncometabolite for this genetic mutation[15]. Magnetic resonance spectroscopy (MRS) has been identified as a tool in the diagnosis of IDH mutant gliomas via the non-invasive detection of 2-HG[16–19]. Alongside diagnostic applications, non-invasive detection and quantification of 2-HG levels within the mutant gliomas is highly desirable for the development of targeted treatment and response monitoring. The mutated pathway has been identified as a candidate target for novel therapies, and imaging strategies capable of accurately examining 2-HG levels will aid in longitudinal assessments of treatment response, whilst accelerating pharmaceutical translation from bench to bedside[20–22].

1.5 Biochemistry of IDH Mutation in Glioma

IDH is known to occur in three structural isoforms, namely IDH1, 2 and 3; however, only mutations in IDH1/2 have been identified in human gliomas[15]. IDH1 is localised to the cytoplasm and peroxisomes, whereas IDH2 is localised to the mitochondria[23]. Both of the IDH1/2 isoforms function to catalyse the oxidative decarboxylation of isocitrate into α-ketoglutarate (αKG) using NADP+ as a cofactor[24]. Mutations in both IDH1 and IDH2 are ubiquitously expressed within the tumour tissue; moreover, they consistently occur in a heterozygous manner and are most commonly observed in the IDH1 isoform. Remarkably, within gliomas, less than 90% of these point mutations occur at a single residue within the

active site of the IDH1 isoform, where the Arginine 132 residue is replaced with a histidine (IDH1R132H)[14]. The much less frequent mutations in the IDH2 analogue occur at Arginine 172,
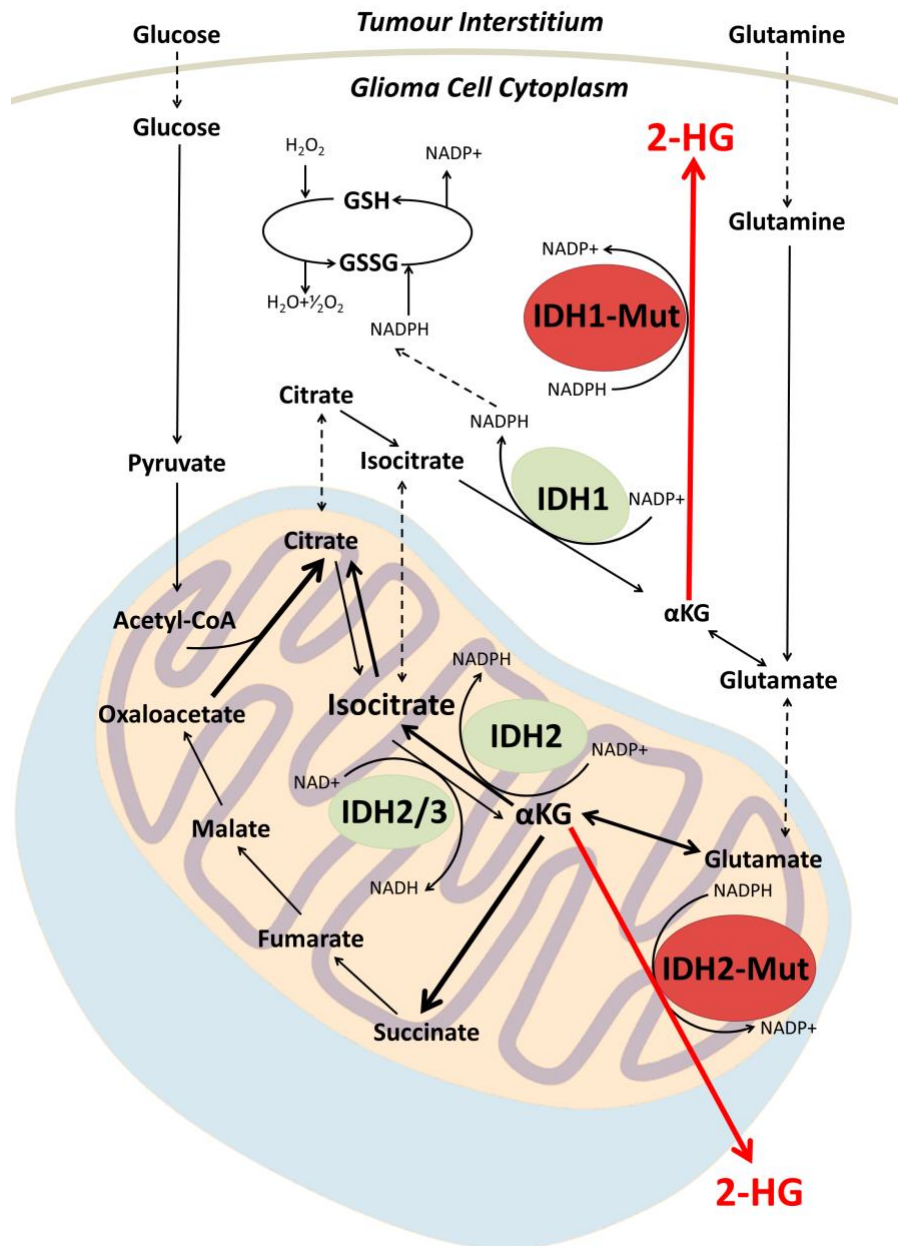
Fig 1.1. Simplified schematic of metabolic pathways associated with IDH1/2. Mutations in IDH1/2 catalyze the production of the oncometaboite 2-HG via reduction of αKG. Accumulation of 2-HG within the cytoplasm surpasses the detection threshold of MRS at clinical filed strengths, and therefore may be utilized as a biomarker for the non-invasive detection of the mutation. H_2O_2 = hydrogen peroxide, GSH = glutathione, GSSG = glutathione disulphide, H2O = water, NADP = nicotinamide adenine dinucleotide phosphate, NADPH = nicotinamide adenine dinucleotide phosphate hydrate, Acetyl-CoA = acetyl coenzyme A IDH1 = isocitrate dehydrogenase 1, IDH2 = Isocitrate dehydrogenase 2, 2-HG = 2-hydrozygluarate, αKG= alpha-ketoglutarate[25].

where the most common mutation is IDH2R172K. Alterations in the conformation of the catalytic site of the enzyme confer a gain-of-function neomorphic enzymatic activity, whereby mutant IDH1/2 catalyses the reduction of αKG into the oncometabolite 2-HG (Figure 1.1)[15]. The subsequent accumulation of the metabolite within mutated cells is facilitated by the heterozygous manner of the mutation and the dominant effect that the gain-of-function mutant imparts over the remaining wild type (WT) allele, which serves to provide the mutant enzyme with a continuous supply of αKG to be aberrantly reduced into 2-HG.

The full impact of accumulation of 2-HG within tumour cells is yet to be elucidated, but recent work suggests that it may have a pivotal role in altering the genetic, epigenetic and metabolic profile of the IDH mutant cells, driving the phenotype towards a more malignant state. One well-established mechanism of phenotypic alteration is inhibition of α-ketogluarate-dependant deoxygenases and histone demethylases, including the Ten Eleven Translocation (TET) family of 5-methlycytosine hydroxylases, leading to genome-wide alterations in methylation and histone patterns[15]. This proposed inhibition allows for IDH mutant gliomas to be classified into a distinct subgroup amongst intracranial tumours, whereby they exhibit a CpG island methylator phenotype and vastly altered histone methylation pattern[26]. This aberrantly methylated phenotype has been suggested to contribute to tumourgenesis. Interestingly, the occurrence of IDH mutations is highly restricted to a narrow spectrum, encompassing an exclusive list of malignancies. For example, these mutations are frequently observed in grades II to III gliomas and secondary GBM but not in primary GBM. The observed pattern in mutation occurrence suggests that its involvement in tumourgenesis may be linked to prevention of histone demethylation normally required for linage specific differentiation into terminally differentiated progenitors[27,28], in addition to altering the activity of chromosomal topological-regulating proteins[29]. A further molecular mechanism by which mutations in the IDH1/2 genes have been suggested to contribute to tumourgenesis is via the ability of accumulated 2-HG to stimulate activity of the egl-9 family hypoxia inducible factor (EGLN) prolyl 4-hydroxylases. Activation of these enzymes facilitate- the ubiquitination and proteasomal degradation of hypoxia inducible factor, a process that has been shown to enhance the proliferation of astrocytes in culture[30]. 2-HG has also been suggested to indirectly contribute to increased exposure to oxidative stress, as the conferred neomorphic enzymatic activity observed in mutant cells incurs a reduction in NADPH levels, a molecule that normally functions as a prerequisite for the reduction of glutathione disulphide into the antioxidant

glutathione (Figure 1.1)[31]. Although a decrease in glutathione levels increases exposure to free radicals and reactive oxygen species, it has been suggested that the cellular depletion of glutathione contributes to the increased sensitivity to adjuvant radiotherapy in IDH mutated gliomas[31].

One approach that has been adopted for probing IDH mutation uses ^{13}C MRS to monitor the production of 2-HG from its substrate αKG in a rat glioma model[32]. This preclinical study used Dynamic Nuclear Polarization labelled 1-^{13}C αKG to examine the metabolic activity of tumour cells, providing information pertaining both the metabolic origin of the aberrantly produced 2-HG, as well as tracking of real-time metabolism. It was demonstrated that the injected hyperpolarized 1-^{13}C αKG was converted to hyperpolarized 1-^{13}C 2-HG in both cell lysates and *in vivo* in orthotopic tumours. In comparison to ^{13}C, ^{1}H MRS provides total, steady-state 2-HG levels present throughout the tumour environment, inclusive of both the intracellular and extracellular compartments. Alternatively, ^{13}C MRS provides real-time, dynamic information on the metabolic fate of ^{13}C α-KG, and thus directly probes the enzymatic activity of cytosolic mutant IDH1 in glioma cells. The study of dynamic metabolite processes permitted by ^{13}C MRS offers the opportunity to instantly gain a picture of the metabolic activity taking place within a tumour. The monitoring of active mutant IDH enzymes, as well as 2-HG synthesis, confirms the proposed biochemical pathway leading to the accumulation of 2-HG within glioma cells. In addition, utilization of a non-radioactive hyperpolarized 1-^{13}C αKG probe highlights clinical applicability of such detection methods for the accurate tracking of oncogenic metabolic activity when monitoring disease progression; moreover, as IDH targeted therapies emerge, such techniques delivering high-quality information on 2-HG synthesis may be invaluable in the development and monitoring of novel compounds through clinical trials[33,34].

Taken together, the production and accumulation of 2-HG within IDH mutated glioma cells not only offers an interesting avenue for investigation into the underlying pathological events that contribute to glioma formation and progression, but correspondingly presents as a potential biomarker for diagnostic confirmation of IDH mutational status, in addition to prospective opportunities to monitor treatment and develop targeted therapies.

1.6 2-HG as a Biomarker

Although the downstream consequences of metabolic reprogramming observed in IDH mutated gliomas are yet to be fully elucidated, the incorporation of IDH status into the WHO diagnostic criteria indicates a significant diagnostic and prognostic role of the genetic abnormality. IDH status is currently confirmed by immunohistochemistry or genetic sequencing, requiring tissue samples obtained at surgery.

As cancer treatments move towards personalised medicine, the need for robust and quantitative biomarkers becomes increasingly important. The ability of 2-HG to act as a surrogate marker of IDH status is highlighted by the fulfilment of several criteria. Firstly, it is known that > 99% of all IDH mutated cells exhibit 2-HG levels several orders of magnitude above the physiological trace amounts observed in healthy brain cells. In fact, accumulation of the metabolite occurs in such an extensive manner that its concentration reaches well into the millimolar (mM) range (5–35 mM) [35], far surpassing the sensitivity threshold of MRS at clinical field strengths (~1 mM)[36]. Secondly, there is virtually no background 2-HG present in healthy brain tissue. The 2-HG metabolite contains an asymmetric carbon within its backbone and is therefore produced as two enantiomers, either D-R- or L-S- confirmation. The D-R-2-HG stereoisomer is the metabolite that is produced by the neomorph enzymes resulting from IDH mutation, whereas the L-S-2-HG isomer is produced as a result of limited oxygen availability within a tissue. Both stereoisomers function to inhibit αKG-dependant enzymes. Physiologically, both enantiomers of this metabolite are rarely produced, and form only as transient errors of normal metabolism to be swiftly degraded by endogenous 2-HG dehydrogenases[15]. Finally, there is only one other known cause of elevated levels of 2-HG within brain tissue, a rare genetic disorder known as hydroxyglutaric aciduria, a condition recognised early in brain development due to the presentation of a range of additional clinical manifestations[37].

1.7 Prognostic Implications of IDH Mutation

Overall survival (OS), treatment response and probability of malignant transformation are all factors that correlate with IDH mutation[34,38,39]. The prognosis of glioma is dependent on several factors including tumour grade as defined by the WHO classification[1]. Infiltrative low grade gliomas (LGGs) (grade II) have a median survival of 7–10 years, whilst high grade gliomas

(HGGs) (WHO grade III and IV) have a poorer prognosis with a median OS ranging from a few months to a few years[39]. LGG (including oligodendroglioma, astrocytoma and mixed glioma) have an incidence rate of 0.63/100,000 adults per year and hold the potential to undergo malignant transformation into WHO grade III anaplastic glioma or grade IV secondary GBM[40]. Mutations in IDH1/2 within LGGs are highly prevalent, and can be observed at a rate of 70–80% in all tumours within this subgroup. The influence of IDH mutation on LGG patients has been highlighted in the stratification of oligodendroglioma, astrocytoma and mixed glioma[41]. That study evaluated the accumulation of a number of additional genetic aberrations alongside IDH mutation, including hypermethylation of O-6-methylguanine DNA methyltransferase (MGMT) promotor region, 1p/19q co-deletion and TP53 mutation. It was reported that all genetic subsets, inclusive of IDH mutation, showed hazard ratios 3–5 times more likely to undergo malignant transformation. Although LGG have been stratified with regards to the prognostic influence of IDH mutation, the importance of the genetic abnormality in this subset of tumours is still a topic of debate, as others have failed to identify such a significant correlation[42,43]. In addition to the increased risk reported by Leu et al.[43], a study conducted by Sanson et al.[44] observed a significant increase in OS in patients harbouring an IDH mutation. Conversely, other studies have not found a link between IDH status and OS[42,43,45]. Kim et al. reported no significant influence of IDH on progression free survival (PFS) or OS in a cohort of nearly 200 LGGs[45]. Furthermore, Juratil et al. failed to identify any predictive value of the mutation with regard to OS, PFS and mean time to malignant transformation (MTT) in a multivariate analysis[46]. This investigation did, however, identify that a higher rate of patients with mutated IDH were to develop into a secondary high grade glioma. A comprehensive analysis of cellular tricarboxylic acid cycle metabolites was conducted on a range of glioma and non-glioma patients, with both IDH mutant and WT genetic profiles. However, this study failed to identify 2-HG as an early biomarker in determining the prognostic course of glioma patients[45].

Whilst the influence of IDH mutation on the malignant progression of grade II gliomas is yet to be fully understood, the favourable impact on patient outcome for grade III gliomas is recognised. A number of studies have outlined the prognostic, rather than predictive influence of IDH mutation on grade III gliomas[42,43]. A prospective study by the European Organization for Research and Treatment of Cancer (EORTC) determining the prognostic significance of IDH mutation in anaplastic oligodendroglioma patients found that IDH mutant oligodendrogliomas

had a better OS. However, the study did not find a link between IDH status and response to adjuvant radiation and chemotherapy[47]. Comprehensive, multiplatform integrative genomic analysis of low and intermediate grade gliomas demonstrated that lesions histologically identified as grade III display only a modest difference to those identified as grade II, whereby the majority of observed differences results in a slight increase in frequencies of chromosome 9p and 19q losses[48]. Such analyses highlight the suggested molecular progression through grade classification, initiated by mutation in IDH1/2 and acquisition of the CpG island methylator phenotype. Despite the modest differences in genetic profile, the confirmation of IDH mutation in grade III is considered to reflect improved PFS and OS[47].

In WHO grade IV GBM, IDH mutation has been repeatedly shown to be predictive of improved response to temozolomide treatment and radiotherapy in secondary GBM malignancies that have transformed from lower grade gliomas[48,49]. The presence of IDH mutation has additionally been associated with improved PFS and OS of secondary GBM[50]. Secondary GBM that also harbour hypermethylation of the MGMT promoter region alongside IDH1/2 mutation have the best prognosis[51]. Although the underlying molecular mechanisms behind IDH mutation and its proposed influences on diagnosis, prognosis and the suggested predictive capacity of the mutation are yet to be identified, there appears to be overwhelming evidence that this genetic alteration has a significant role to play in understanding the underlying pathology of gliomagenesis and tumour progression. There is therefore an unmet need for the development and implementation of a reliable and robust, non-invasive assay for the detection of this mutation.

2. Principles of Experimentation

There are numerous applications of nuclear magnetic resonance for the study of glioma metabolism and diagnostics with reference to IDH mutation. In this chapter I will introduce and discuss the principles that underpin the MR methods explored in this work, including nuclear magnetic resonance imaging and both *in-* and *ex-vivo* MR spectroscopy.

2.1 Basics of Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) refers to the phenomenon of specific atomic nuclei absorbing and remitting radiation of specific frequencies when exposed to an externally applied magnetic field. Quantum theory can aid in understanding the principles underpinning NMR, and may allow us to explain how we are able to generate images (Magnetic Resonance Imaging) and spectra (Magnetic Resonance Spectroscopy) in order to probe the human body to asses a vast range of biophysical properties[52].

2.2 Nuclear Spin Properties

The nucleus of an atom is comprised of a given number of protons and neutrons. These subatomic particles are fermions, and as such, according to spin-statistic theory from within any reasonable relativistic quantum field theory, possess a half integer spin value. Spin is an intrinsic property of any quantum particle, and here we may think of this property as a particle rotating about its own axis. Atoms consisting of an odd number of protons and neutrons are considered to possess a non-zero spin, and any spinning charged particle is known to emit electromagnetic radiation. One fundamental property of electromagnetic waves dictated by Maxwell's wave theory, is that they are comprised of an electric field E and a magnetic field B, positioned perpendicular to each other. Therefore, the nucleus of a hydrogen atom consisting of a single proton, with a non-zero spin emitting electromagnetic radiation, possesses a magnetic moment along the axis of that spin. This nuclear magnetic field attributed to hydrogen atoms is analogous to that of a bar magnet.

The interaction between a nuclear spin placed within a magnetic field can be explained by the Hamiltonian operator in equation 1:

Equation 1

$$\hat{H}_{one\ spin} = -\gamma B_0 \hat{I}_Z$$

Where γ is the gyromagnetic ratio, which is a fundamental property of the spin. $\hat{I}_Z$ is an operator representing the *z*-component of the nuclear spin angular momentum, which is the component that interacts with the applied magnetic field. The eigenvalues of this Hamiltonian describe the energy of a spin state where $\hbar$ is Planck's constant and γ is the gyromagnetic ratio (equation 2):

Equation 2

$$E = -m\hbar\gamma B_0 \qquad m = +\frac{1}{2}\ or\ -\frac{1}{2}$$

These are the two energy levels of a single spin-half in a magnetic field B_0 (Fig. 2.1).
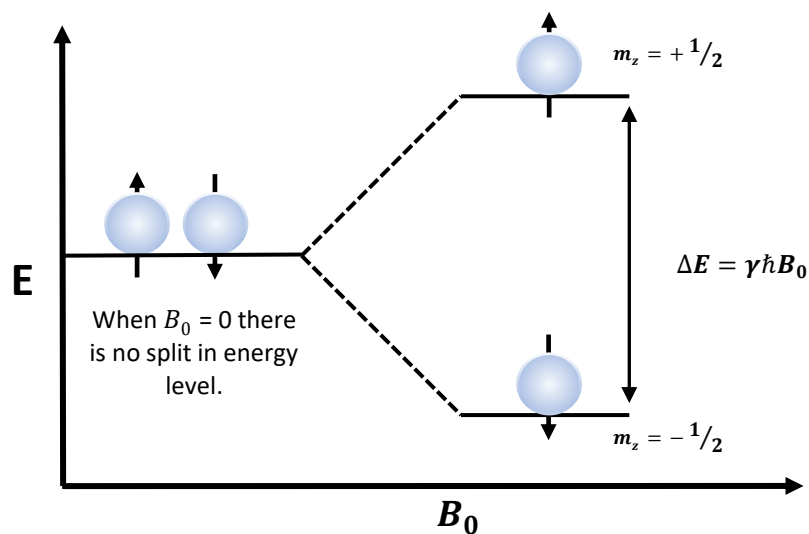
Fig 2.1 Energy level split for a population of ¹H spins (*I* = 1/2). In the absence of a magnetic field the energy levels are almost absent, however when a magnetic field is applied along the Z-axis there is a splitting of spin states.

Following this rule of quantum theory, it can be determined that a hydrogen atom has 2 energy states, -½ and +½. This was classically demonstrated and quantised in an experiment conducted by Stern and Gerlach in 1922, where a beam of silver ions was passed through a

magnetic field gradient[53]. They found that a given number of silver ions were deflected into one of 2 paths, which dictated the number of ions in the spin up (+½) or spin down (-½) state. When considering the previous description of spinning magnetic moments of protons, such as hydrogen atoms that are found in abundance in biological tissue, these tiny magnets can either be positioned within the higher energy state (+½) whereby they will point upwards, or in the low energy state (-½) whereby they will point downwards[54].

One key intrinsic property of a nucleus is the magnetic moment μ caused by nuclear magnetism. The nuclear magnetic moment μ and the nuclear spin angular momentum I are proportionally linked to the constant γ which is termed the gyromagnetic ratio. Magnetic moments can be defined as vectors defining the spin polariazation (equation 3) and may take on the value of:

Equation 3

$$\mu = \gamma . \hbar . I$$

Where γ is the gyromagnetic ratio, $\hbar$ is Planck's constant over 2π and I is the nuclear spin angular momentum(kg-m^2/s)[55].

Hydrogen atoms can be found in vast abundance throughout a wide variety of tissues in the human body. When considering the above description of nuclear spin in hydrogen atoms, these tiny magnets can assume a position within either the high energy state (+½) or the low energy state (-½). Under normal conditions, where such atoms have no externally applied magnetic field, the distribution of protons occupying either one of the two energy states is equal. When an external magnetic field is applied ($B_0 > 0$) the spin state possessing the net magnetic moment aligned to the B_0 field is energetically favourable, consequently there is a preference for the majority of the spins to align parallel to the B_0 field. This anisotropic spin distribution ($\Sigma\mu > 0$)(euqation4), which results in a positive net magnetization at thermal equilibrium is termed the macroscopic magnetization:

Equation 4

$$M_0 = \Sigma\mu$$

The net magnetization produced by the protons aligned within the magnetic field is subject to a torque force. This force, referred to as precession, results from the phenomenon observed when a proton is exposed to an external magnetic field, not only does it rotate about its own axis, but it begins to 'wobble' or precess about the axis of B_0[56]. The rate ω_0 (rad.s^{-1}) at which this rotation occurs is directly related to the strength of the magnetic field B_0, and is denoted by the equation 5, otherwise known as the Larmor frequency:

Equation 5

$$\omega_0 = -\gamma B_0$$

$\omega_0/2\pi$ will simply define the Larmor frequency in Hz in equation 6:

Equation 6

$$\omega_0 = -\gamma B_0/2\pi$$

The Larmor frequency can be either positive or negative, suggesting either clockwise (positive) or anti-clockwise (negative) precession about the magnetic field. For the purpose of this text we are considering H^1 nuclei, which incur a positive gyromagnetic ratio, and therefore a clockwise precession about the magnetic field (Fig 2.2).
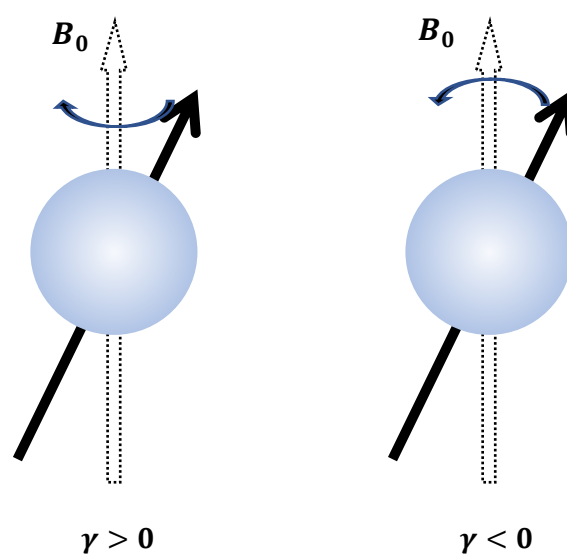
Fig 2.2 Precession of the nuclear spin magnetic moment about the B_0 externally applied magnetic field incurring a positive or negative gyromagnetic ratio.

2.3 Polarization at Thermal Equilibrium

The alignment of spins to the applied B_0 is not a ubiquitous arrangement within the system due to thermal motion of particles enforcing a dynamic balance between the two energy states, which may be determined by the strength of the magnetic field and the temperature within the system. The excess of spins aligned parallel to the B_0 may be explained by the Boltzmann's distribution equation, which can be given by equation 7:

Equation 7

$$\Delta N = N(\uparrow) - N(\downarrow) = \frac{1}{2}N\left(e\frac{\hbar\omega}{\kappa T}\right) = \frac{1}{2}N\left(e\frac{\Delta E}{\kappa T}\right)$$

Where N is the number of spins, $\hbar$ is the reduced Planck constant, ω is the Larmor frequency (Hz), K is Boltzmann's constant and T is temperature in kelvin. Taking into account equation 8, this means the equilibrium spin density $\rho_{eq,i}$ at a given spin state *i* will be equal to:

Equation 8

$$\rho_{eq,i} = \frac{N_\iota}{\Sigma_\iota N} = \frac{\exp(-E_\iota/\kappa_B T)}{\Sigma_{-I}^{I}\exp(-E_\iota/\kappa_B T)}$$

The polarization of a spin system is proportional to the longitudinal magnetization over its maximum value. Therefore, equation 9 shows that in a half spin system (e.g. ^{1}H), where $N\uparrow$ is the number of spins aligned along the B_0 and $N\downarrow$ is the number of spins aligned opposite to the B_0, the system polarization P is equal to equation 9 where γ is the gyromagnetic ratio, $\hbar$ is reduced Planck's constant, κ is the Boltzmann's constant and T is temperature in Kelvin:

Equation 9

$$P = \frac{N\uparrow - N\downarrow}{N\uparrow + N\downarrow}$$

$$P = \frac{\exp\left(-\gamma\hbar B_0/2\kappa_B T\right) - \exp\left(\gamma\hbar B_0/2\kappa_B T\right)}{\exp\left(-\gamma\hbar B_0/2\kappa_B T\right) + \exp\left(\gamma\hbar B_0/2\kappa_B T\right)}$$

The thermal energy is always much greater than the split between the two energy states of the spins, making the thermal equilibrium population very small. The high temperature limit is met, where the exponential is less than 1 i.e. $\frac{E_i}{\kappa_B T}$, the polarization equation stated above, for a half spin, may be simplified to equation 10:

Equation 10

$$P = \frac{|\gamma|\hbar B_0}{2\kappa_B T}$$

Using this equation, we are able to understand why it is possible to generate increased amounts of signal at higher magnetic fields, as polarization is linearly related to the B_0, and the gyromagnetic ratio. The spins are able to absorb radiofrequency radiation equal to the difference in energy states. Therefore, at higher B_0 more spins will become aligned with the magnetic field, resulting an increased magnetic moment and increased interaction with detection coils.

The longitudinal magnetization M_z, represents a very small excess of nuclear spins due to the random thermal motion, and may be represented by equation 11:

Equation 11

$$M_z = M_0 = \frac{N\gamma\hbar}{2}P = \frac{N\gamma^2 B_0}{4\kappa_B T}$$

2.4 Excitation and Detection of Signal

The precession of the magnetization vector is the phenomenon that we actually detect during an NMR experiment. A radio-frequency coil, often composed of a transmit and receive component is placed around the sample from which a measurement is to be taken. The coil is aligned along the *xy*-axis. An oscillating field within the radiofrequency range, with the frequency ν tuned and matched to the Larmor frequency of the sample spins we wish to measure, will induce an energy level transition. Equation 12 describes the energy absorbed by the system in order to induce the transition is equal to the energy difference between $E_{-\frac{1}{2}}$ and $E_{\frac{1}{2}}$.where ω_0 is the Larmor frequency in Hz(equation 12):
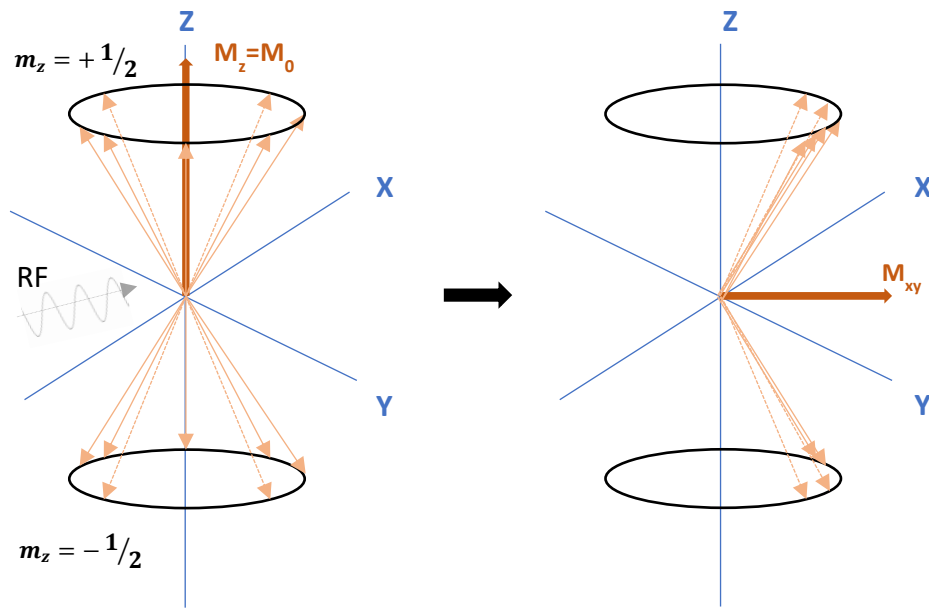
Fig 2.3 Excitation of spins 1/2. On the left the Boltzmann distribution of spins dictates that the bulk magnetization occurs along the Z axis (B_0) and the spins have random phases. An RF pulse is appled and the spins undergo phase coherence such that the bulk of the magnetization now lies in the transverse plane.

Equation 12

$$\Delta E = E_{-\frac{1}{2}} - E_{\frac{1}{2}} = \gamma \hbar B_0 = -\gamma \omega_0$$

The introduction of the RF pulse causes the precessing spins aligned along the magnetization vector M_z to undergo phase coherence, producing a non-zero vector in the transverse plane M_{xy}, shifting the majority of the previously defined polarization away from the Z-axis. As the precessing magnetization vector cuts through the detection coil placed over the sample perpendicular to the position of the coil, an oscillating current is generated which can then be amplified and detected to produce the free induction decay. As the RF pulse is switched off the transverse magnetization M_{xy}, will begin to decay to its zero value M_0 at a rate dictated by the constant T_2^*, whilst the longitudinal magnetization will begin to decay to its equilibrium value M_0. Within the stationary reference frame this relaxation process occurs as a dense spiralling precession about the B_0 where M_0 spins at the Larmor frequency whilst undergoing

both longitudinal (z-axis) recovery and transverse (x,y-plane) relaxation (figure 2.3) If we view this process within the rotating frame, then it is simplified and the M_0 returns to equilibrium along a linear path whereby both longitudinal and transverse relaxation occur exponentially.

2.5 Relaxation Mechanisms

Relaxation can be outlined as the process by which the bulk of the magnetization within a system returns to equilibrium over time. At equilibrium, as previously established, magnetization occurs only along the direction of the applied magnetic field, otherwise known as z-magnetization. Rather than measuring this longitudinal magnetization, in NMR experiments the transverse magnetization, M_{xy}, is measured.

Following the application of a 90° RF pulse, the bulk magnetization (M_z) is flipped into the transverse plane. As the RF pulse is turned off, the spins, which are now in the transverse plane, will begin to precess about the B_0, and it is this precession that we detect in the form of an FID. Rather quickly, the spins begin to lose this phase coherence inducing a relaxation in transverse magnetization termed the spin-spin relaxation, or the T$_2$ relaxation. If this relaxation is measured and plotted it will show an exponential decay, which may be described as:

Equation 13

$$M_{xy}(t) = M_0 e^{-t/T_2}$$

Where t is the time (milliseconds) following the flip into the transverse plane, and M_0 is the initial magnetization. This transverse relaxation occurs much more rapidly than longitudinal recovery. Immediately following RF pulse, the spins can be considered as pressesing in phase with each other. Due to slight variations in magnetic field strength at both the microscopic and nanoscopic levels, these spins experience slightly different magnetic fields and therefore begin to process at slightly different frequencies. These differences in precessional frequencies lead to magnetic interactions between neighbouring nuclei inducing further dephasing until the transverse component of the net magnetisation vector has a value of zero. Transverse relaxation follows and exponential decay curve described by equation 13. In any real-world experimental situation, this transverse relaxation occurs at a faster rate than we would normally predict via inter-spin nuclear interactions. The reason for this increase in rate of decay

is the inhomogeneities within the main magnetic field B_0. These inhomogeneities arise due to a number of reasons including poor shimming, magnetic susceptibility at boundaries between different tissue types and the presence of magnetic particles. This accelerated T_2 is referred to as T_2^* and follows the same relaxation equation as previously described. This T_2^* relaxation can be recovered using a spin echo acquisition, and the recovery is possible because the T_2^* relaxation is not random, but rather is dependent on position within the field and is therefore not a true relaxation process.[55]

The signal that is obtained from the FID obtained by measuring this relaxation is actually the T_2^*, which takes into account the inhomogeneities within the B_0. Equation 14 shows that this T_2^* can be expressed as:

Equation 14

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_{2\,inhomo}}$$

As the M_{xy} magnetization begins to relax, the z-magnetization that was flipped into the transverse plane begins to slowly recover so that it reaches the thermal equilibrium value of M_0. This phenomenon is termed the spin-lattice, or T_1 relaxation, as it is influenced by the lattice of nearby spin systems. It may be defined by equation 15:

Equation 15

$$M_Z(t) = M_0 \left(1 - 2e^{-\frac{t}{T_1}} \right)$$

Where M_Z is the component of magnetization along the longitudinal (z) axis, M_0 is the unperturbed magnetization vector, t is the time since recovery began and T_1 is the relaxation constant that determines recovery[54].
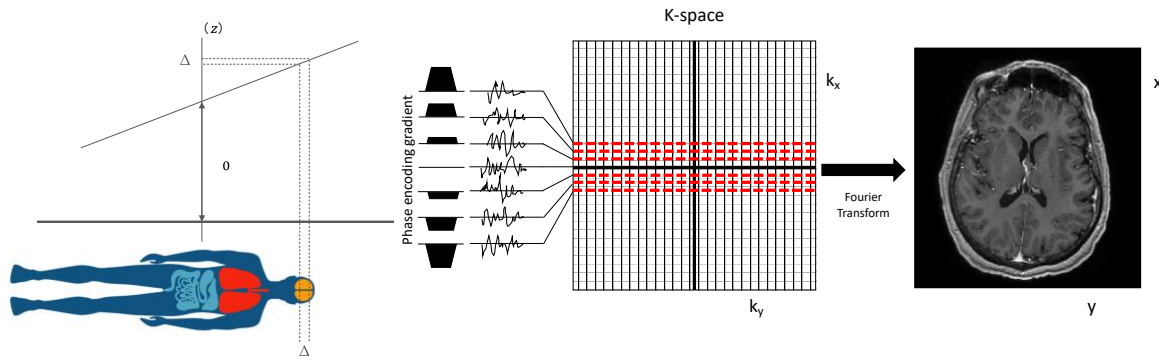
Fig 2.4 When a magnetic field gradient in the orientation of z is superimposed onto the B_0 every local spin at position z is exposed to a unique magnetic field strength. The result is a location-dependent Larmor frequency. When the RF excitation bandwidth is chosen ($\Delta\omega$) only magnetisation of the spins in slice Δz will be excited into the transverse plane. In cartesian acquisition, the k-space is filled line by line. Each phase encoding gradient accounts for a separate line of the k-space. The data is Fourier transformed in order to create the image.

2.6 Image Formation and Reconstruction

The concept of utilising the NMR phenomenon in order to create an image can be traced back to the original experiments of Mansfield and Lauterbur in 1973[57]. They showed that by applying spatially varying magnetic fields, or gradients, spatially dependent precession frequencies can be used to generate tomographic imaging planes[57,58].

In an MRI system the gradients are resistive electromagnets that are embedded within the main body of the magnet and are used to modulate the main magnetic field creating spatial dependence in precession frequency. Three distinct gradient systems are used to modulate the field in the x, y and z planes respectively. These gradient systems are referred to as the phase encoding, frequency encoding and slice select gradients depending on the orientation of the imaging plane. The gradients combine in a linear fashion in combination with the B_0 to produce magnetic field gradients that can be oriented in any direction. The three orthogonal directions created by these gradients spatially localise the NMR signal within a 3D space by altering the Larmor frequency with modulations of the magnetic field experienced by the spins as a function of distance along the given gradient. The spatial encoding procedure can be described in three steps.

First the slice selection gradient is applied (Gss). This is turned on simultaneously with the RF pulse and localises the excitation to a 2D plane with a thickness of Δz. The frequency of the coincident RF pulse will have the bandwidth Δf which is the range of frequencies that will

centre on the Larmor frequency within a slice of tissue with thickness Δz. Where Δz depends on the excitation bandwidth and the steepness of the Gss.

The second gradient applied in the acquisition of an image is the phase encoding gradient (Gpe). Application of this gradient causes the spins that have been excited by the RF pulse and Gss (the imaging plane) to precess at different frequencies along the direction of the gradient. Where the Gpe is lower than the B_0 these spins will precess slower, and where it is higher than the B_0, these spins will precess faster. The result of this is a phase shift that is dependent on the length and amplitude of the pulsed Gpe. The spins continue to precess at the same frequency after the Gpe is turned off, maintaining their phase shift with respect to their position along the Gpe.

The final spatial encoding gradient is the frequency encoding gradient, which is often referred to as the read out gradient (Gro), as it is applied for longer than the other two spatial encoding gradients and coincides with the readout of the RF coil. Much like the Gpe, the Gro causes the spins to precess faster at higher gradient strength and slower at the lower end of the gradient. The RF coils detect the precession frequencies during the application of this gradient, resulting in a composite detection signal encompassing many different frequencies corresponding to all points along the Gro. This process is repeated for a series of different Gpe's thus allowing the Gro to read out one line of phase encoded data per repetition. This process is continued until the whole field of view (FOV) has been sampled.

All the data acquired during the above process is done so within the time domain, and is stored as complex data points in a mathematic matrix known as k-space. Images are comprised of points distributed along x and y planes, whereas the datapoint within the k-space are distributed along the K_x and K_y. Each point in k-space contains both frequency and phase information (figure 2.4). In order to transform the data from the time domain in to the frequency domain, and therefore produce an image, Fourier transformation must be applied[54].

2.7 Gradient Echo and Spin Echo

The order and composition in which gradient pulses are applied during an MR examination is referred to as a pulse sequence. The most common types of pulse sequence are gradient echo (GRE) and spin echo (SE) sequences. Following excitation into the transverse plane, spins undergo dephasing due to local inhomogeneities within the magnetic field. During an SE

experiment, a 180° pulse is applied shortly after the excitation pulse. This causes the rotational direction of the dephasing spins in the transverse plane to become reversed, causing the spins to rephase again. As a result, spins will be in phase again, producing a spin echo exactly after twice the time between the excitation pulse and the 180° refocusing pulse. This is termed the time to echo (TE).

In a GRE sequence, the echo is achieved by intentionally dephasing and rephasing the frequency encoding gradient by application of a positive and negative gradient lobe, as opposed to application of a 180° refocusing pulse. During a GRE sequence only 1 RF pulse is applied, and the echo may be recorded much quicker when compared to a SE acquisition. This generally results in shorter TE and TR values, and this can facilitate rapid signal acquisition. Further to this, the gradient reversal is only capable of refocusing those spins that have been dephased by the application of the gradient itself. Therefore, phase shifts resulting from tissue susceptibility, magnetic field inhomogeneities or chemical shift are not cancelled out as they are during the SE acquisition. Image contrast is therefore dictated by inherent T_2^*.

2.8 MR Spectroscopy

Up until this point the discussion has been focused around the precssion of protons at a specific Larmor frequency which is determine by the gyromagnetic ratio and the strength of the magnetic field. However, it is important to note that the movement of electrons from other nuclei, both around and within a given molecule, has the ability to create an induced magnetic field B_i that changes the local magnetic field B_{loc} that is experienced by the observed nucleus. This may be defined by equation 17:

Equation 17
$$B_{loc} = B_0 - B_i$$

As a result, the observed resonance frequency also changes and can be described by equation 18:

Equation 18
$$\omega_0 = \gamma B_{loc}$$

23

Where ω_0 is the Larmor frequency in Hz, of the observed nucleus and γ is the gyromagnetic ratio. This deviation of the Larmor frequency of the observed nucleus away from the reference resonance frequency is known as the chemical shift. There is a requirement for the chemical shift to be independent of field strength, so that spectra acquired across different spectrometers can be compared. In order to achieve this field independence, chemical shift values (δ) are expressed with respect to the frequency of a reference nucleus proportional to the magnetic field in parts per million, which can be represented by equation 19:

Equation 19

$$\delta = \frac{\omega_0 - \omega_{ref}}{\omega_{ref}} \cdot 10^6$$

Where ω_{ref} is the Larmor frequency of the reference molecule - Trimethylsilyl propanoic acid (TSP) for ^{1}H spectroscopy. The difference in electron distribution between molecules perpetuates different spin environments. These small differences in resonance frequency manifest as chemical shift and allow us to determine structural molecular properties. When a scan is acquired in the absence of magnetic field gradients the resultant FID contains all the frequency components. When the Fourier transform of this FID is taken, the result is a graphical representation of the observed frequencies from the individual nuclei of molecules along the horizontal axis[52].

2.9 Coupling

The presence of two or more neighbouring spins either within the same molecule, or from an adjacent molecule, can cause the spins to perturb each other. Scalar, or J-coupling arises from interactions that take place between via bonding electrons. This coupling gives rise to the splitting of peaks within the spectrum into multiplets. If two half spin nuclei are coupled, then
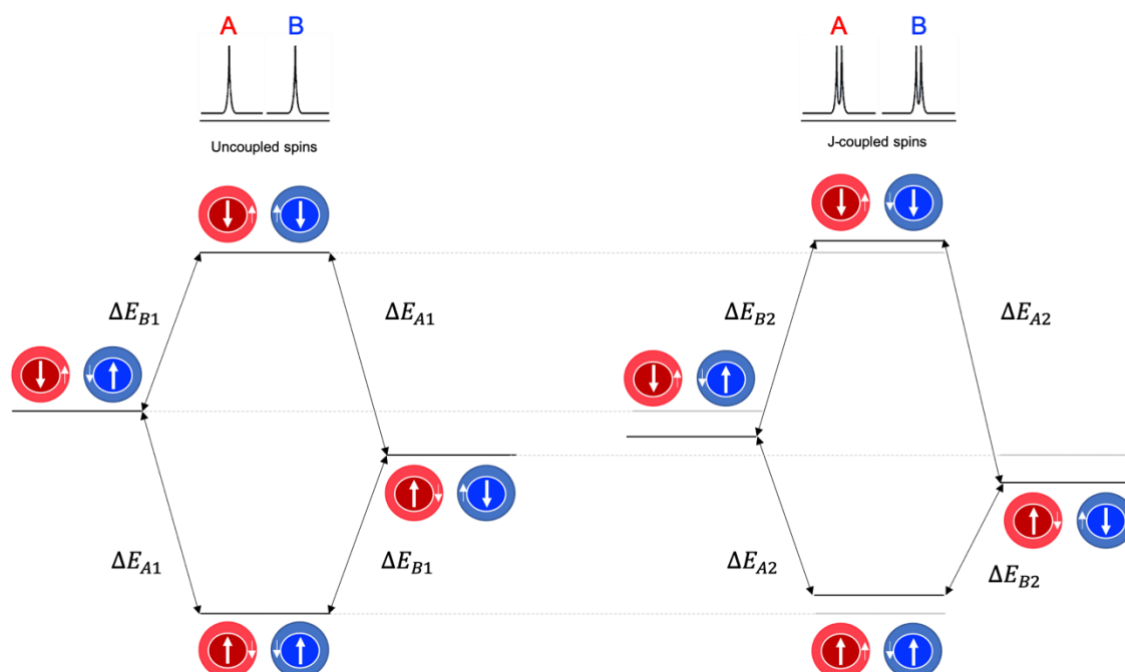
Fig 2.5 An energy diagram of both uncoupled and coupled spin systems, A (red) and B (blue). The large white arrow in the centre represents the nuclear spin state and the smaller white arrow in the periphery represents the electron spin state. Within an uncoupled spin system, the energy difference between the different energy states of A (ΔE_{A1}) are equal, as are the energy differences of the transitions of nucleus B (ΔE_{B1}). The two different energy differences are then reflected as single peaks on the NMR spectrum. When the spin systems are coupled, the nuclear spins of A and B are always aligned anti-parallel to each other. This can force nuclear spins to be parallel with their electron spins, which is energetically less favourable. When both nuclear spins are anti-parallel to each other, and their electron spins are anti-parallel to their nuclear spin, this energy state is more favourable. This results in 4 distinct energy differences, namely $\Delta E_{A1} \neq \Delta E_{A2}$ and $\Delta E_{B1} \neq \Delta E_{B2}$. This coupling results in the spitting of both the A and B peaks.

the resonance from each of the spins split the peak into two symmetrical lines about the chemical shift known as a doublet. *J*-coupling is typically only observed when the nuclei are in close proximity and are not identical. There are two types of coupling, that seen between very similar nuclei, known as homonuclear coupling, and that seen between different nuclei, known as heteronuclear coupling[52].

In order to explain the coupling phenomenon we must first revisit the understanding that spins may reside in two different energetic states, either as a high energy state (+½) otherwise known as spin up (↑), or they may occupy the lower energy state (-½) otherwise known as spin down (↓), where the cumulative difference between these two energy states ΔE gives rise to a single resonance frequency. One simple way of thinking about the two lines seen within a doublet is that they each represent the two possible spin states of the coupled spin, whereby one line represents the up (↑) state and the other represents the down (↓).

Let us now consider two spins, A and B, that are not coupled to each other (figure 2.5). There are four possible energy states that these spins may occupy, these are (A-up B-up ↑↑, A-up B-down ↑↓, A-down B-up ↓↑, A-down B-down ↓↓), and possess the same probability of

occurrence. As the spin states may transition from one to another, there is an energy difference, ΔE_A corresponding with the spin-up to down transition of proton A from ↑↑ to↓↑, which is equal to the energy difference of the spin-down to -up transition (↓↓ to ↑↓). Similarly, there is also a distinct energy difference associated with the transitions of proton B between the different spin states ΔE_B. These two individual energy differences give rise to two separate peaks within the spectrum at the resonant frequencies of A and B (figure 2.5)[52].

When spins are coupled to each other, the interaction between the nuclear spin and the electron spin can affect the energy states that are possible. In quantum mechanics it is stated that when electron spins are coupled, they must always be oriented anti-parallel to each other. A result of this rule is that nuclear spins that are coupled to each other are forced into a parallel orientation along with the electron spins (↑↑ or ↓↓), which in turn results in increased energy levels. When we consider the antiparallel nuclear spins (↑↓ or ↓↑), both the coupled electron spins, and the electron nuclear spins are antiparallel, which is energetically favourable (lower energy). The consequences of this are that the energy differences associated with the spin transitions are no longer equal to each other, and we now have four distinct spin transitions with corresponding energy differences. These four separate energy differences give rise to four individual resonant frequencies that can be observed as two split peaks, or doublets. The magnitude of this splitting and shifting away from the central frequency is typically expressed by the coupling constant J, which is field independent[52].

2.10 ^{1}H-^{13}C 2D Heteronuclear Single Quantum Coherence Coupling (HSQC) NMR

Although 1D NMR experiments are often sufficient for the identification of a wide variety of small molecules, larger, more complex molecules, and other small molecules that are severely overlapped by others, may be significantly harder to resolve from a 1D experiment. 2D NMR allows the supplement of an further set of experimental parameters that in turn result in the addition of a second dimension, which may facilitate ease in interpretation of such spectra.

^{1}H-^{13}C 2D HSQC NMR is a technique that is used to correlate the ^{1}H chemical shift with the ^{13}C chemical shift using the $^1J_{CH}$ coupling. This experiment relies on the transfer of detectable signal from nuclei with a low gyromagnetic ratio, to nuclei with a higher
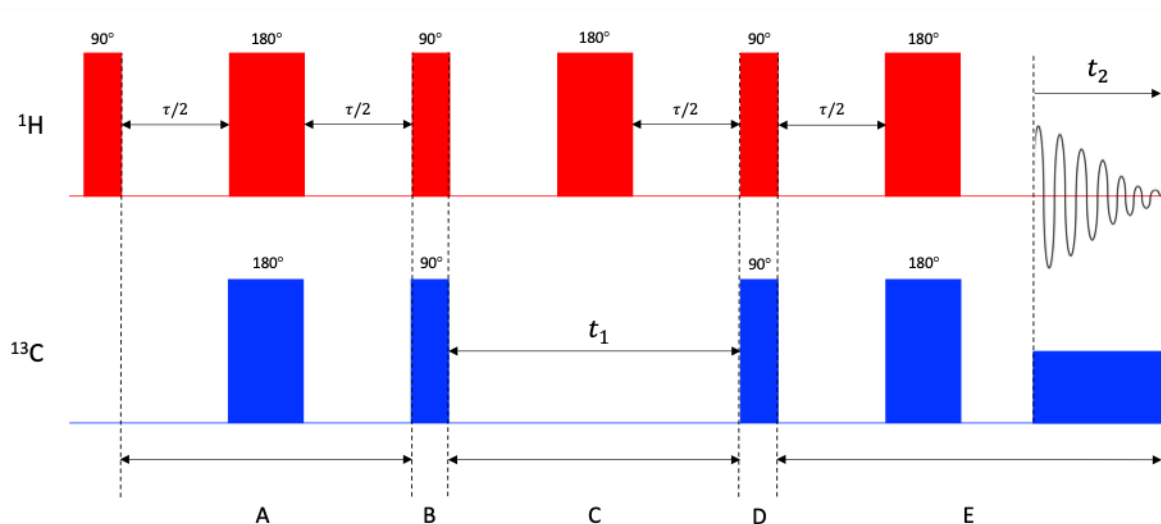
Fig 2.6 The pulse sequence for a typical HSQC experiment. The sequence begins with an equilibrium magnetisation on the ^{1}H which is transferred to the ^{13}C via an IEPT sequence (parts A and B). The ^{13}C spins evolve over time period t_1 (part C) with a centrally placed 180° pulse on the ^{1}H in order refocus the evolution of the coupling. The magnetisation is then transferred back from the ^{13}C to the ^{1}H (part D). The spin echo allows the anti-phase signals to become in-phase and can then be observed using broadband ^{13}C decoupling

gyromagnetic ratio in order to increase the detectable signal. The HSQC experiment achieves this by applying an Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) sequence, allowing a specified evolution time, and then a retro-INEPT sequence in order to transfer the magnetization back onto the ^{1}H[52]. A diagram of the HSQC pulse sequence can be seen in figure 2.6.

The application of the first INEPT sequence involves a 90° pulse on the ^{1}H followed by a 180° refocusing pulse of $\tau/2$ on both the ^{1}H and the ^{13}C channel. Following the 90° pulse and prior to the 180° pulse, an antiphase state is generated on the ^{1}H spins. The 180° pulse allows the transfer of this antiphase state from the ^{1}H to the ^{13}C spins. Subsequent to the magnetisation transfer, following a time period of $\tau/2$, a 90° pulse is applied in order to flip the ^{13}C spins into the transverse plane. This completes the first INEPT sequence, and the ^{13}C spins now evolve for t_1, whilst a centrally placed 180° pulse on the ^{1}H is applied refocusing the evolution of the coupling. Now the retro-INEPT sequence is applied in order to transfer the magnetisation from the ^{13}C back to the ^{1}H. This begins with the application of a 90° pulse on both channels to flip the ^{1}H into the transverse plane, followed by a spin echo period, allowing the antiphase signals to become in phase, followed by detection on the ^{1}H channel[52].

The following chapters will discuss a number of investigations that both directly and indirectly use the techniques described above.

3. Radiomics for the Detection of IDH Mutation

3.0 Chapter synopsis

Radiomics is a rapidly emerging and evolving field of image analysis, and has shown great promise in medical imaging diagnostics. This study explores the use of a radiomics approach in order to classify MR images of glioma into IDH Mut or WT. In order to achieve this a database of patient images that were acquired between 2016-2019 was analysed, and all glioma patients whom underwent IDH investigation, as well as T_1 post-contrast and T_2 imaging were included. All imaging studies that were included in the study underwent a series of normalisation steps in order to account of unwanted variability within the images.

This investigation specifically focuses on the utility of image texture features generated in python using third party software. These mathematically derived features reflect the distribution and interrelationship of the greyscales within the image. Here it is hypothesised that the presence of IDH mutation may induce micro-environmental changes within the imaged tissue that may, in turn, altered the spin environments of the protons that are manipulated during MRI acquisition in order to produce the signal.

The extraction of large amounts of imaging variables inherently facilitates the production of highly dimensional data. This study has explored a number of avenues for the handling of such data, including the use of genetic algorithms (GA) as well as random forests in conjunction with multi-way ANOVA. Although GAs have seldom been explored with respect to medical imaging data, they have been previously shown to perform well when dealing with micro-array data, which is often also plagued with the curse of high dimensionality[59]. Here the novelty of the use of GAs in imaging has been demonstrated, and future investigation may facilitate their more widespread use. Although GAs provide an elegant way of handling data with huge numbers of variables, a more convenient and less computationally demanding was required here, forming the rationale for the use of random forests. Although these popular classifiers allow for both feature selection and classification, it was shown that metadata associated with the acquisition of the images was introducing a large amount of unwanted variability, so further feature selection steps incorporating principal component analysis and multi-way ANOVA were used.

The results of this work may aid in the building body of evidence for the non-invasive classification of MR images into IDH Mut and WT.

3.1 Introduction

3.1.1 Radiomics

A biomarker may be defined as "a characteristic that is objectively measured and evaluated as an indicator of a normal biological process, pathological development or pharmacological response to therapeutic intervention"[60]. Biomarkers can be measured or obtained from a number of sources including tissue biopsy, blood, urine or imaging. The latter source provides both qualitative and quantitative information pertaining to a tissue, process or disease that one wishes to probe whilst remaining entirely non-invasive. Qualitative information requires interpretation by an expert clinician following years of training and is often subjected to fluctuating degrees of interobserver variability, whereas quantitative biomarkers are established on mathematical definitions that are standardised and transferrable.

The role of medical imaging in the assessment and management of cancer has evolved extensively over the past decade from a primary diagnostic tool into a central role in the context of individualised cancer treatment[61]. The use of investigatory methods such as genomics, metabolomics and proteomics have been the primary focus of personalised treatment strategies for many years[62]. However, the true spatial and temporal heterogeneity of tumours that is facilitated by local variations in factors such as oxygenation, blood supply, gene expression profiles, subpopulation evolution and metabolism are both common tumoral features, as well as integral to the treatment and management of the pathology[63–65]. Biopsy samples obtained through invasive and often risk-prone surgical procedures, taken from localised regions of the tumour, fail to reflect the inherent heterogenous landscape of the lesion[66]. Conversely, medical imaging is able to sample the entire pathology, non-invasively, multiple times over. A number of studies have postulated that genetic and molecular characteristics may be echoed in phenotypic manifestations that can be captured via medical imaging[67–69].

The exponential growth in computing power and recent developments in machine learning techniques have allowed for the emergence of new processes for the extraction and analysis of quantitative features arising from the conversion of medical images into mineable datasets[70]. The term radiomics was first employed in 2010, when quantitative imaging findings were used to characterise the phenotype and environment of the tumour, reflecting gene expression[71]. A number of radiomics studies immediately followed, some of which highlighted

that the reliability and reproducibility of the calculated features was heavily dependent on a number of factors including image acquisition, reconstruction, pre-processing and segmentation[72–76]. The framework for this -omics based approach to imaging science was explicitly outlined by Lambin et al in 2012, and has since been utilised in almost all areas of medical imaging[61].

The utility of radiomics in contributing to predictive and prognostic oncological models depends strictly on a basic 3 step workflow, namely 1. Image acquisition 2. Computation of radiomic features 3. Statistical analysis and machine learning. To fully extend any proposed model into a clinical setting, a prospective multicentre study would ideally be employed as a fourth and final step. Radiomics is offered as a supportive tool in clinical decision making, providing auxiliary evidence alongside other patient characteristics upon which diagnosis and prognosis can be suggested. Although the technique has been applied to a variety of different conditions, both within and outside of medical imaging, its application in oncology is probably the best described.

The mining of radiomic data in pursuit of correlation with genomic profile and mutational status has gained much traction since the advent of radiomics The value of this application is rooted by the fact that almost every cancer patient will undergo some form of imaging study during their care, and in fact most will have multiple imaging studies across several timepoints to monitor and track disease progression. However, not all of these patients will undergo genomic profiling. Moreover, when tissue samples are obtained for profiling, they are gathered from a distinct area of the tumour at a single timepoint, subjecting such investigations to sampling error.

3.1.2 Methodology and Implementation

Quantitative features of an image may be referred to as image texture, which is a reflection of pixel intensities, how they are distributed and the interrelationship between neighbouring groups of pixels. In this study the rationale for investigating if texture features are able to discern IDH mutation using the radiomics workflow stems from the fact that the mutation facilitates intracellular changes that are detectable via magnetic resonance spectroscopy (MRS). If those changes in cellular environment are detectable via MRS, then the same intracellular changes may invoke both macro and microscopic changes within images acquired of the same tumour. The differences within the images acquired may give rise to alterations in

image texture when comparing Mut tumours with WT. Eventually, by combining both MRS and MRI findings, there would be a building body of evidence to non-invasively detect the presence of IDH mutation within clinical decision making. Texture analysis has the potential to measure whole tumour heterogeneity by correlating mathematically derived parameters with clinical outcomes in order to predict tumour biology[77,78].

There is a requirement for the procurement of data to be used within the radiomics workflow to be acquired under stable conditions in a standardised manner. Clinical MRI data, however, is subject to a wide range of variabilities. The exponential advancement of the technology has led to an enormous range of both hardware and software for the acquisition and reconstruction of images being used in institutions worldwide. Radiomic features are subject to influence from parameters such as magnet strength, pulse sequence variation, signal-to-noise ratio and spatial resolution. There is some debate over the robustness of radiomic features and the models generated from them due to the presence of acquisition induced variability. Conversely, some studies, such as that conducted by Mayerhoefer et al, suggest that variations in acquisition protocols such as repetition time, echo time and sampling bandwidth have little-to-no effect on the features generated[76]. All medical images, including MR images, must undergo pre-processing prior to feature extraction. Pre-processing steps include image segmentation, interpolation, intensity normalisation, bias field correction, grey level quantization and filtration[79].

The radiomics workflow has been outlined in figure 3.1 and consists of a number of pre-processing and feature calculation steps. All steps involved within the workflow can be carried out in commercial, open source or locally constructed software. Here we have used only opensource software that is freely available to download for any potential user. This is to hopefully facilitate the reproducibility of the study, so that any other centre is available to recapitulate the investigations free of charge. One of the first steps is image segmentation, which involves drawing a region of interest (ROI) either manually, automatically or semi-automatically. It appears that features calculated from 3D volume of interests (VOI) are more reliable than those that are determined from 2D single slice areas of an image[80]. Furthermore, if the full potential of the radiomics workflow to reflect tumour heterogeneity is to be achieved, it would stand to reason that a 3D VOI would be more representative of the overall pathology.

To improve image matrix resolution and deal with potential mismatch of image acquisition parameters, image interpolation is applied to remap the original matrix onto an

isotropic grid spacing. For the generated texture features to be rotationally invariant, and so that images from different samples are comparable between cohorts, samples and batches, isotropic spacing is essential[81]. Texture features are sensitive to changes in voxel size, therefore image interpolation has the potential to heavily influence the end result of any radiomics investigation[82]. To enable reproducibility, consistent isotropic voxel spacing across all measures within a study is imperative. There is no clear consensus as to whether up-sampling or down-sampling is preferable in any given case, although there have been some studies, such as those conducted by Mayerhoefer et al that suggest up-sampling by transforming the original image into a higher matrix size can expedite improved texture classification[83].

Following interpolation, image grey levels must undergo discretisation to reduce the intensity scale to make feature calculation tractable[84]. There are two distinct methods for image discretisation, namely fixed bin number and fixed bin size[81]. The optimal method of discretisation that yields the least inter- and intra-observer variability is modality dependant. It should be noted however, that only the fixed bin number method is suitable for images consisting of arbitrary units, such as those that make up most conventional MR images. The details of this method and its implementation will be explored later in this chapter.

Feature generation can encompass either or both agnostic or semantic texture features to be used in the radiomics workflow. Semantic features include location, shape, vascularity and necrosis, all of which may be quantified via a scoring system, however these are still subject to varying levels of inter-observer variability. Agnostic features are mathematically derived from assessment of the distribution, inter-relationship, and local variation of grey scales within the segmented region of the image. Hundreds, if not thousands of features may be generated per image using different combinations of filters and grey level matrices. One significant issue when dealing with such datasets, whereby the number of features to be analysed is greater than the number of samples is data dimensionality. Highly dimensional data is often subject to overfitting and inherently contains large amounts of redundant data. To overcome this, the use of classifier models such as support vector machine, regression models and random forests are used to reduce false discovery rate and to select the most discriminative features. We have explored a number of classifier options to probe the data for texture features that may be able to discern IDH Mut and WT from standard MR imaging.

Radiomics uses the agnostic features that reflect grey level value, interrelationship, and spatial distribution[79]. First-order statistics in radiomics originates from the histogram

projection of grey level intensities from within the segmented region of the image. Values that are calculated include the mean, median, standard deviation, skewness, kurtosis and entropy of the histogram. The mean is the average intensity of the image histogram and can be reflected in the brightness of the image. Standard deviation is the spread of intensities about the mean, whereby higher values indicate a large range of intensity values within the segmented ROI and may reflect increased heterogeneity of the underlying tissue within the ROI. Skewness is representative of the asymmetry of the histogram, and kurtosis is reflective of the magnitude of pixel distribution, whereby a more positive kurtosis indicates that the height of the histogram is taller than a normal gaussian distribution. Entropy depicts the irregularity of intensity distribution within the segmented region, where a high entropy value represents a heterogeneous distribution and may be indicative of a heterogenous tumour environment.

Second-order statistics look at examining the spatial relationship of intensity values within the segmented region by construction of grey level dependency matrices[70,85]. The concept of dependence matrices was first investigated with the advent of the grey level cooccurrence matrix (GLCM) by Haralick et al[86]. The GLCM explores the combinations of grey level intensities from voxels that make up the 3D volume in all directions. Over time the development of further dependence matrices has allowed for other intra-regional intensity relationships to be investigated. The grey level run length matrix (GLRLM) examines the number of consecutive voxels with the same intensity value, and the grey level zone length matrix (GLZLM) allows the quantification of groups of voxels with given values to assess the homogeneity of an area. Further to both first-order and second-order statistics, are higher-order statistics that come from the application of filters or transforms, prior to feature extraction. The details surrounding the calculation of texture features and the various statistic levels will be discussed in further sections of this chapter in more detail.

3.1.3 Radiogenomics in Glioma

Radiomics refers to the general field in which patient images are converted to large mineable quantitative datasets that may be leveraged in pursuit of decoding tumour phenotype for clinical diagnostic and prognostic purposes. Radiogenomics refers to the application of radiomics by which imaging features are correlated with gene expression profiles/molecular phenotypes. The genetic profile of a tumour is used clinically to determine prognosis, diagnosis

and treatment strategy, as the genomic instability of a lesion represents one of the hallmarks of cancer. It therefore stands to reason that there has been a recent surge in popularity of radiogenomic studies that involve the application of the relatively new and rapidly evolving field of radiomics correlating with genetic profiles.

Despite the basic outline, and the great promise offered in contributing towards clinical decision making, translation of radiogenomics into the clinic has been somewhat restricted by its own practice. The lack of reproducibility in a number of recent studies has highlighted that there is a clear unmet need to standardise the process, so that this emerging technology can eventually become incorporated into routine clinical practice. Computation of radiomic features is a convoluted and complex procedure, incorporating a great number of steps (figure 2.1), which if not fully conveyed when published, can result in the misinterpretation and inaccurate reporting of both methodologies and results. Other key limitations accompanying a large number of radiomics-based studies include incorporation of high false positive rate, limited patient numbers and overfitting of proposed models.

As a result of these shortcomings the image biomarker standardisation initiative (IBSI) was formed. The IBSI is an independent international collaboration with the aim of standardising the computation, extraction and usage of imaging biomarkers with the outlook of faster and more relevant clinical translation of quantitative imaging markers identified by high-throughput image analysis[81].

3.1.4 Scope of Study

This study was set out to investigate the utility of radiomics in the classification of MR images into IDH Mut or WT based on imaging texture features for diagnostic purposes. Several methods have been employed in order to examine the possibility of constructing a translatable model in order to aid in the non-invasive determination of the presence of IDH mutation. The use of standard MR imaging alone (T_1 and T_2 -weighted) forms an important role in the translatable capacity of any potential models, due to the ubiquitous use of such sequences in glioma diagnostics.

3.2 Methods

3.2.1 Data Collection

This retrospective study was approved by the Walton Centre Research Tissue Bank research ethics committee (REC) approval (Reference number 15/WA/0385). All patients whose images have been included in this study submitted informed consent to be included in the study. Patients were originally identified for this study based on completion of IDH pathology test. All patients included in the study had been imaged between 2016 and 2019. IDH mutational status was confirmed by either histopathology or sequencing, or a combination of the two. The final pathological diagnosis, as given by neuropathology consultant, was taken as the mutational status for each patient included in the study. 486 patients were first identified, of which 303 had 3D T_1 contrast enhanced spoiled gradient echo imaging studies with repetition times (TR) 8.1-9ms and echo times (TE) 1.4-3.1ms. Of these, 92 imaging studies were discounted due to poor image quality from motion and high levels of susceptibility artefact. This left a total of 211 3D T_1 contrast imaging data sets to be included in the investigation, of which 152 patients carried WT IDH, and 59 carried Mut IDH. Of the 486 patients, 208 Axial T_2 imaging data sets were selected whereby the data was acquired using a spin echo sequence TRs 3000-6680ms and TEs 80-103ms. Of these axial T_2 studies, 52 were discounted, again due to poor image quality due to the inclusion of imaging artefacts, leaving a total of 156 axial T_2 imaging studies including 62 IDH Mut and 94 IDH WT patients. All images were anonymised and exported as DICOM images before being converted to NifTi format using MRICron® software. All image acquisition parameters and patient demographics can be found in appendix 1.

3.2.2 Pre-processing

A number of pre-processing steps preceding radiomic texture feature extraction were optimised for this investigation. Pre-processing steps were carried out using a combination of Advanced Normalisation Tools (ANTs)[87] toolbox, 3D slicer[88] and Pyradiomics[89]. An outline of the steps followed in the processing of images and extraction of features can be found in figure 3.1. The first pre-processing step was to correct for intensity inhomogeneity within the images. Intensity inhomogeneity, bias or gain field is a potential confounder in any MR image analysis[90]. The term refers to the low frequency, non-anatomical variation in signal intensity of the same tissue class over the image domain, and its presence can heavily impede on quantitative image
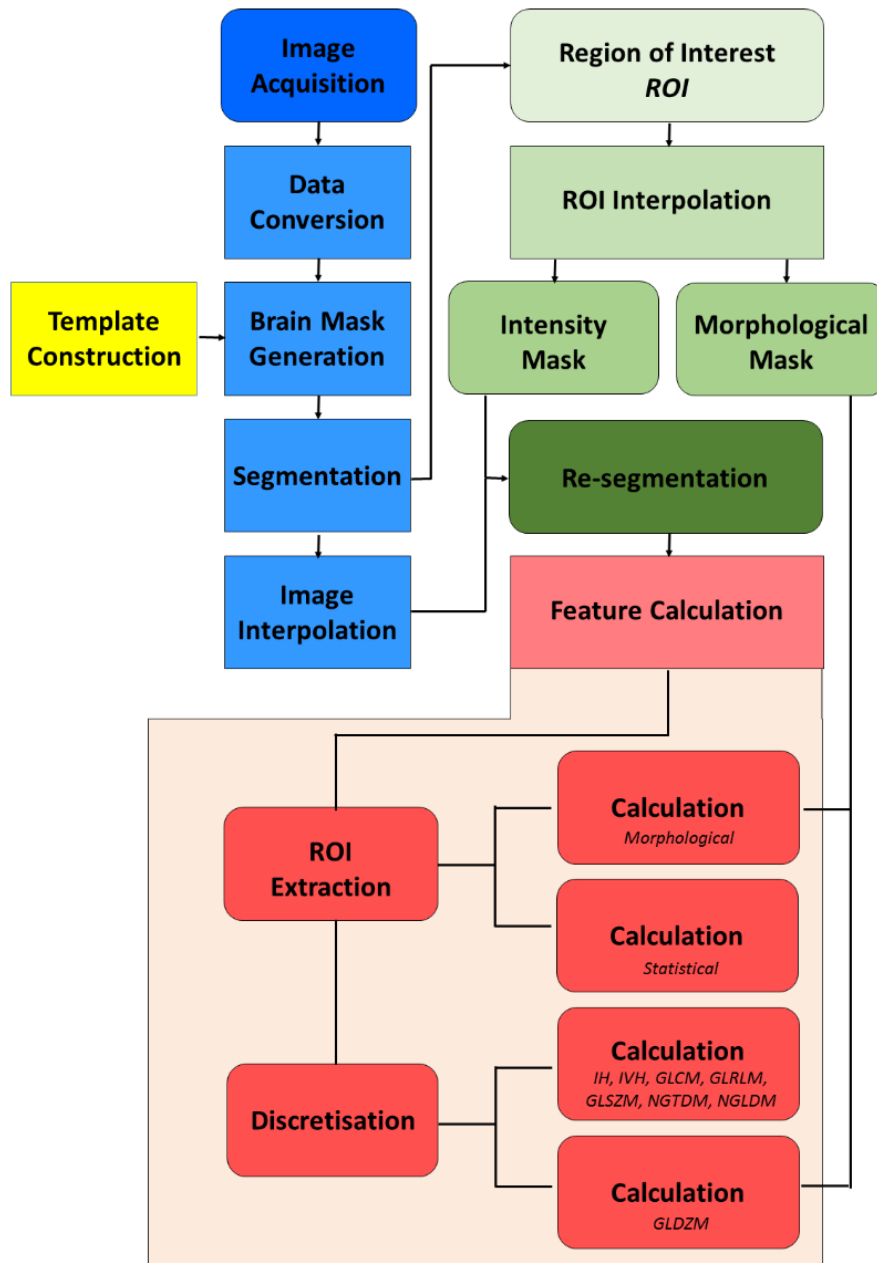
Fig 3.1 A scheme to demonstrate the steps taken to achieve image feature calculation whilst following the IBSI guidelines for image biomarker identification. Boxes in blue describe the number of pre-processing steps to be carried out across the entire raw data image stack. Boxes in green indicate pre-processing steps prior to feature calculation stipulated by the IBSI. Following image segmentation (into both whole brain and tumour volume) this ROI may then be processed into two entities, intensity mask and morphological mask. This is only required for absolute units, since this study is concerned with the arbitrary units of MRI, only the morphological mask is required. Both the image and the morphological mask must be interpolated to produce rotational invariance. Feature calculation can be seen in the red boxes and classified into a number of families: IH: intensity histogram; IVH: intensity volume histogram; GLCM: grey level co-occurrence matrix; GLRLM: grey level run length matrix; GLSZM: grey level size zone matrix; NGTDM neighbourhood grey tone difference matrix; NGLDM: neighbourhood grey level dependence matrix; GLDZM: grey level distance zone matrix. Adapted from Zwanenburg et al.[81]

biomarker calculation and extraction[91]. For this study the N4BiasFieldCorrection algorithm within the ANTs toolbox was used to correct for this potential source of error[92]. The N4BiasFieldCorrection uses a B-spline approximator that allows greater control at magnetic field strength of 3T and above than its predecessors, permitting the specification of a mask to allow smaller control point spacing. The advantages of these features include multiresolution approximation strategy which employs hierarchical fitting of sequentially higher levels of frequency variation of the bias field, and an iterative incremental update of the bias field[92]. To implement these features and achieve the optimal intensity inhomogeneity correction output, a region mask must be generated prior to application of the N4 correction algorithm.

To generate a binary brain mask the AntsBrainExtraction.sh wrapper function, part of the ANTs toolbox, was used[87]. This function requires an atlas, an intensity mask and an optional extraction mask along with the input image, to carry out parameterised B-spline registration step of the extraction process. There are a number of publicly available datasets that exist to generate such atlases using the BuildTemplateParallel.sh function within the ANTs toolbox. A number of templates were constructed from publicly available datasets to both generate individual whole brain binary masks for each patient, as well as accurately removing the skull and meninges from around the brain. Multi-subject templates are designed to possess features that are representative of an entire population, therefore the higher the number of datasets contributing to the creation of a template, the more likely such a template will possess mutual information relating to the subject dataset that the template is being warped to. Initial experiments were carried out using the LPBA40 dataset[93] to create an atlas for use in the ANTs brain extraction algorithm. Further optimisation steps used both the Kirkby[94] and NKI[95] datasets, consisting of 21 and 10 subject respectively, to create suitable atlases. The template that was found to produce the optimal brain extraction result and therefore the most accurate binary brain mask was created from the IXI dataset[96]. This dataset contained T_1 pre- and post-contrast images along with T_2 images from 600 patients. T_2 FLAIR templates were downloaded from the brainder database[97]. To assess the performance of each of the template on the ability to delineate the brain region from the skull on each of the images, the extracted images from the BrainExtraction.sh wrapper function used to create the binary brain mask were visually inspected.

Following construction of the optimal template, intensity mask and extraction mask, each patient dataset was corrected for intensity inhomogeneity. To ensure that each patient
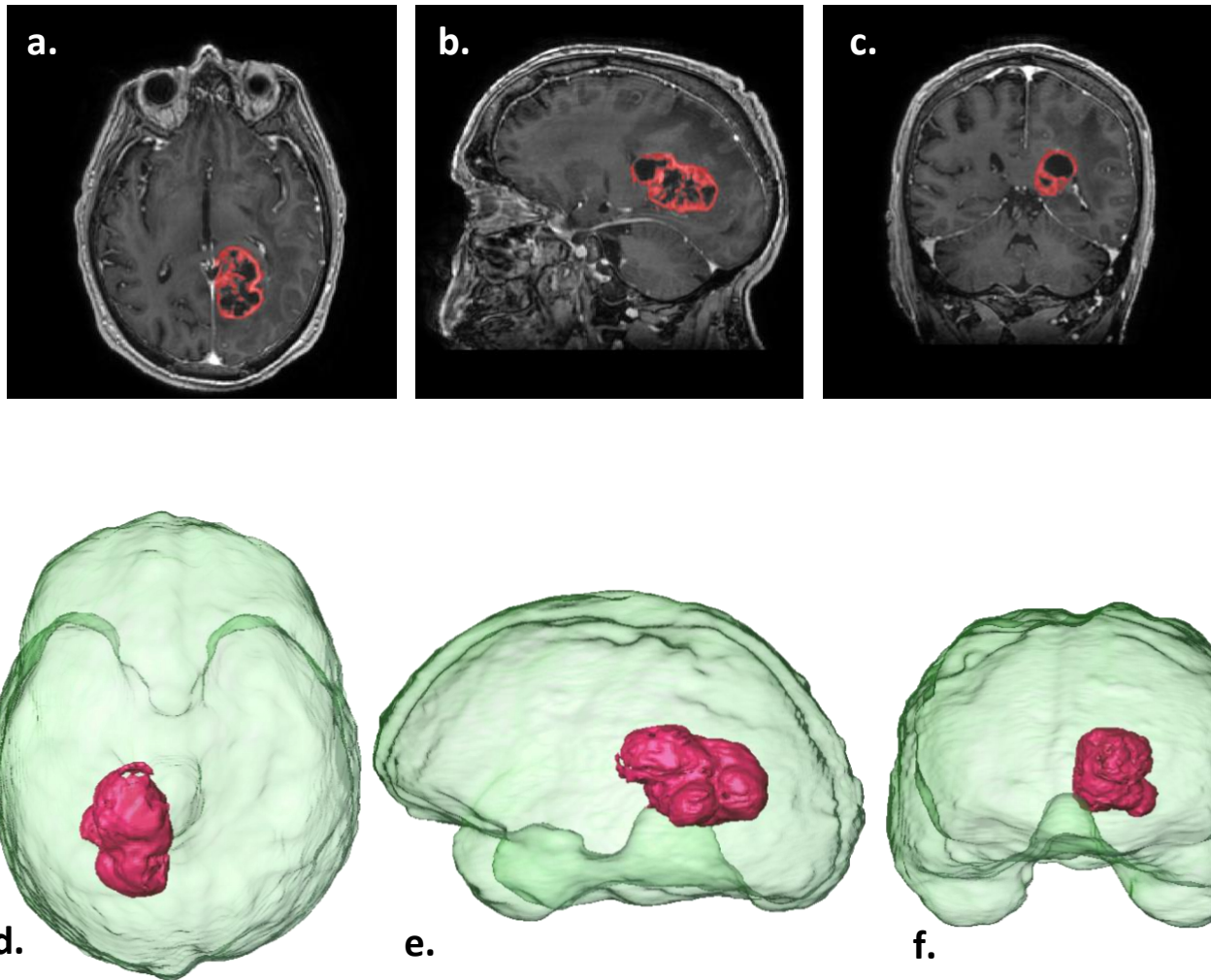
Fig 3.2 Example of tumour segmentation result. Only the enhancing region of the tumour was included in the segmentation, as this is the active part of the tumour, and will provide the least amount of inter-patient variability. The central region of the tumour was not included in the segmentation for feature calculation as it can contain varying amounts of active tumour and necrotic regions. Inclusion of this region is a potential source of bias when considering statistical analysis. **(a)** GBM patient axial view of segmented tumour whereby the enhancing region has been included and all other parts negated. **(b)** sagittal view of segmented tumour. **(c)** Coronal view of the segmented tumour **(d)** Superior view 3D rendering of the extracted brain (transparent green) with a 3D rendering of the segmentation result *in situ* (solid pink). **(e)** lateral view of the 3D tumour and brain. **(f)** Anterior view of the 3D tumour and brain.

dataset and its mask were in the same physical space, the header information from the patient image was copied onto the mask image. The parameters used for the N4 algorithm were: number of iterations 100x100x100x100, convergence factor 0.00000001 and spline distance 200. The rationale for the conduction of the intensity inhomogeneity correction prior to brain extraction and removal of the skull and meninges is such that the subsequent registration steps involved in the brain extraction algorithm use a combination of rigid, affine and deformable B-

spline registration models to map the atlas image onto the patient data. If low frequency field inhomogeneity were to persist throughout the image during these steps, the registration would be sub-optimal as the brain extraction algorithm presumes a given grey level intensity range for a certain tissue class.

For this step the previously constructed extraction template, built using the IXI dataset was used. The brain probability mask, also generated using the antsBuildTemplateParallel.sh function was specified. In addition, a brain extraction registration mask was specified to limit the metric computation to within a specific region. All other parameters were kept as default.

Subsequent to brain extraction, a region of interest in which radiomic features were to be calculated was generated. Here we looked at two regions of interest, both the whole brain, for which the brain mask generated as an output from the previous brain extraction step was used, as well as the enhancing region of the tumour. For this study we chose to include only the enhancing region of the tumour, as seen on T_1 post contrast images, as this region represents the most aggressive area of the disease[98]. Furthermore, calculation of radiomic features from within this region alone should allow the most homogenous sampling, as the inherent heterogeneity of gliomas means that inclusion of the entire tumour will introduce varying proportions of necrotic regions, proliferative regions and invasive borders. In a further attempt to adhere to the previously described image standardisation methods, and to reduce variability and eliminate bias from the application of statistical methods, only the enhancing region of the tumour was used. To segment the enhancing region of the tumour on all T_1 post contrast images, a semi-automated approach was adopted. For tumour segmentation 3D slicer was utilised along with the segmentation wizard module[99]. For this process an initial ROI encompassing the entire tumour volume was delineated using the model marker placement tool. Next a thresholding range was determined to eliminate non-enhancing voxels. The resultant segmentation was then reviewed and edited using the paint tool to exclude any blood vessels or residual meningeal tissue that remained from brain extraction, that may have been included within the enhancing threshold range. A number of T_1 post-contrast images were excluded at this point, as the tumour was too small and close to the skull and not distinguishable from meninges, or the enhancing region was too small to calculate texture features across. A total of 131 IDH WT tumours were included, along with 24 Mut tumours, giving a total of 155 patient images to be included in the intra-tumoral analysis. An example of a segmented T_1 post contrast tumour region can be seen in figure 3.2

Once the region for quantitative feature generation had been determined, the pre-processing steps required to normalise the images with respect to intensity distribution and spatial division ahead of feature calculation can be carried out. For this process a .yaml text file was written to be used as an argument in the command line implementation of pyradiomics, the software of choice for feature calculation[100]. Image texture features require isotropic voxel spacing so that they are rotationally invariant and to facilitate evaluation of images derived from different patients and cohorts[81]. A large number of imaging texture features are susceptible to alterations in voxel size, therefore image interpolation can have a significant effect on the texture features generated. To facilitate reproducibility, it is crucial to maintain isotropic voxel spacing. Interpolation algorithms aim to translate image intensities from the original image grid to a revised interpolation grid. When considering such grids, it is important to remember that voxels are spatially represented by their centre. Here we used ITK B-spline interpolator due to the computational efficiency and increased accuracy of point estimation using the generation of third order polynomials derived from large neighbourhoods of intensity values at every voxel centre. The use of this interpolation method can incur out-of-range intensities in acute intensity transitions, however it is also likely to enhance the preservation of tissue contrast differences. The resampled pixel spacing was set at 1.0x1.0x1.0 for T_1 post contrast images and 2.97x2.97x2.97 for T_2 images. These spacings were determined by calculating the mean voxel spacing in the x and y plane from a sample of 50 representative images included within each cohort.

To render the calculation of image features tractable, image discretisation is required[101]. There are two potential approaches to discretisation, one involves the discretisation to a fixed number of bins and the other to a fixed bin width. However, the fixed bin number method is not recommended when dealing with arbitrary image intensities such as in MRI, as this method maintains a direct relationship between the original image scale, which in MR images can vary immensely between different patient, scanners, protocols and centres. For this reason, the fixed bin number method was be used, whereby the grey level intensities X_{gl} are discretised to N_g a fixed number of bins. The following equation describes the process:

$$X_{d,k} = \begin{cases} N_g & X_{gl,k} = X_{gl,min} \\ \left\lceil N_g \dfrac{X_{gl,k} - X_{gl,min}}{X_{gl,max} - X_{gl,min}} \right\rceil + 1 & X_{gl,k} > X_{gl,min} \end{cases}$$

Let $X_{gl,k}$ be the intensity of voxel k that has been corrected to the lowest occurring intensity value $X_{gl,min}$ observed within the ROI, divided by the bin width $X_{gl,max} - X_{gl,min}/N_g$ and rounded up to the nearest integer. This method allowed us to impart a normalisation effect on the data, which is very useful when handling arbitrary units in MRI where an image intensity range can be enormous. It also allows us to maintain contrast within the image. Other variations of intensity discretisation have been described and used, such as intensity histogram equalisation[102], however this method alters contrast and may impact image texture. Here we used 64 bins to discretise the images.

Other additional setting parameters were also used to standardise the images ahead of feature extraction. The 'correct mask' parameter was selected, which uses a nearest neighbour interpolator to resample the mask to the image geometry. As a Laplacian of Gaussian filter was applied, a pad distance of 3 was selected. Filters were also applied to images for additional feature extraction. Laplacian of Gaussian filter was applied for edge enhancement, with sigma vales of 1.0 and 3.0. The lower sigma value enhances finer features within the image, and the greater sigma value will enhance the more coarse features. A wavelet bandpass filter with a bin width of 10 was also applied, yielding 8 decompositions per level, which is equal to all possible combinations of applying either a high or low pass filter in each of the three dimensions. Pyradiomics utilises the parameter file containing the specifications for pre-processing settings as an argument during the input of each image. The above listed settings were applied to each image immediately prior to feature extraction.

3.2.3 Feature Calculation

Feature calculation is the final image processing step to generate quantitative data that can be assessed using machine learning. This step is where feature descriptors are used to quantify attributes within the defined ROI. Here we generated features from the following classes:

1. First order statistics (19 features)
2. Grey level cooccurrence matrix (24 features)
3. Grey level run length matrix (16 features)

4.   Grey level size zone matrix (16 features)

5.   Grey level dependence matrix (14 features)

All feature calculation was carried out in Pyradiomics[89]. All of the feature classes were calculated on both the original images, as well as all images derived from application of the aforementioned filters. All matrices from which the texture features are calculated are both translationally and rotationally invariant.

Here we chose to pursue texture features in order to classify the images into IDH Mut or WT as these are mathematically defined, well reported and easy to calculate with third party software such as PyRadiomics[100]. Further to this, the different filters that were applied were chosen due to their ability to enhance and supress different clusters of greyscale groups within the images.

3.2.3.1 First order statistics

First order statistics describe how grey levels within the specified ROI are distributed.

3.2.3.2 Grey level cooccurrence matrix

The grey level co-occurrence matrix (GLCM) is a matrix that is able to quantify the relationship that discretised grey level intensity values of neighbouring voxels by expressing how their combinations are distributed along a given direction. When we are considering a 3D volume, such as in this study, the GLCM is a neighbourhood that consists of 26 connected voxels in all three dimensions. Within the 3D volumes in this study, there are therefore 13 individual direction vectors within such a neighbourhood when we consider the voxel distance as $\delta = 1$. These vectors are: (0, 0, 1), (0, 1, 0), (1, 0, 0), (0, 1, 1), (0, 1, −1), (1, 0, 1), (1, 0, −1), (1, 1, 0), (1, −1, 0), (1, 1, 1), (1, 1, −1), (1, −1, 1) and (1, −1, −1).

Once the above spatial relationship has been defined, we calculate the GLCM. Let $\mathbf{M_X}$ be the N_g x N_g co-occurrence matrix, where N_g is the number of discretised grey levels seen within the ROI intensity mask, and $\mathbf{x}$ is a specific direction vector. Element (i, j) of the matrix describes the frequency at which the arrangements of grey levels i and j arise in neighbouring voxels along the direction $\mathbf{x_+} = \mathbf{x}$ and along the direction $\mathbf{x_-} = -\mathbf{x}$. Therefore we have $\mathbf{M_X} = \mathbf{M_{X+}} + \mathbf{M_{X-}} = \mathbf{M_{X+}} + \mathbf{M_{x+}^T}$ [103]. The following two dimensional example demonstrates how GLCM for $\mathbf{M_{X+}} = \rightarrow$ and $\mathbf{M_{X-}} = \leftarrow$ are calculated:

$$I = \begin{bmatrix} 1 & 2 & 2 & 3 \\ 1 & 2 & 3 & 3 \\ 4 & 2 & 4 & 1 \\ 4 & 1 & 2 & 3 \end{bmatrix}$$

The GLCM for $\mathbf{M_{X+}} = \rightarrow$ then beomes:

$$P = \begin{bmatrix} 0 & 3 & 0 & 0 \\ 0 & 1 & 3 & 1 \\ 0 & 0 & 1 & 0 \\ 2 & 1 & 0 & 0 \end{bmatrix}$$

And the GLCM for $\mathbf{M_{X-}} = \leftarrow$ becomes:

$$P = \begin{bmatrix} 0 & 0 & 0 & 2 \\ 3 & 1 & 0 & 1 \\ 0 & 3 & 1 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

To achieve rotational invariance, the GLCM texture feature values are calculated following aggregation of information from the different matrices. There are a total of 24 different texture features that are calculable from the GLCM in accordance to the IBSI guidelines[81].

3.2.3.3 Grey level run length matrix

Grey level run length based features (GLRLM) were introduced by Galloway[104] and are an additional image texture feature to asses homogeneous runs of grey level intensities throughout a discretised image stack. A run length is defined as a number of consecutive voxels with the same intensity value along a given direction $\mathbf{x}$ as previously defined above. In the GLRLM of $\mathbf{P}(i, j|\mathbf{x})$, the $(i, j)^{th}$ element describes the number of occurrences of runs with the length j for the discretised value i. The two dimensional example below demonstrates the principle using a 5x5 matrix with 5 discrete grey levels:

$$I = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix}$$

The GLRLM where the given vector direction $\mathbf{x} = 0$ then becomes:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 3 & 0 & 1 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 3 & 0 & 0 & 0 & 0 \end{bmatrix}$$

If we consider the above matrix, $\mathbf{M_X}$ is the $N_g \times N_r$ GLRLM, in which N_g is the total number of discritised grey level intenisites present within the defined ROI, and N_r is the longest possible run length direction along vector $\mathbf{x}$. The matrix element $\mathbf{P}(i,j|\mathbf{x})$ of the GLRLM is the occurrence of the given grey level i of run length j. $N_z(\mathbf{x})$ is the total number of runs along the given direction $\mathbf{x}$ which is equal to:

$$\sum_{i=1}^{N_g} \sum_{i=1}^{N_r} \mathbf{P}(i,j|\mathbf{x})$$

A total of 16 high order statistic texture features can be calculated from the GLRLM when following the IBSI guidelines[81]

3.2.3.4 Grey level size zone matrix

In addition to the GLCM and GLRLM, texture feature may also be calculated from the grey level size zone matrix (GLSZM). This matrix allows quantification of groups, or zones of grey levels. A zone may be defined as a group, or neighbourhood of connected voxels that share the same grey level intensity[105]. Whether a voxel lies within a given neighbourhood is dictated by its connectedness. Here we are looking for voxels connected to all 26 neighbouring voxels in all 3 dimensions. By definition the GLSZM is always rotationally independent, whereby one matrix is calculated for all directions within the ROI. Consider the two dimensional 5x5 matrix below:

$$\mathbf{I} = \begin{bmatrix} 5 & 2 & 5 & 4 & 4 \\ 3 & 3 & 3 & 1 & 3 \\ 2 & 1 & 1 & 1 & 3 \\ 4 & 2 & 2 & 2 & 3 \\ 3 & 5 & 3 & 3 & 2 \end{bmatrix}$$

From this the GLSZM calculated is:

$$
\mathbf{P} = \begin{bmatrix}
0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 \\
3 & 0 & 0 & 0 & 0
\end{bmatrix}
$$

There are a total of 16 features calculable from the GLSZM that agree with the IBSI guidelines[81].

3.2.3.5 Grey level dependence matrix

The GLDM is focussed around the concept of a neighbourhood of connected voxels within a specified distance δ that are dependent on a central voxel. A neighbouring voxel j is deemed to be dependent on the central voxel with the grey level i if $|i - j| \leq a$. In a grey level dependent matrix $\mathbf{P}(i,j)$ the $(i,j)^{th}$ element describes the number of times a voxel with the grey level i and j dependent voxels in its neighbourhood appears in image. Consider the 5x5 matrix below:

$$
\mathbf{I} = \begin{bmatrix}
5 & 2 & 5 & 4 & 4 \\
3 & 3 & 3 & 1 & 3 \\
2 & 1 & 1 & 1 & 3 \\
4 & 2 & 2 & 2 & 3 \\
3 & 5 & 3 & 3 & 2
\end{bmatrix}
$$

From this the GLDM is calculated as follows:

$$
\mathbf{P} = \begin{bmatrix}
0 & 1 & 2 & 2 \\
1 & 2 & 3 & 0 \\
1 & 4 & 4 & 0 \\
1 & 2 & 0 & 0 \\
3 & 0 & 0 & 0
\end{bmatrix}
$$

There are a total of 14 features that may be derived from the GLDM in accordance to the IBSI guidelines[81].

3.2.4 Feature Selection Using GALGO Genetic Algorithm

The application of filters and the utilization of all calculatable features from the above four matrices, along with the inclusion of first order features, resulted in the generation of 973 variables for each patient image. As this number supersedes the number of participants within the study (211 T_1 and 156 T_2 datasets), there is a requirement to reduce the data dimensionality to eliminate redundant features that may not contribute to the molecular classification of tumours. Here we have used a software package GALGO, which is implemented in the statistical programming environment R and utilises genetic algorithms (GA) to solve variable selection problems.. Although this group of algorithms has seldom been explored within the context of medical image analysis, especially in the field of radiomics, they have shown to be effective when applied to highly dimensional micro array data, and the dimensionality of the dataset in this study shares some characteristics[59]. The value of this approach is rooted in the splitting of data into a large number of small models and finding the best combination of variables over several iterations.

The primary goal of this study is the selection of texture features whose profile, in some way, associates with the IDH mutational status of the glioma that was imaged. In statistical language this is known as 'supervised classification', and there are a number of different methods that can carry out 'tests' to determine if the texture features that we have generated from the images are able to distinguish the tumour phenotype. The methods of testing can be subdivided into either univariate or multivariate methods. Univariate methods look to evaluate each variable (e.g. texture feature) one at a time, to determine its ability to discriminate between to the testing groups (e.g. IDH Mut vs IDH WT). Such methods have been shown to perform well in certain circumstances, such as applications in functional genomics[106,107]. These methods, however, have been found to be ineffective when applied to complex biological questions. One conceptual limitation when applying such methods to highly dimensional imaging texture data is that the derivation of features is not always obvious, and there may be varying factors that can be attributed to a myriad of physiological processes within the tissue being imaged that contribute to their eventual calculation. It is therefore important to assess how these features perform as a group in being able to distinguish between the two histological subtypes. Multivariate methods may be more suited for the analysis of such data, since variables are considered in combination to identify interactions between imaging
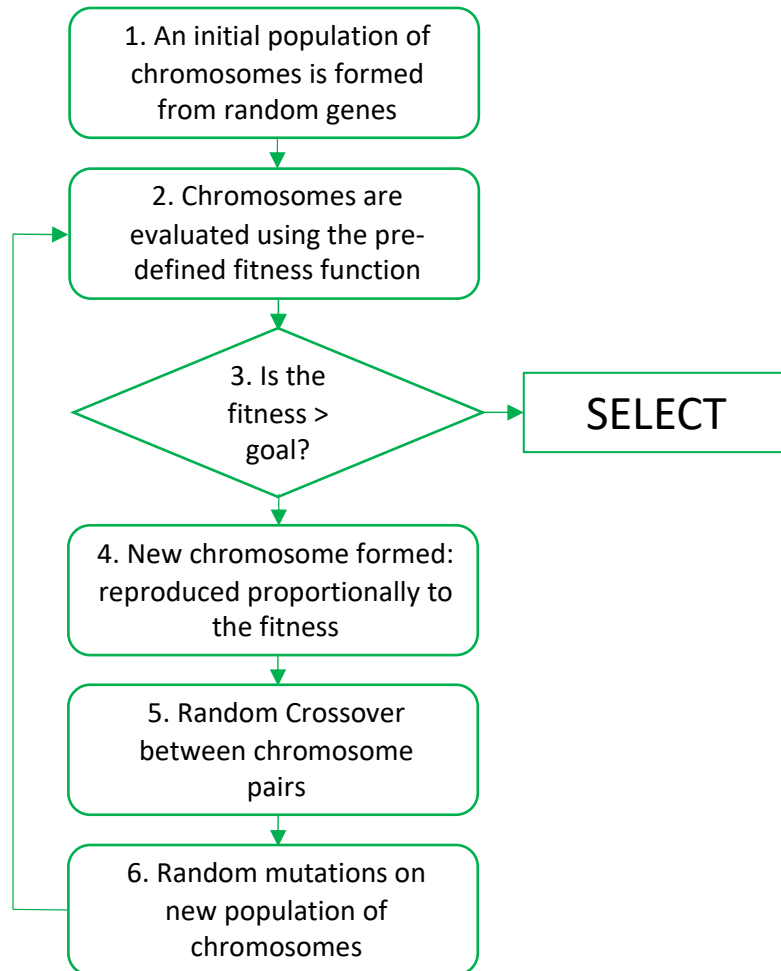
Fig 3.3 a Flow diagram to demonstrate the stages of the GA approach to feature selection prior to classifier construction. If the chromosome fitness function score, awarded based on accuracy of classification, reaches above the pre-defined level, then the process is terminated as a solution has been reached.

features and disease state. There are, however, an excessively large number of possible models that may be constructed from different combinations of the hundreds of texture features and it is not possible to critically evaluate these using the available computing resources. One possible alternative to this in depth scrutinization of each of the individual model, is to use a search procedure that is able to 'explore' looking for sets of variables that are good at classification, but perhaps not optimal. This procedure has been demonstrated and applied successfully in the context of microarray data[108–110]. GALGO is a software package that supports the development of statistical models using multivariate variable selection strategies[111].

GAs work by applying *in silico* models of evolution by natural selection as search procedures for variable selection. This process works by evolving sets of chromosomes (groups

of variables) that are able to fit a certain set of criteria (fitness function) from an initial random population through cycles of differential replication, recombination and mutation of the fittest chromosomes. The concept of applying a model of evolution to explore sets of variables to find those that are 'good' for discrimination between testing groups was first introduced by John Holland in 1975[112]. The popularity of GAs increased phenomenally once sufficiently powerful computers became readily available. Below is a description of how general GAs work, broken down into stages (figure 3.3):

Stage 1: From the input data, a number of random sets of variables (chromosomes) are created. These sets of variables (chromosomes) can form a population of chromosomes (niche).

Stage 2: The ability of each of the chromosomes within the niche to predict the group membership of each of the samples is assessed (fitness function). This is accomplished by training a statistical model. The GA then tests the accuracy of prediction of each of the chromosomes and assigns a score based on the accuracy, which is referred to as the fitness function.

Stage 3: When a chromosome is awarded a score that is higher than the predefined fitness function, this chromosome is then selected and the procedure stops. If the cut off score is not reached, then the procedure continues onto the next stage.

Stage 4: The population of chromosomes is replicated, whereby the strongest chromosomes with the highest fitness function scores will generate more offspring.

Stage 5: The information contained within the chromosomes is combined through the process of genetic crossover. This involves taking two randomly selected chromosomes and swapping sections of variables between them. By employing this method, it is possible to increase the search capacity for possible solutions.

Stage 6: Mutations are randomly introduced into chromosomes, by way of adding, deleting and swapping individual variables.

Stage 7: The process loops back to stage 2 until a chromosome that is accurate enough is produced.

For the calibration of GLAGO the following parameters were used. K-nearest neighbour was the classifier of choice, chromosome size:10, max solutions: 200, goal fitness: 0.80, save frequency:30. Due to the imbalance in sample size of each testing group, whereby there were much fewer Mut patients when compared to the WT patients (which is representative of the

normal distribution), we attempted two approaches at correcting for this. The first approach was to subsample the data by creating a subsample index in R, and the second method was to introduce a weighted error.

3.2.5 Random Forest for Radiomic Classification of IDH Mutation

The large number of texture features generated (973) compared to the relatively smaller sample size can result in overfitting of classifiers. Furthermore, within the data there may be varying degrees of correlation and redundancy between many of the generated features. In a further attempt to optimise feature selection with respect to classification of tumour IDH status we employed a random forest classifier approach to select and optimise the feature set. Random forests was chosen as an alternative route of investigation to the GA approach as it also works by assessing a vast number of small models to find the best combination of variables for classification, however it is much less computationally expensive and results can be generated much more rapidly.

Although canonical classification approaches such as support vector machines have demonstrated utility in a wide variety of classification problems, the large and heterogenous nature of the training data available from texture feature generation in radiomics investigations is considered to be beyond the capacity of single classifiers[113]. This has led to the adoption of ensembles of classifiers, bagging and boosting, to address such problems. This approach uses a great number of simple models that are trained on parts of the data, followed by aggregation of their predictions onto new data. Random forests uses the bagging method, whereby the algorithm draws on training examples to a subset of features when training the simple, 'weak learner' models[114]. This forms a major part of the rationale as to why random forests were the classifier of choice for this part of our investigation, as it provides consistent proficiency in the estimation of information value of each of the individual predictor variables for accurate classification.

Here we used the 'mlr' package within R to fit a random forest model to these data. The structure of the random forest algorithm is such that feature selection comes under the category of embedded methods, whereby such algorithms combine the qualities of filter and wrapper methods. Random forests are constructed from large groups of decision trees, whereby each of the decision trees are built under a random extraction of observations form the dataset, and a random extraction of features. Each tree within the forest can only select

variables from a subset of features, forcing variation amongst the trees within the model and facilitating feature selection and filtering. Furthermore, the construction of an ensemble of uncorrelated models allows us to build an overall classifier from a large number of variables that can allow accurate predictions.

An initial random forest model was fit with default parameters where the data was split 2/3 train 1/3 test. A receiver operator curve (ROC) was fit and the mean misclassification error was generated. We then carried out K-fold cross validation to assess the performance of the model. K-Fold cross validation is a popular method of carrying out model validation as it generally results in a less biased model compared to other methods, as it ensures that every observation from the original dataset has the chance to appear in the training and test set[115]. Here we used 10 fold and 5 repetitions resulting in a total of 50 iterations. A ROC was fitted for each of the iterations, and a mean ROC was fit to display overall performance. Accuracy, precision (equation 20), recall (equation 21) and F1 scores (equation 22) were calculated. Accuracy is calculated by dividing the number of correct predictions by the number of total predictions. Precision can be defined as:

Equation 20

$$Precision = \frac{True\ positive}{True\ Positive + False\ Positive}$$

Recall is calculated using equation 21:

Equation 21

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

F1 score is calculated using the formula described in equation 22:

Equation 22

$$F1 = 2\ \times \frac{Precision \times Recall}{Precision + Recall}$$

51

Factors of tumour size, cellular pathological origin (oligodendroglioma, astrocytoma and primary/secondary glioblastoma) and scanner manufacturer were also taken into account. Due to the dependence of some of the features on scanner and tumour size, a further variable selection approach was adopted. Spearman correlations were calculated to assess the influence of tumour size over the descriptors. A multiple linear model approach was adopted
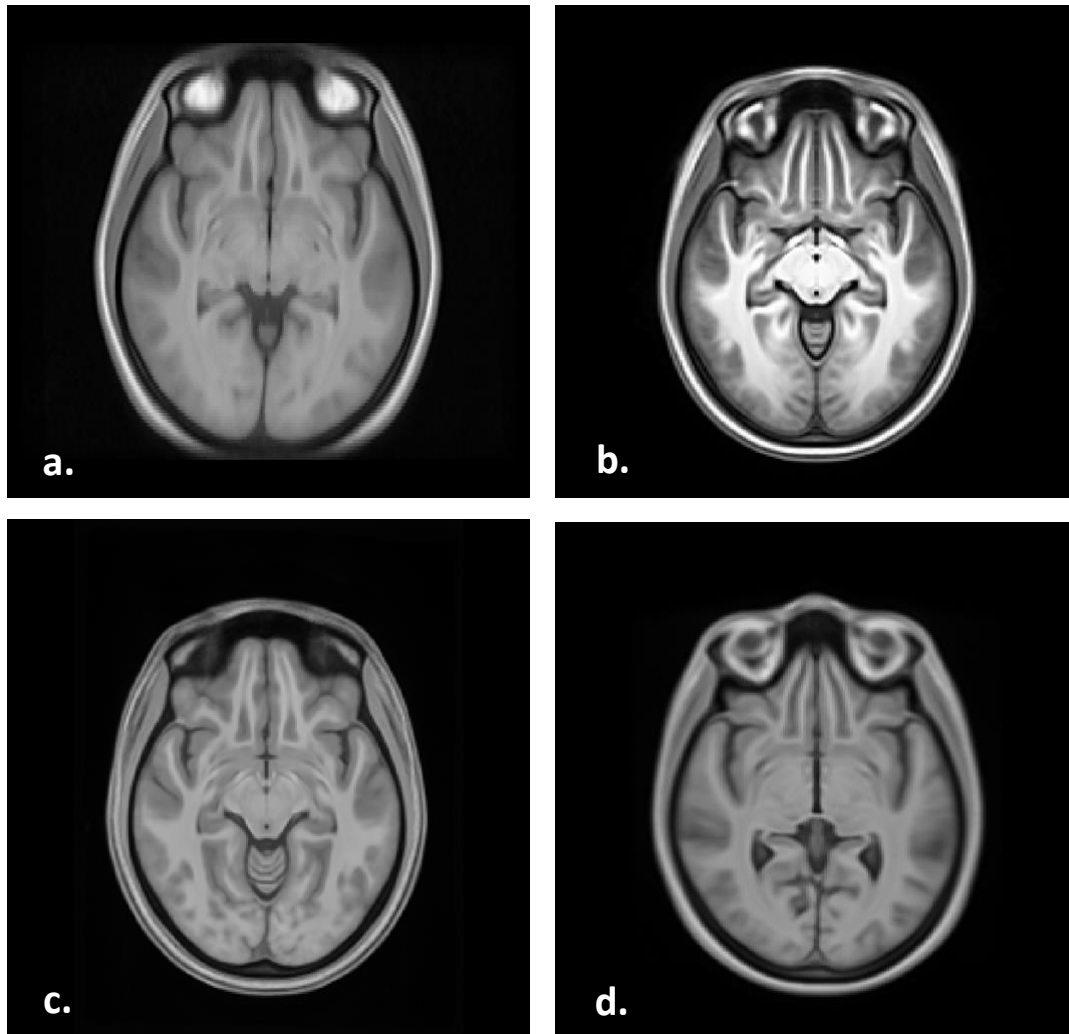
Fig 3.4. Axial slices from the 4 templates constructed from public MRI dataset libraries of varying sample size. **(a)** Template constructed from the LPBA40 dataset consisting of 40 healthy volunteer images. **(b)** Template constructed from the Kirkby dataset consisting of 21 healthy volunteers. **(c)** Template constructed from the NKI dataset consisting of 60 datasets. **(d)** Template created from the IXI dataset, consisting of 600 healthy scans. As the number of datasets increases the registration mismatch errors become averaged out and the likelihood of sharing mutual information with the target image increases.

to assess the potential influence of factors such as tumour pathology, scanner manufacturer, tumour size as well as IDH status. For this we used a multi factor ANOVA. A second random

forest model was fit following the same procedure as previously described using variables where IDH status was important but other factors were not.

3.3 Results

3.3.1 Brain Extraction Optimisation

The template created from the LBPA40 dataset was formed by averaging 40 datasets from healthy volunteers. The quality of multisubject templates created by averaging various subject image datasets is dependent on how well they are registered to one another. Potential mismatch errors are more likely to be averaged out as the sample size increases, thereby increasing the quality of the template and the likelihood of the presence of mutual information with the subject dataset to which the template is being registered to. The performance of each of the brain extraction attempts using the various constructed templates has varied, whereby varying levels of skull and meninges remained in the images post-extraction. Figure 3.4 shows axial slices from each of the constructed templates. Upon visual inspection the template constructed from the IXI dataset seemed to perform the best as there was the minimum amount of skull remaining on these images when compared to other templates such as the one constructed from the LPBA40 dataset (figure 3.5).

3.3.2 GALGO for Feature selection

3.3.2.1 T$_1$ post-contrast whole brain analysis

Figure 3.6 shows the fitness function plot from the initial GALGO investigation using T$_1$ post-contrast whole brain texture features. This plot allows us to analyse whether or not we are managing to get solutions from the GA. Here we are able to see the evolution of the fitness function value across the generations of chromosomes. We were able to determine that on average, we are reaching a solution at generation 69. The two lines show the average fitness for all the chromosomes, as well as those that have not reached the pre-defined fitness function respectively. Once the chromosomes have been selected by determining those that surpass the fitness function, we need to assess the classification accuracy. GALGO uses the bootstrapping method to estimate the error of the models, whereby the error is determined by averaging estimates calculated from multiple small samples of the data. The samples of the

53

data from which the individual errors are estimated are constructed by drawing observations from the dataset one at a time and then replacing them to the dataset once they have been chosen. This is an optimal strategy for error estimation when sample sizes are relatively low. Figure 3.7 shows the confusion matrix box plots for the models derived from whole brain T_1 post contrast texture features. The low sensitivity of Mut detection (0.28) seen in figure 3.7 (a) may be attributed to the imbalance in the sample size of the testing groups.
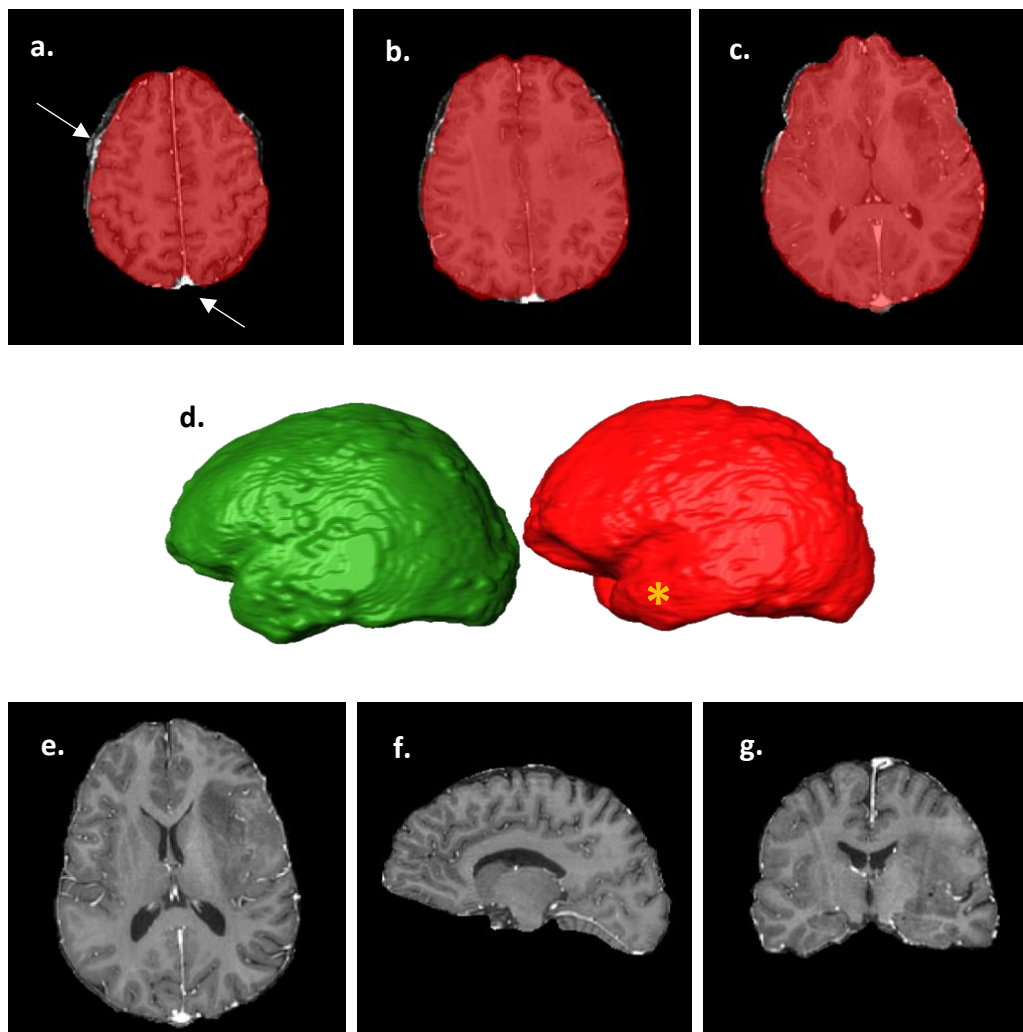

Fig 3.5. Comparison between the brain extraction result and mask generation results generated from the LPBA40 and IXI constructed templates. **(a)** Superior section of extracted brain, the MR image underneath is the result from extraction using LPBA40 generated template, and the red overlay is the result from the IXI generated template. It is evident that there remains a relatively large proportion of skull and meningeal tissue from the LPBA40 extraction (white arrows). **(b)** Mid axial view **(c)** Inferior axial view. **(d)** 3D render of extraction result. Green: LPBA40 extraction. Red: IXI extraction. The comparison here is most evident when looking at the temporal horn (yellow star). **(e)** 3D Extraction result from IXI template. There is no evident meningeal of skull tissue remaining. **(f)** Sagittal view. **(g)** Coronal view procedure, the images experienced a large amount of residual skull and meninges.
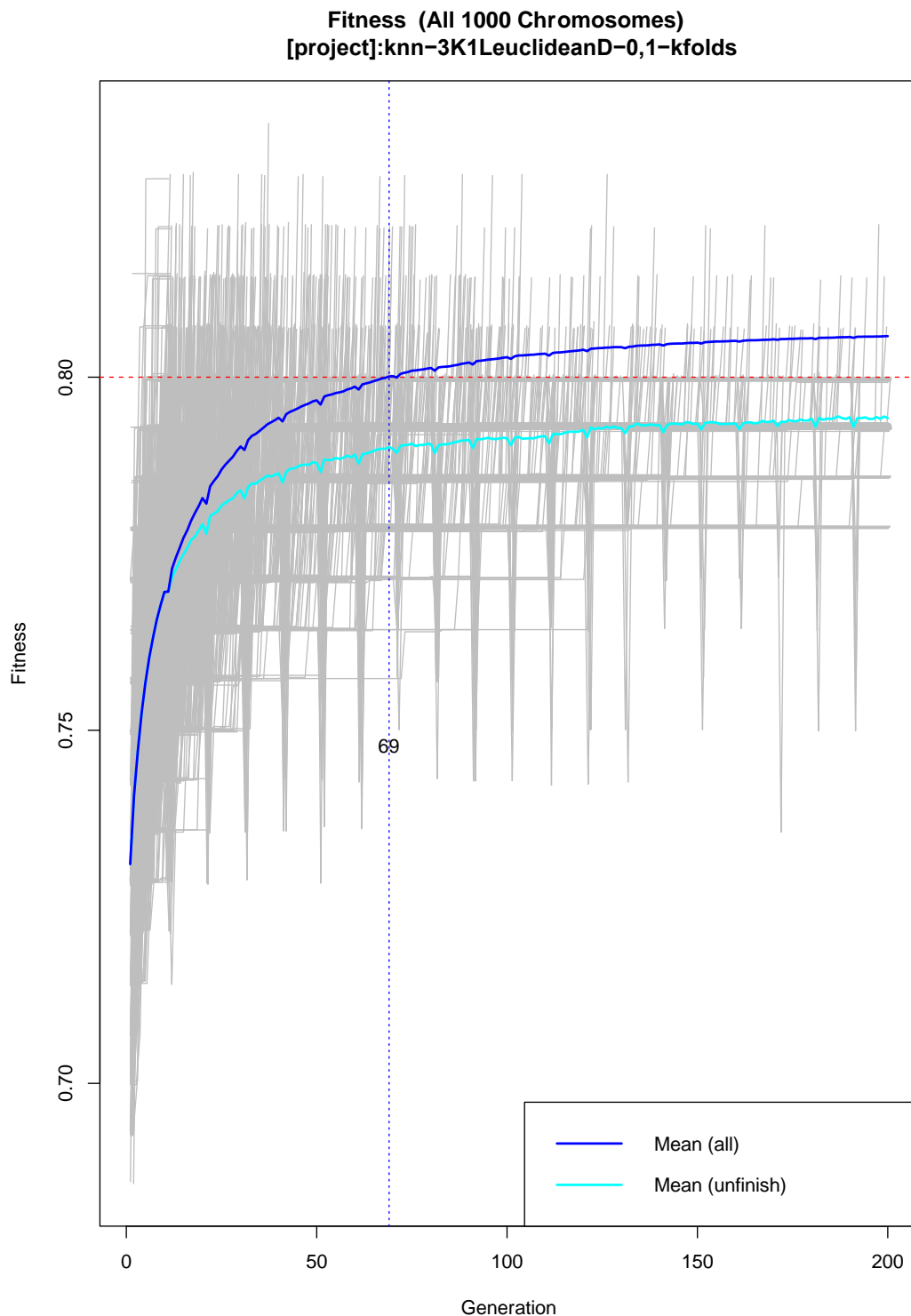
Fig 3.6 A fitness function curve for the initial GALGO experiments using data derived from T_1 post-contrast whole brain texture features. Here we are able to determine that we reach the fitness threshold of 0.8 by generation 69 of the chromosomes.
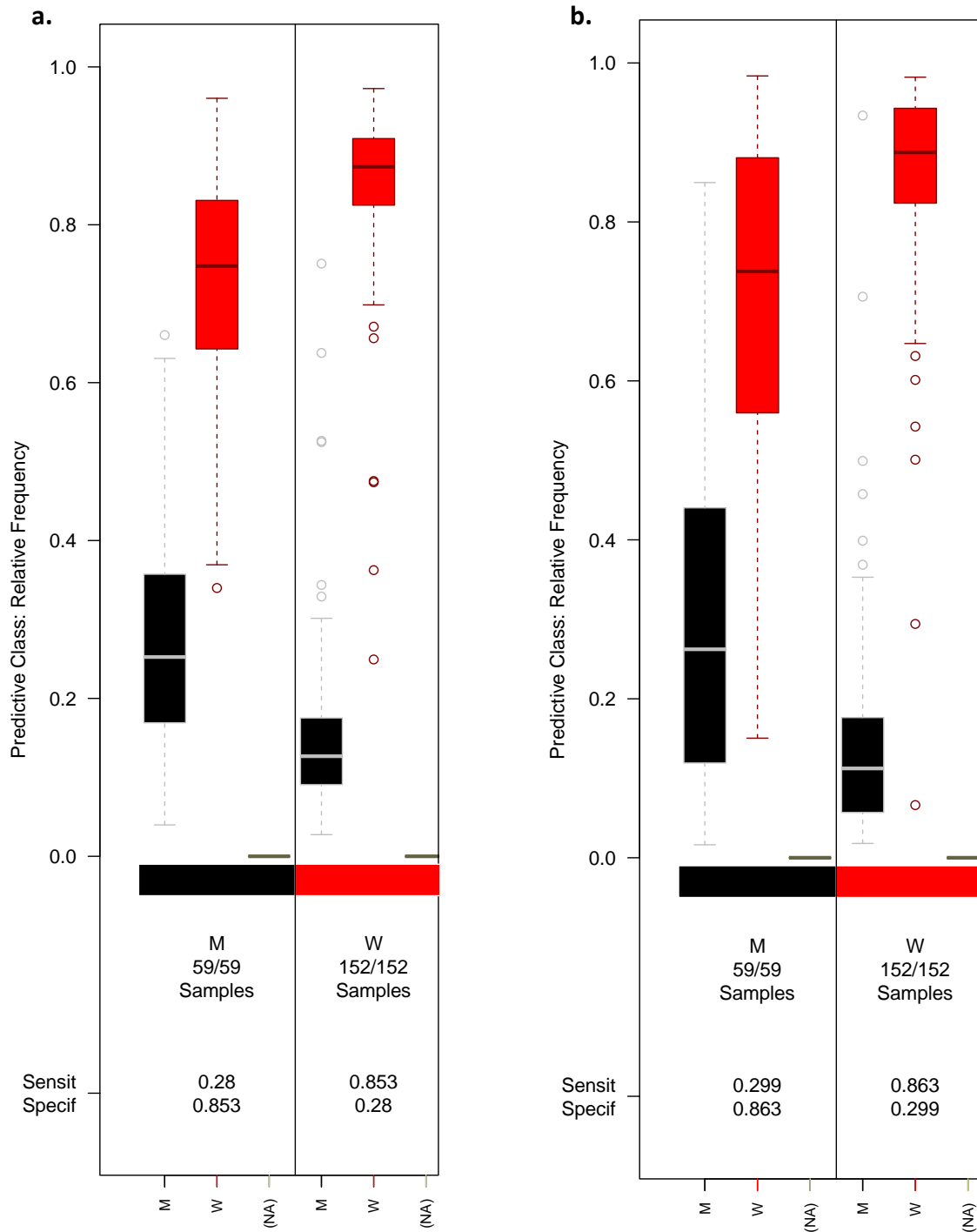
Fig 3.7 A box plot to highlight the distributions of the averages of class prediction for each class. **(a)** class confusion calculated from whole brain T_1 post-contrast images without application of subsampling in order to account for the imbalance in Mut vs WT. The poor sensitivity of the model when predicting Mut tumours may be accredited to this imbalance **(b)** class confusion calculated from the same dataset however this time a subsampling routine was introduced prior to the configuration of GALGO in order to account for the imbalance in group sizes. Despite the attempt to correct for this imbalance the model performed to a similar standard as without subsampling, with a negligible increase in sensitivity of Mut prediction.
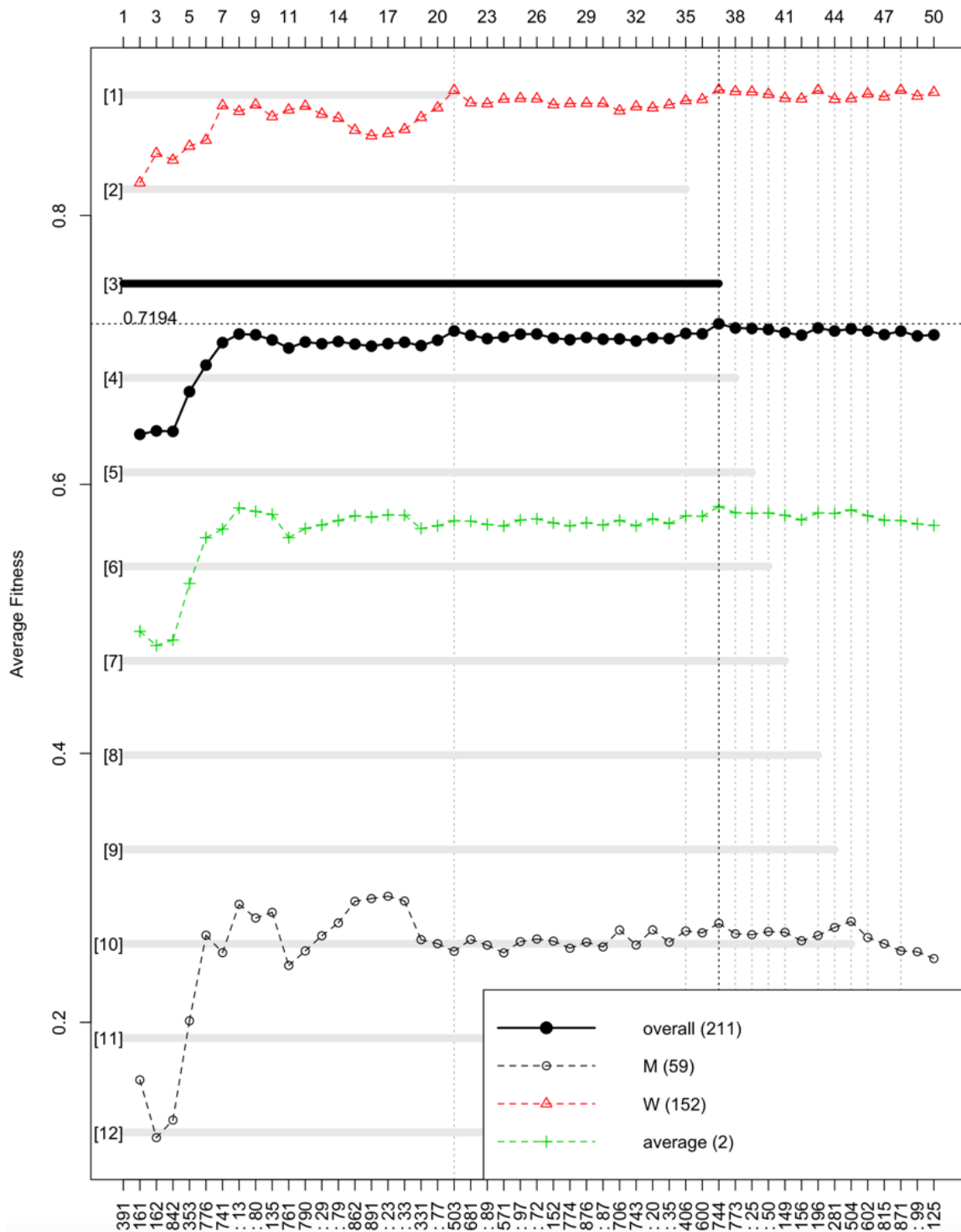
Fig 3.8 Plot to demonstrate the forward selection strategy for the models generated from the GA using the most frequent genes. The horizontal axis represents the genes ordered by their rank and the vertical axis displays the overall classification accuracy. The coloured dashed lines represent the accuracy per class. The thick black line represents the one model that resulted from selection whose fitness value is maximum (number 3). However, as we can determine from here, the classification accuracy of this model is only 0.7194, which is below our specified 0.8 fitness function value.

In an attempt to account for this, a subsampling routine was introduced prior to the configuration of the GA, artificially balancing the size of the testing groups. Despite this approach the sensitivity of Mut detection remained very low, as the classification error of each of the proposed models was high. This can be seen in figure 3.7 (b).

Despite the poor performance of the models generated from the configuration of the GA in classification of Mut tumours from whole brain texture features, we wanted to explore the capabilities of GALGO further and determine how either successful or unsuccessful the resultant model would be at classification. The outcome of the GA is a large number of chromosomes (groups of features), all of which provide good solutions that ideally surpass the pre-defined fitness function, however it is not clear which one should be chosen to develop a classifier. Therefore, there is a requirement to build one model that is representative, in some way, of the population. Figure 3.8 is a forward selection plot of the best models using gene frequency and average fitness to determine success. This selection is carried out by assessing the fitness function within all test sets. Model number 3, containing 36 of the most frequent genes (features), performed the best in terms of accuracy, however even though this was the best performing model, the classification error was 0.7194.

Figure 3.9 shows a heatmap generated from the best model as a result from forward selection whereby hierarchical clustering has been carried out on both the genes (texture features) in rows and samples in columns. Although there is clustering of the variables within this heatmap, is evident that this is not correlated with IDH mutational status. Further to this, Figure 3.10 displays a principal component analysis (PCA) plot of the first 4 principal components. There appears to be no separation between the IDH Mut and WT groups based on the PCA plot of the best performing model as a result from forward selection.

Taken together there was no discernible set of texture features, derived from T_1 post-contrast whole brain analysis, that were able to classify the tumours as IDH Mut or WT using this approach for feature selection and model generation. Following on from this, the next step was to assess if features that were calculated from only within the tumour, from the same images, may perform better at radiogenomic classification.
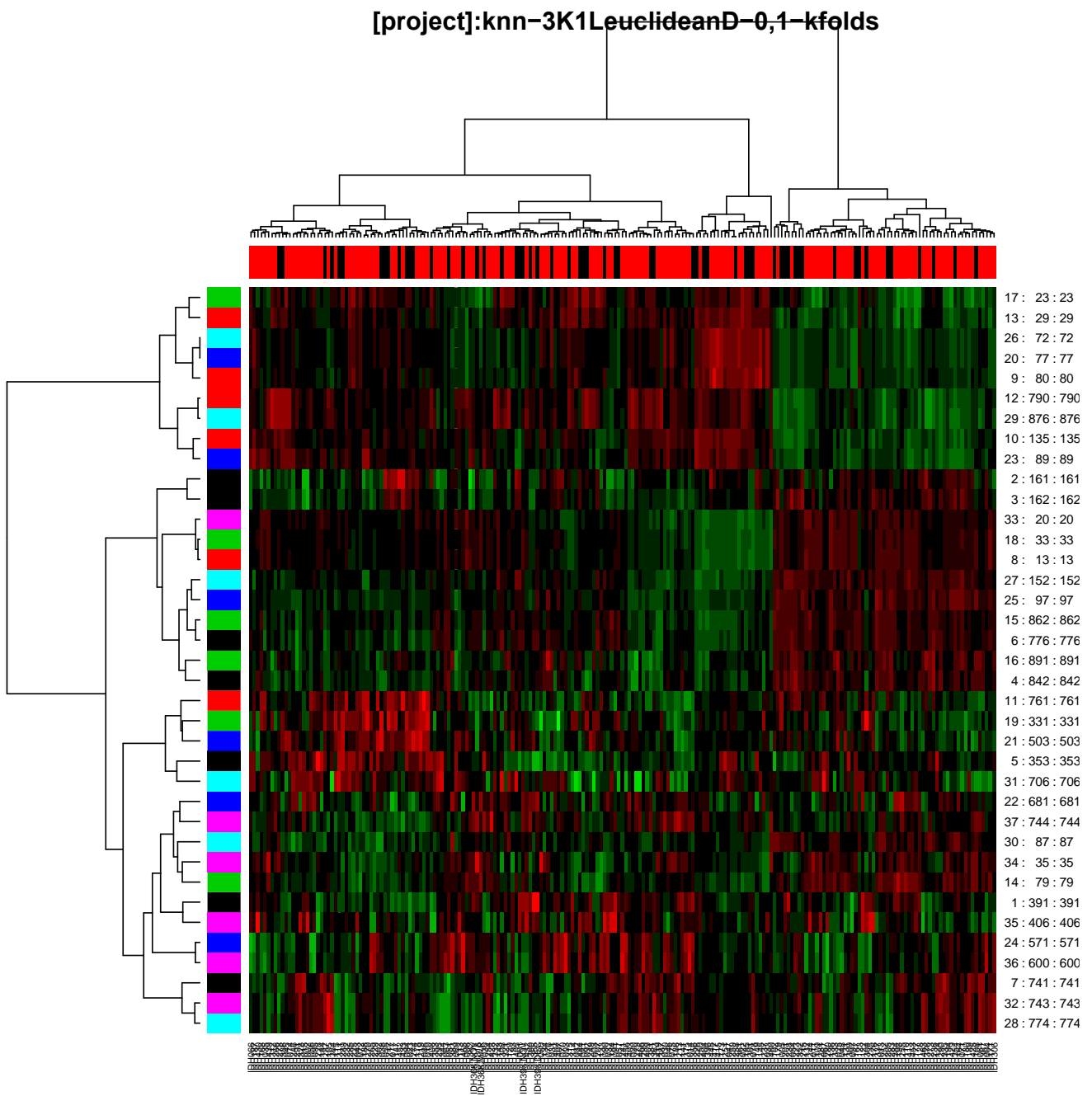
Fig 3.9 A heatmap plotted from the best model as determined by the forward selection procedure. Hierarchical clustering has been implemented on both the genes (texture features) in rows and IDH mutational status in columns. It is evident that even though there seems to be some clustering of variables within the data, these cannot be attributed to IDH mutation.
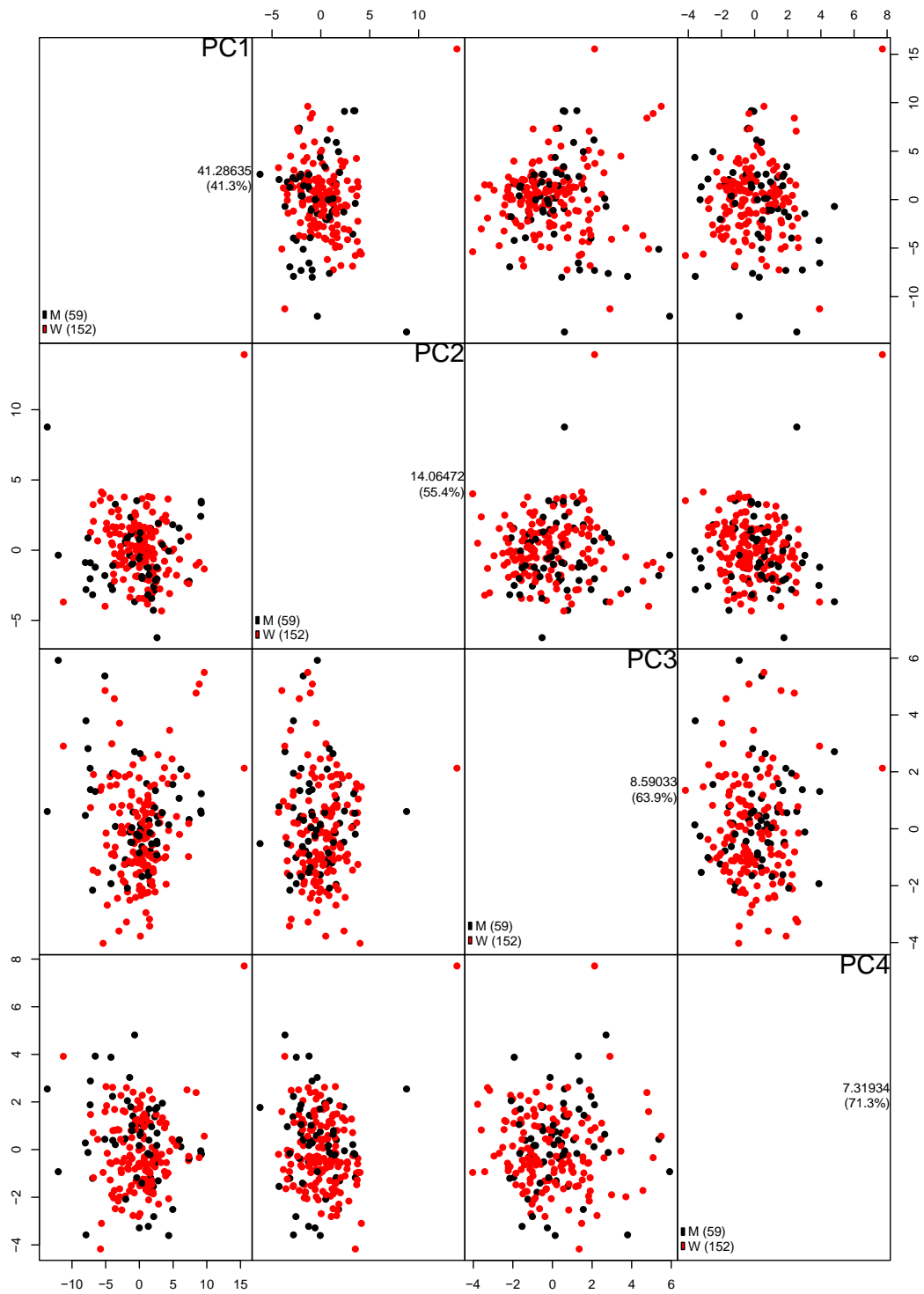
Fig 3.10 A PCA plot of the first 4 PCs of the best performing model as determined by the forward selection procedure. It is evident that there is a distinct inability of the model to classify the tumours within the images into IDH Mut or WT based on texture features as there is no separation between the two groups on any of the above PCs.

3.3.2.2 T_1 post-contrast tumour analysis

The previous investigation into whole brain texture features that may facilitate the classification of tumours into IDH Mut or WT using a GA approach for feature selection to build the most representative model failed to yield a model that had a reasonable classification error. In this section we are taking the same GA approach, but this time only considering texture features that have been generated within the enhancing region of the T_1 post-contrast images. All parameters were maintained as per the previous investigation. Figure 3.11a shows the confusion matrix box plot for the T_1 post-contrast intra-tumoral data both with and without a subsampling scheme applied. The sensitivity of Mut detection was very poor, at only 0.181. When comparing this classification error to the previous investigations, where the texture features from the whole brain were considered, the features calculated from within the enhancing region of the tumour seem to have performed worse following feature selection and model generation using the GA approach. Conversely the specificity of WT classification was calculated at 0.951, suggesting that detection of a true positive when dealing WT images performs well. Despite the poor classification error associated with Mut detection of the models generated from the GA, we explored the forward selection procedure to assess what effect the classification imbalance would have on overall model performance. Figure 3.11b shows the plot to demonstrate the forward selection procedure. The overall classification accuracy is shown to be 0.88, which may be misinterpreted as good in terms of model performance, however this value is calculated by taking the total of misclassified samples, divided by the total number of samples. Therefore, when the distribution of Mut and WT samples within the cohort is taken into account, and the poor classification error of the Mut samples is considered, it may be determined that the model is not representative. Furthermore, the stability of the Mut classification on the graph is poor, adding to the uncertainty of the even the best models' predictive power. To further illustrate this the heatmap seen in figure 3.12 shows that there may be some clustering of variables with respect to IDH status, however these either reflect only WT classification and not Mut, or are not in line with IDH mutational status. The heatmap has been constructed with hierarchical clustering applied. Taken together, these results suggest that applying a GA approach to feature selection and model generation to image texture features calculated from within the enhancing region of the tumour on T_1 post-contrast images does not expedite the accurate classification of tumours into IDH Mut or WT.
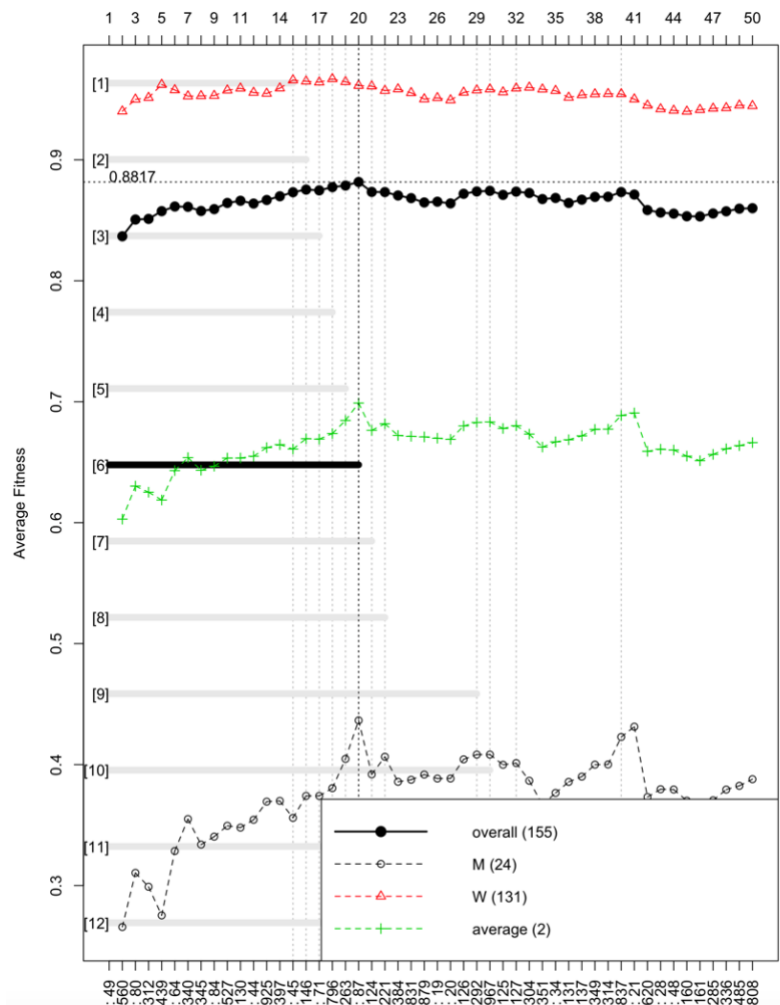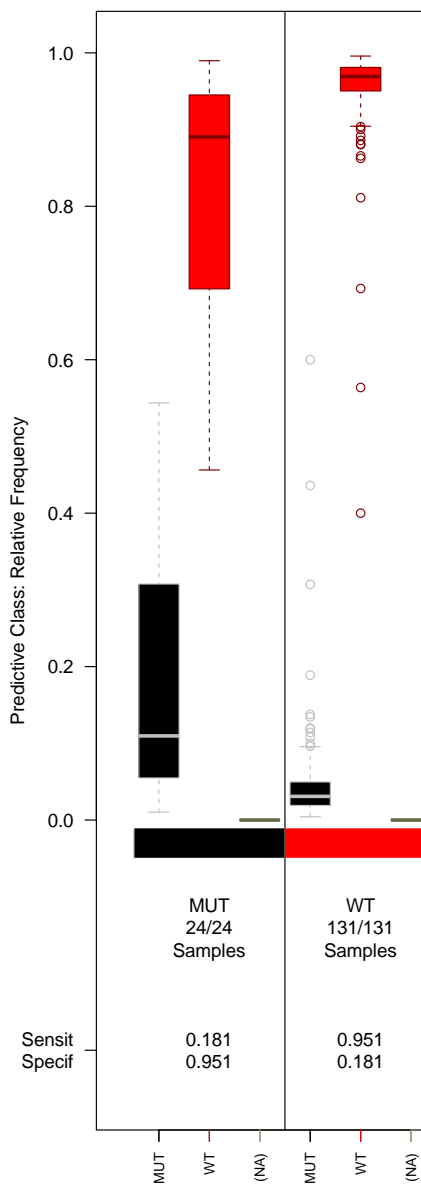
Fig 3.11 (a) Confusion matrix box plot representing the classification errors of the proposed models following GA facilitated feature selection based on texture features that were calculated from within the segmented tumour. The box plot has highlighted that whilst the sensitivity of WT classification was excellent, the sensitivity for Mut classification was in fact very poor, at just 0.181. **(b)** A plot of the result from the forward selection procedure following GA feature selection and classification. Whilst the overall classification accuracy (black line) is computed at 0.8817, this is calculated by taking the number of misclassified samples and dividing by the total number of samples, and therefore does not reflect the poor classification accuracy of the model when dealing with Mut tumours.
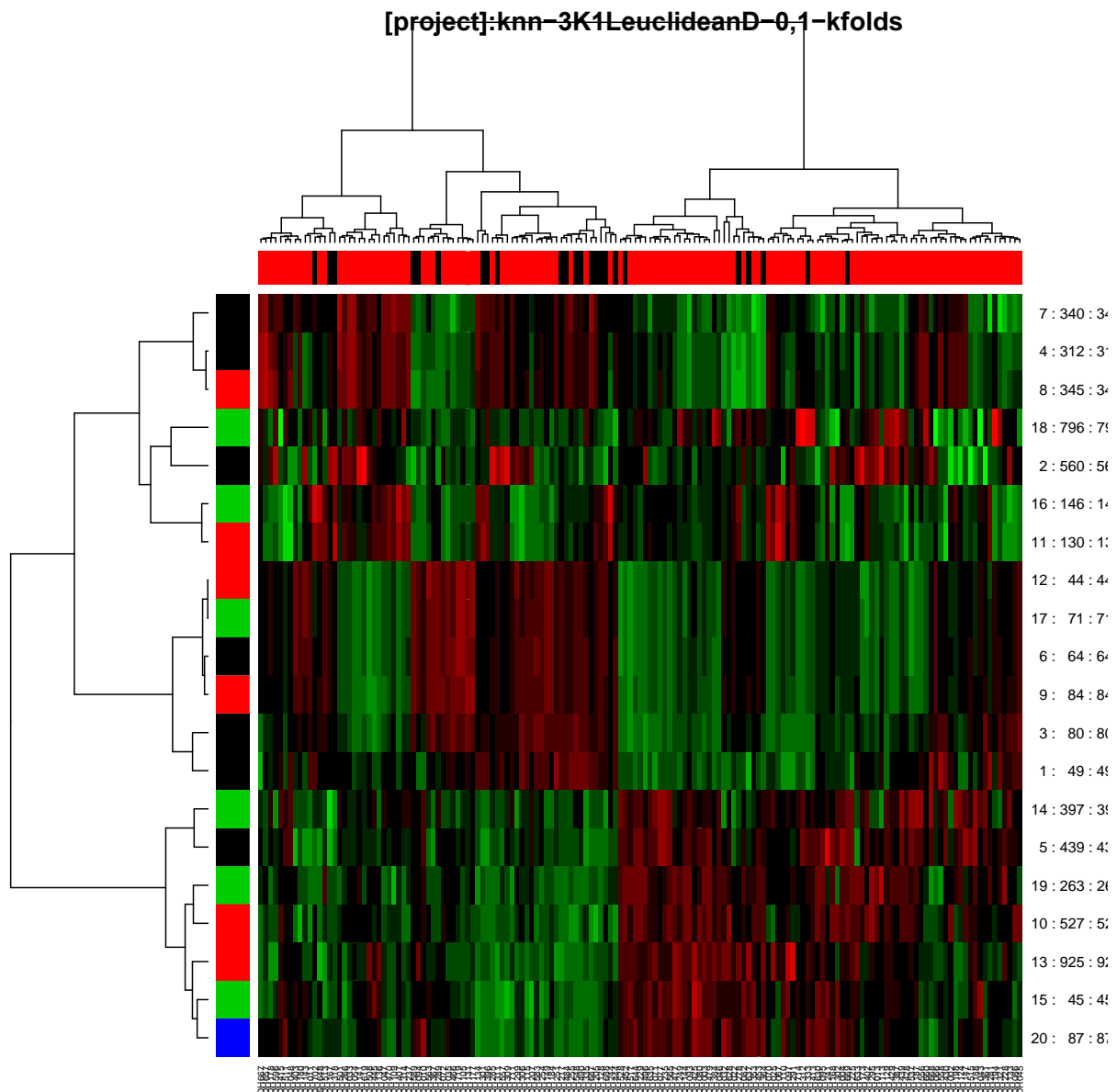
Fig 3.12 A heatmap generated from the best performing model as a result of the forward selection procedure (labelled as 6 on figure 3.10). Hierarchical clustering has been applied with chromosomes ordered in rows and mutational status in columns. There appears to be some very weak clustering within the data, however these are either due only to WT classification, or not in line with IDH mutation.

3.3.2.3 T_2 whole brain analysis

In this part of the investigation we aimed to assess if texture features derived from the whole brain of glioma patients were able to distinguish IDH Mut from WT using the GA approach for feature selection and model generation. For this analysis only whole brain T_2 texture features were used. Furthermore, the number of samples included in this part of the investigation is smaller, as not all patients underwent T_2 imaging investigations. Figure 3.13 displays a plot of the fitness function using the same parameters as the T_1 post-contrast imaging investigation. The plot shows that after the maximum of 200 solutions were reach, none of the chromosomes managed to surpass the stipulated fitness function of 0.8. This represents that none of the groups of texture features within the chromosomes were able to distinguish IDH Mut from WT with a classification error better than 0.8. Additionally, figure 3.14 shows the plot derived from the confusion matrix of the class predictions from all chromosomes. Here we can see that the sensitivity of Mut classification is once again poor, however, in contrast to the T_1 post-contrast investigations, the sensitivity of the WT classification is also quite poor, not managing to surpass the pre-defined lower limit of 0.8 reaching only 0.763. When considering that we are attempting to construct a model for potential translation into clinical decision making, this level of sensitivity is far from adequate. Despite the poor average classification error of the GA search result, we wanted to explore if there were any models that were able to perform with increased accuracy. In order to do this, we carried out the forward selection procedure to assess if there were any particular models that were more accurate at classification. Figure 3.15 shows the results of the forward selection procedure facilitated by GALGO. As was highlighted in the confusion plot, the average classification error failed to surpass the minimum of 0.8. Here we can see that the best performing overall model (labelled as 3 on the graph) had an overall classification error of 0.71, and once again the Mut classification performed more poorly than the WT classification. This may be attributed to the imbalance in sample distribution. Figure 3.16 is a heatmap ordered with hierarchical clustering applied, showing all models used in the forward selection procedure. It is evident that despite the small degree of clustering, this is not attributed to the mutational status of the samples. Taken together, all the aforementioned results derived from T_2 whole brain data demonstrate that we have been unsuccessful in applying the GA approach for feature selection and model generation to classify the samples into IDH Mut or WT based on imaging features.
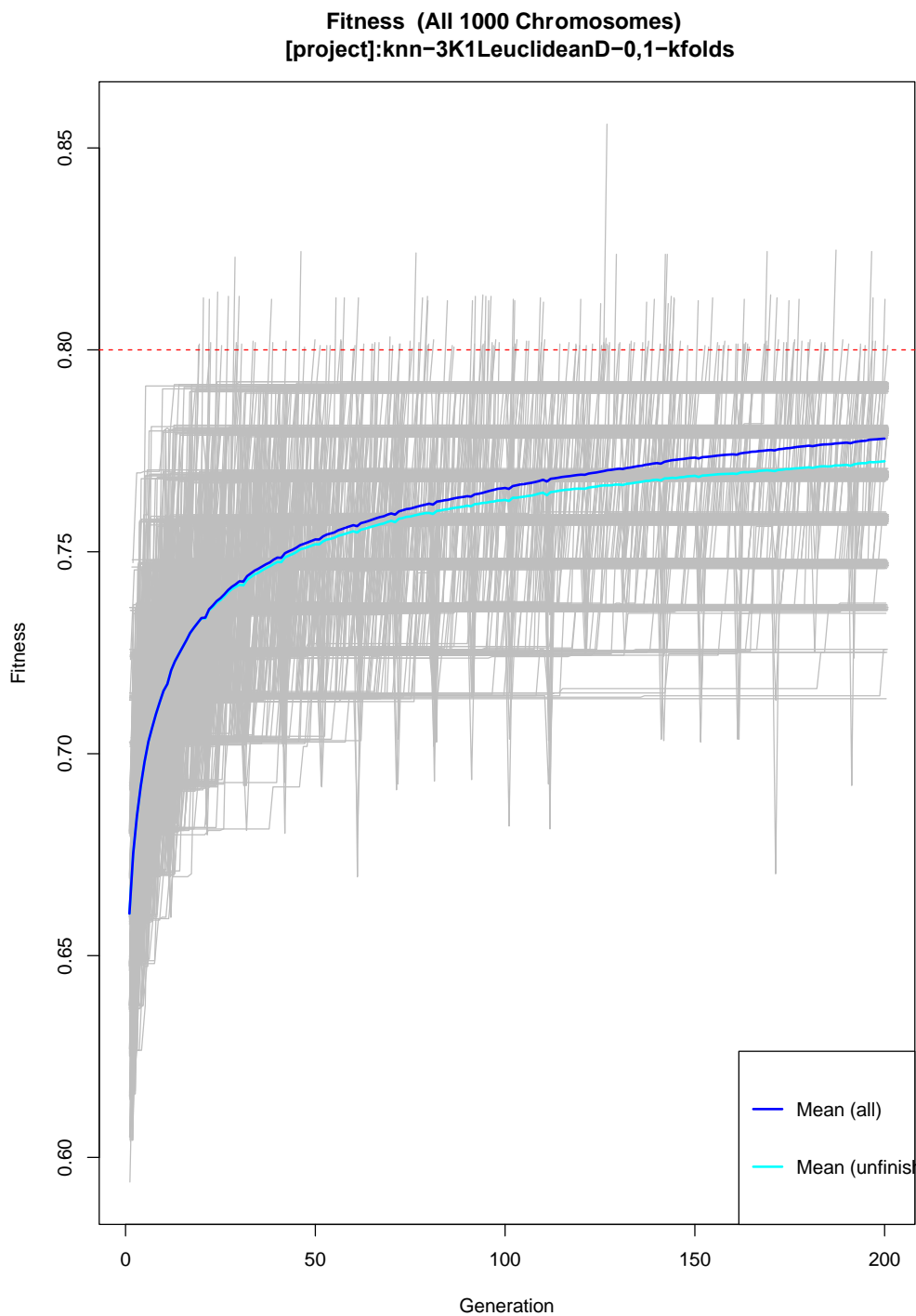
Fig 3.13 A plot of the fitness function over the chromosome generations. From this we can determine that we were unable to reach solutions following 200 generations, as the minimum classification error was unable to surpass the minimum predefined fitness function of 0.8.
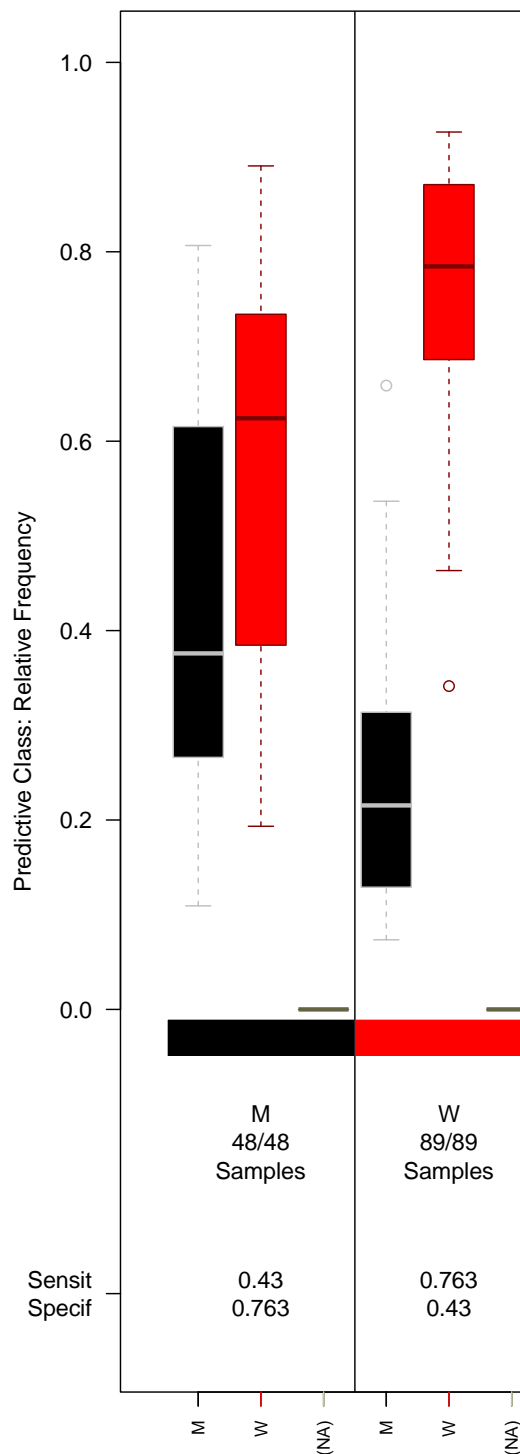
Fig 3.14 A box plot of the confusion matrix derived from the GA search using texture features generated from T_2 whole brain analysis. Here we can see that the sensitivity of Mut classification, although improved form both T_1 post-contrast whole brain and intra-tumoral analysis, it is far below the minimum expected of 0.8 at 0.43. Furthermore, the classification error of WT classification is also below the minimum expected value of 0.8. The forward selection procedure was then followed in order to assess the performance of individual proposed models.
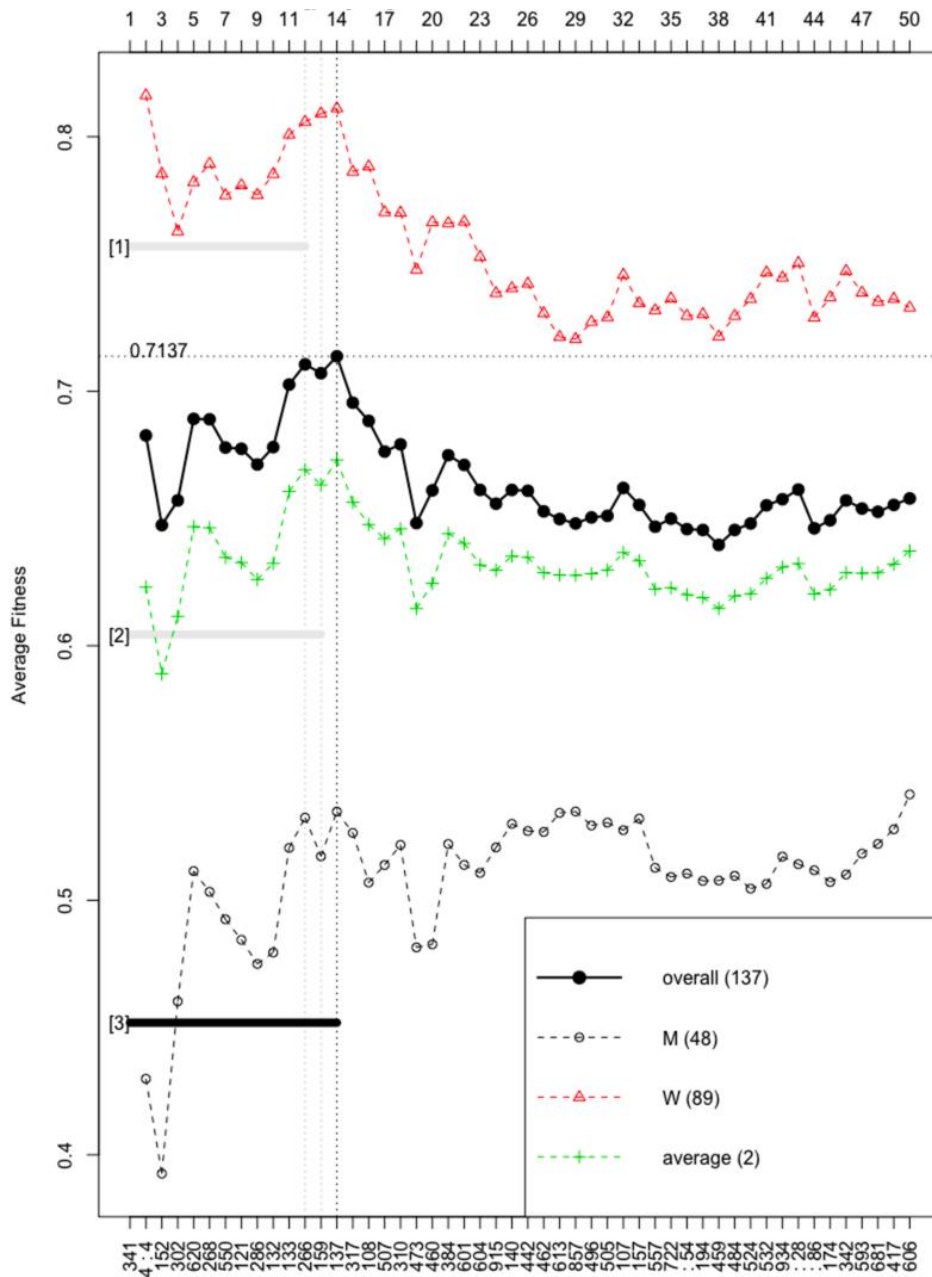
Fig 3.15 A plot of the result from the forward selection procedure in order to assess the performance of individually proposed models. Here we can see that the best preforming models, in both Mut and WT classification, fail to surpass the minimum expected classification error of 0.8. the average classification error of the best preforming model (3) was calculated at 0.71.

3.3.2 Random Forests for intra-tumoral T_1 post-contrast analysis

Here we have used a random forests classifier approach in an attempt to build a representative model that is able to classify the tumours based on image texture features derived from the segmented, enhancing tumour region. As the GA approach to feature selection and classifier construction proved sub-optimal, the first step that was taken was to plot the raw, imaging feature data with respect to IDH mutation, to assess if there are any initial patterns that may be attributed to the mutational status of the patient. Figure 3.16 shows a heatmap (a) and PCA (b) plot to demonstrate the lack of separation between the two testing groups based on raw imaging features. Figure 3.17 shows a receiver operator curve (ROC) for a single random forest model, trained on a 2/3 train 1/3 test set of the data with no cross validation. The area under the ROC (AUC) represents a performance measure for the classification problem in hand, demonstrating the degree, or measure of separability. This value tells us how good the model is at predicting Mut's as Mut's and WT's as WT's. The higher the AUC value, the better the model is at distinguishing between these two different classes based on the input variables. For this random forest model (fig 3.17) the AUC is 0.796, which can be interpreted as, the model has a 79.6% chance of being able to distinguish between Mut and WT. Although, at first, this may seem like a well desired result, we cannot assume that this model is going to perform the same way on future, unseen data. We need some form of assurance that the predictive accuracy of the prediction that the model is producing will be replicated in the future.. Figure 3.18 shows the previous model (fig 3.17) following K-Fold cross validation with 5 folds and 10 repetitions. This led to 50 iterations in total, and the ROC's for each of the models can be seen in this figure. The average AUC for all cross validation iterations is 0.744, meaning that there is a 74.4% chance that the model will predict a WT or Mut tumour correctly based on the inputted texture features.

As our end goal is to construct a model that may have clinical translation capacity, this machine learning approach, whereby we are likely to predict mutational status of the patient wrong 25.6% of the time, is not acceptable. Furthermore, if we assess the distribution of the ROC curves from the K-Fold cross validation procedure, it is clear that there is a large amount
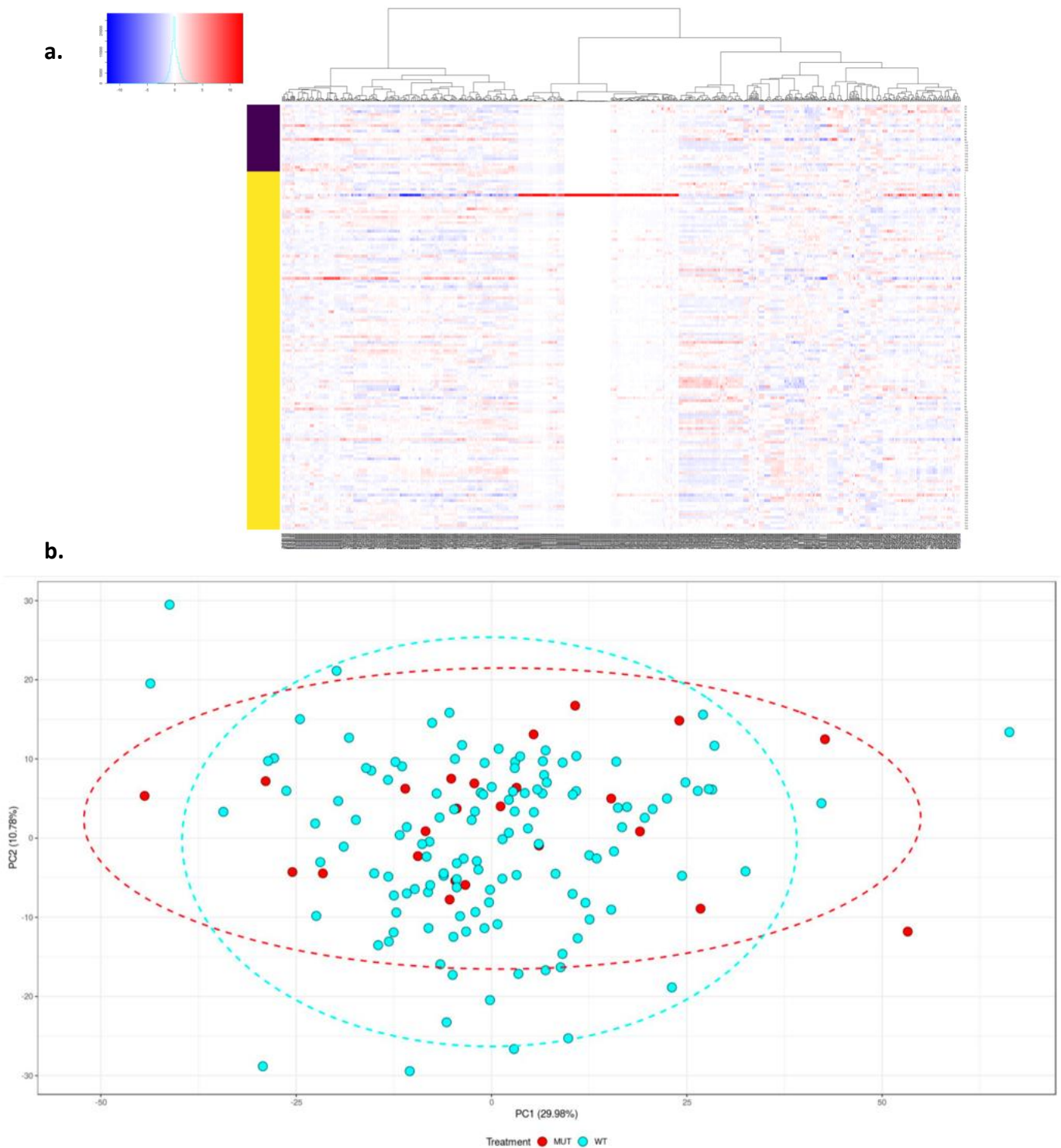
Fig 3.16 (a) a heatmap derived from the raw imaging data, whereby imaging features are on the x axis and samples ordered by mutational status (Mut in purple and WT in yellow) are along the y axis. There is no pattern within the data that seems to be evident. There is the presence of an outlier within the WT's indicated by the thick red horizontal cluster. **(b)** PCA plot using the raw texture features. There is clearly no separation between the Mut (red) and WT (blue) groups.
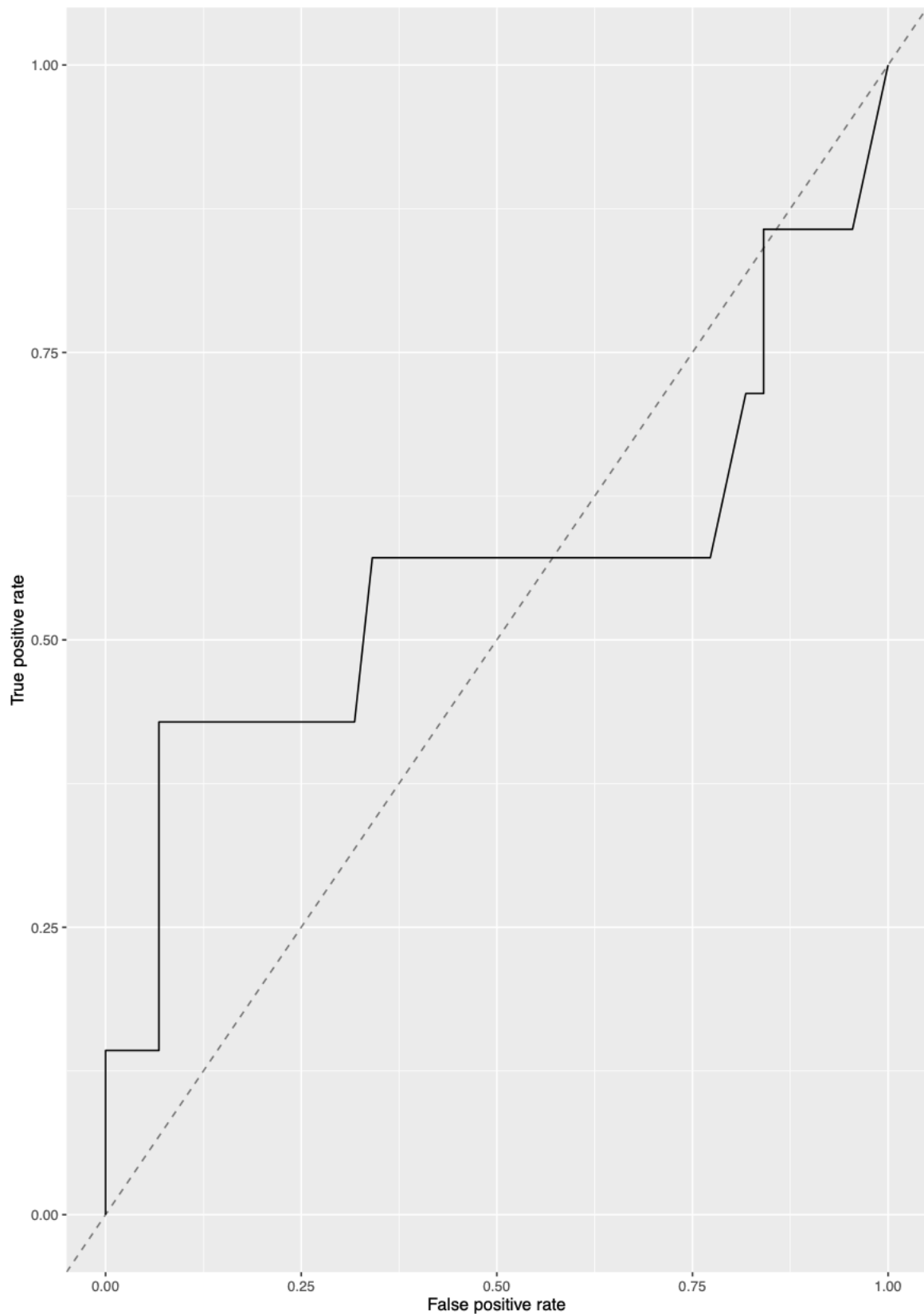
Fig 3.17 A ROC curve from a single random forest model generated using a 2/3 training 1/3 test split of the data. The AUC here is 0.79. Although this is close to the minimum classification error of 0.8, the result must be validated, and it is still too low to allow any translational capacity.
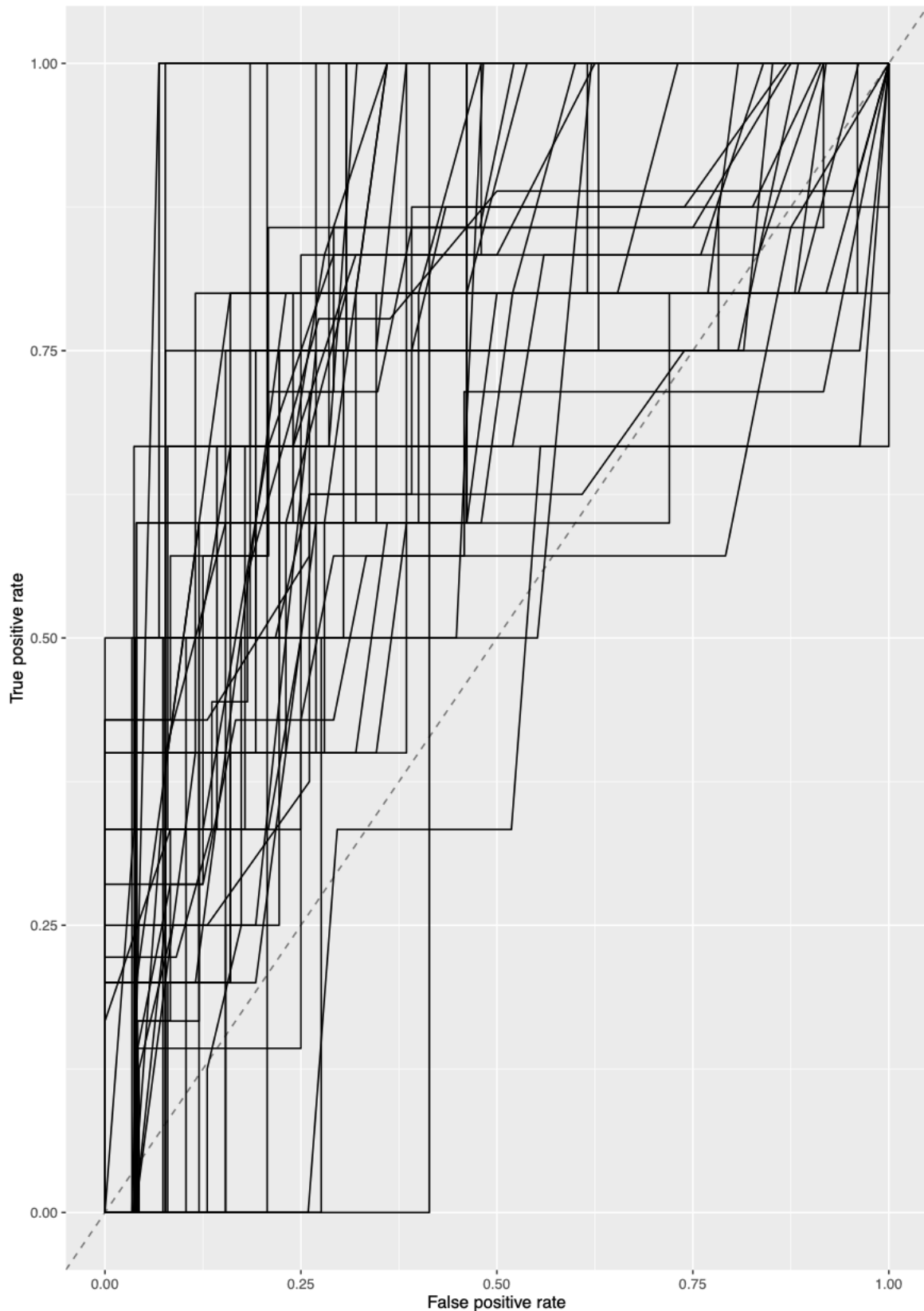
Fig 3.18 ROCs for each of the 50 models fitted as an iteration during the K-Fold cross validation procedure. As is clear from the plot, there is a large variation in the ROCs from all the models. This should be taken into account when considering the predictive power of the overall model. The aggregated AUC here is 0.74, which is considerably lower than the 0.8 absolute minimum expected.
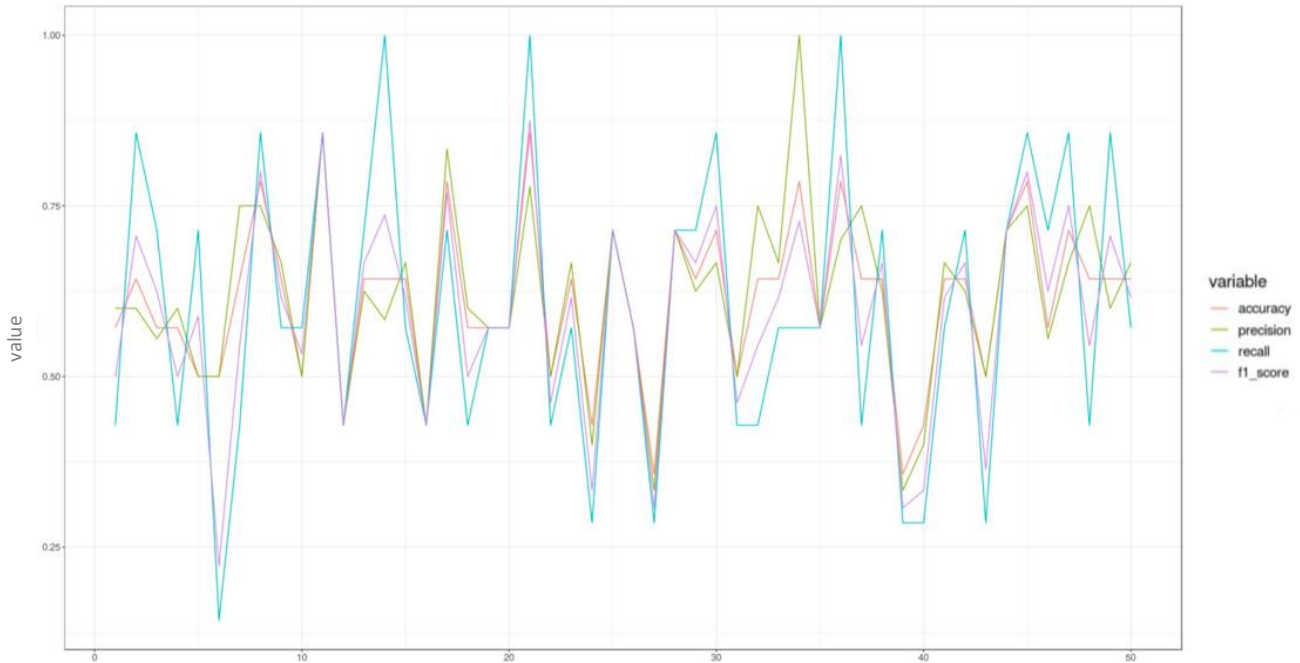
Fig 3.19 Precision metrics from K-Fold cross validation procedure. These include accuracy (red line), precision (green line), recall (blue line) and f1 score (purple line). As can be seen form the plot, there is a large amount of variation between each of the fitted models during this procedure. Sometimes the model does well with good accuracy and f1, but other times it performs very poorly.

of variation throughout the 50 consecutive models that were fitted. To display this further, performance metrics including accuracy, precision, recall and f1 score for each of the models were plotted (fig 3.19). The models have varied severely between datasets, where by sometimes it gains up to 80% predictive accuracy with good f1 scores, and other times it does quite poorly. This suggests that some of the test Mut samples are much more similar to the WTs than others.

As the machine learning algorithms didn't produce the desired results, additional factors, including the pathological origin of the tumour (oligodendroglioma, astrocytoma, or primary GBM), tumour volume and manufacturer of MR scanner were considered. First the data was visualised using heatmaps (fig 3.20) to observe if any of the categorical variables showed any correspondence with the data. None of these factors seemed to have a clear separation. Further to this, to describe the variance in the data, PCA was carried out on mean-centred and scaled data, and visualised coloured by each factor (figs 3.21 and 3.22). The scanner manufacturer seemed to make a difference, whereby the most pronounced alteration
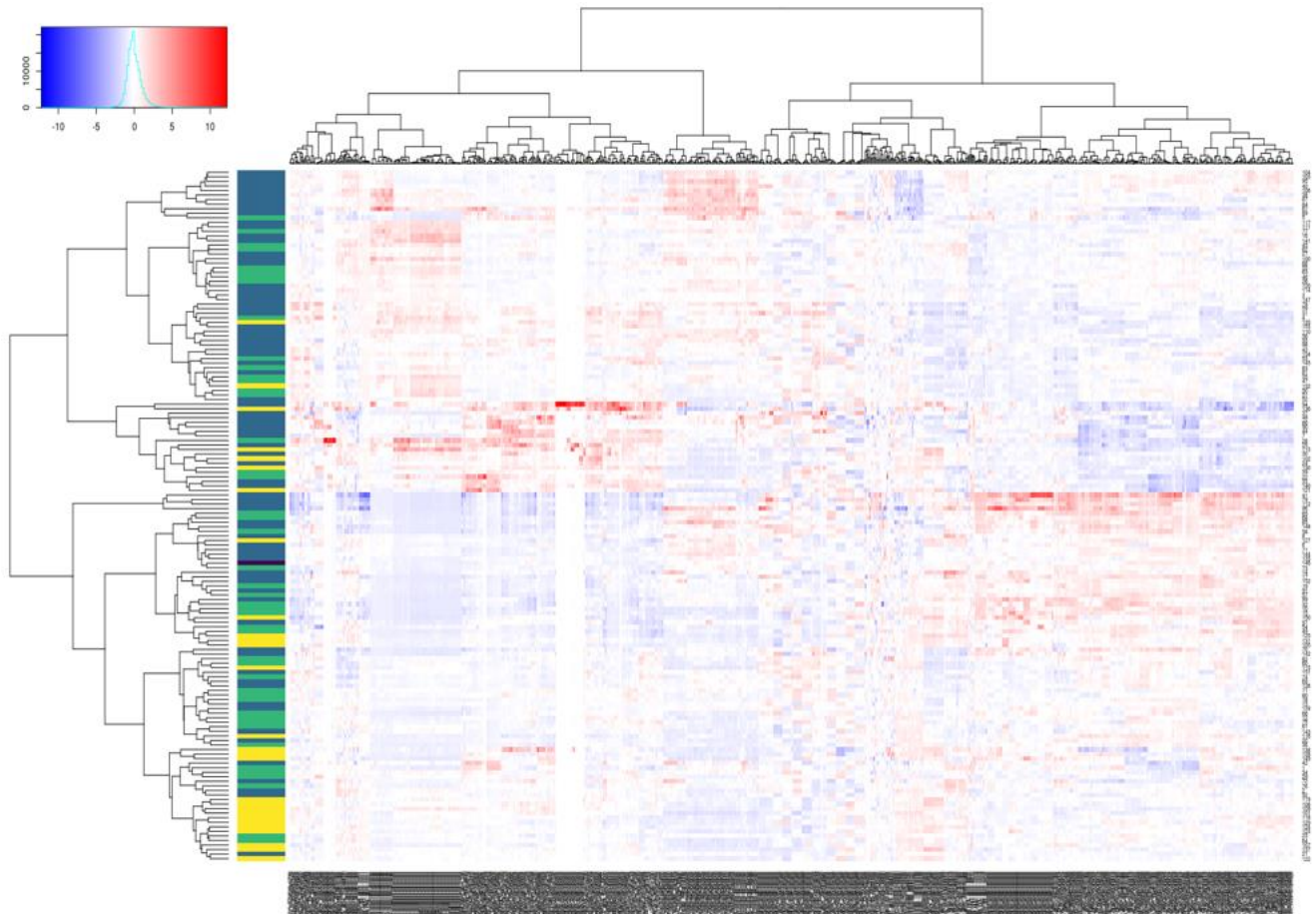
Fig 3.20 A heatmap of the data whereby the categorical data pertains to the scanner manufacturer. Blue represents the GE scanner, green Philips, and yellow Siemens. Hierarchical clustering has been applied. There is clear clustering of the data when this ordering has been applied.

was between GE and Siemens, although there was still a large amount of overlap between these two groups in the PCA plot (fig 3.21).

The PCA plot coloured by tumour size (20%, 40%, 60%, 80% and 100%) displayed that the smaller tumours tend to cluster and do not overlap completely with the larger tumours (fig 3.22). This suggests that some of the texture features were associated with tumour size, and therefore introducing a new, unwanted variance within the data. To assess the impact of this, Spearman correlations were calculated between each feature and the tumour size. Those features that had absolute correlation >0.5 were selected and the PCA was recalculated (fig 3.22). This showed that tumour size clearly influenced some of the descriptors.

As we established that a number of the features have a dependency on scanner manufacturer and tumour size, a further variable selection approach was pursued. A multiple linear model was fit to each variable and evaluated in terms of the variance that each factor
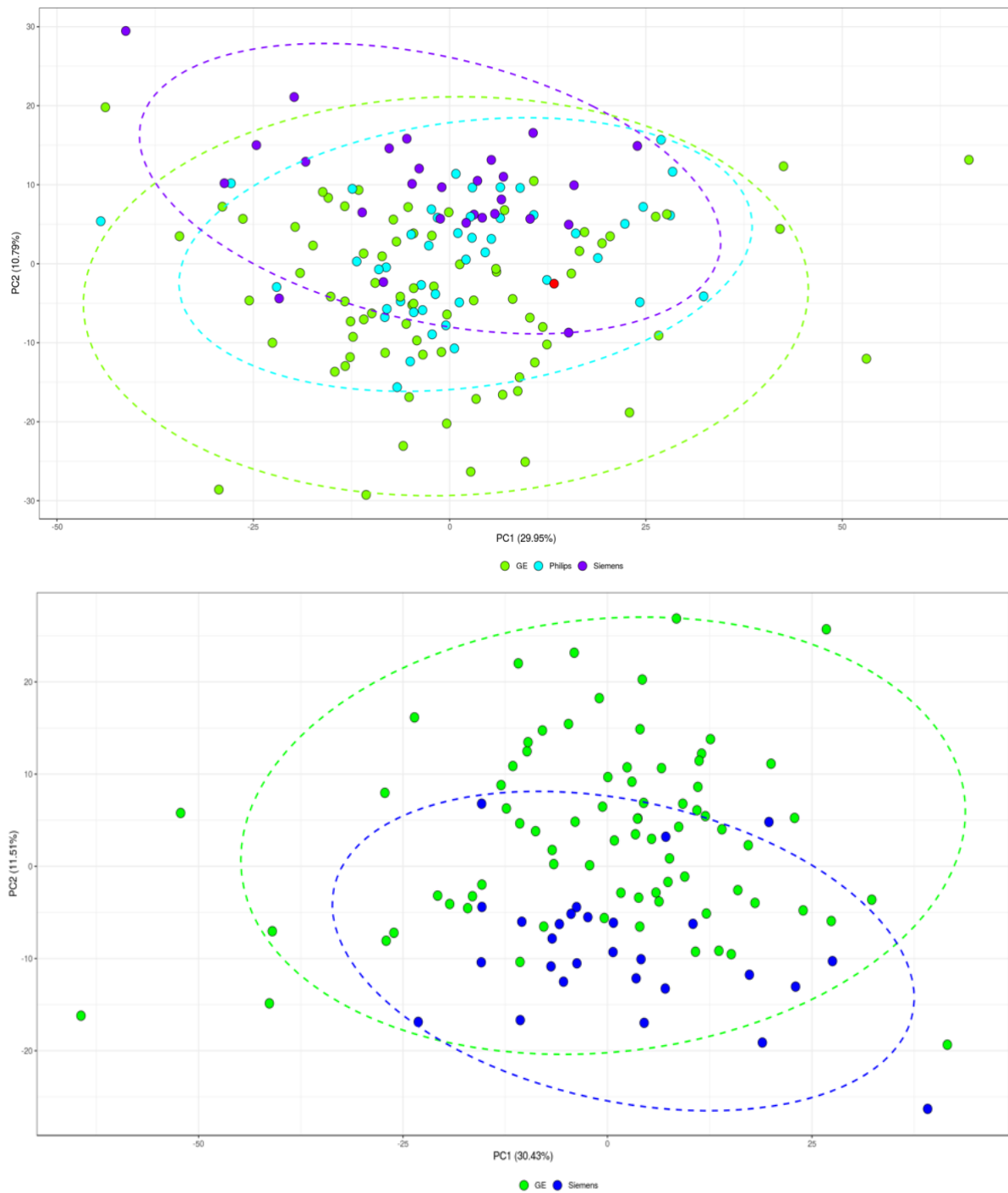
Fig 3.21 (a) PCA plot using scanner manufacture as the categorical variable. GE (green), Philips (light blue) and Siemens (dark blue). There seemed to be some clustering of point, particularly with reference to GE and Siemens. The red point is an unlabelled data point. **(b)** PCA plot, but with only GE and Siemens, there is some clustering, suggesting that scanner manufacturer has introduced unwanted variance within the data.
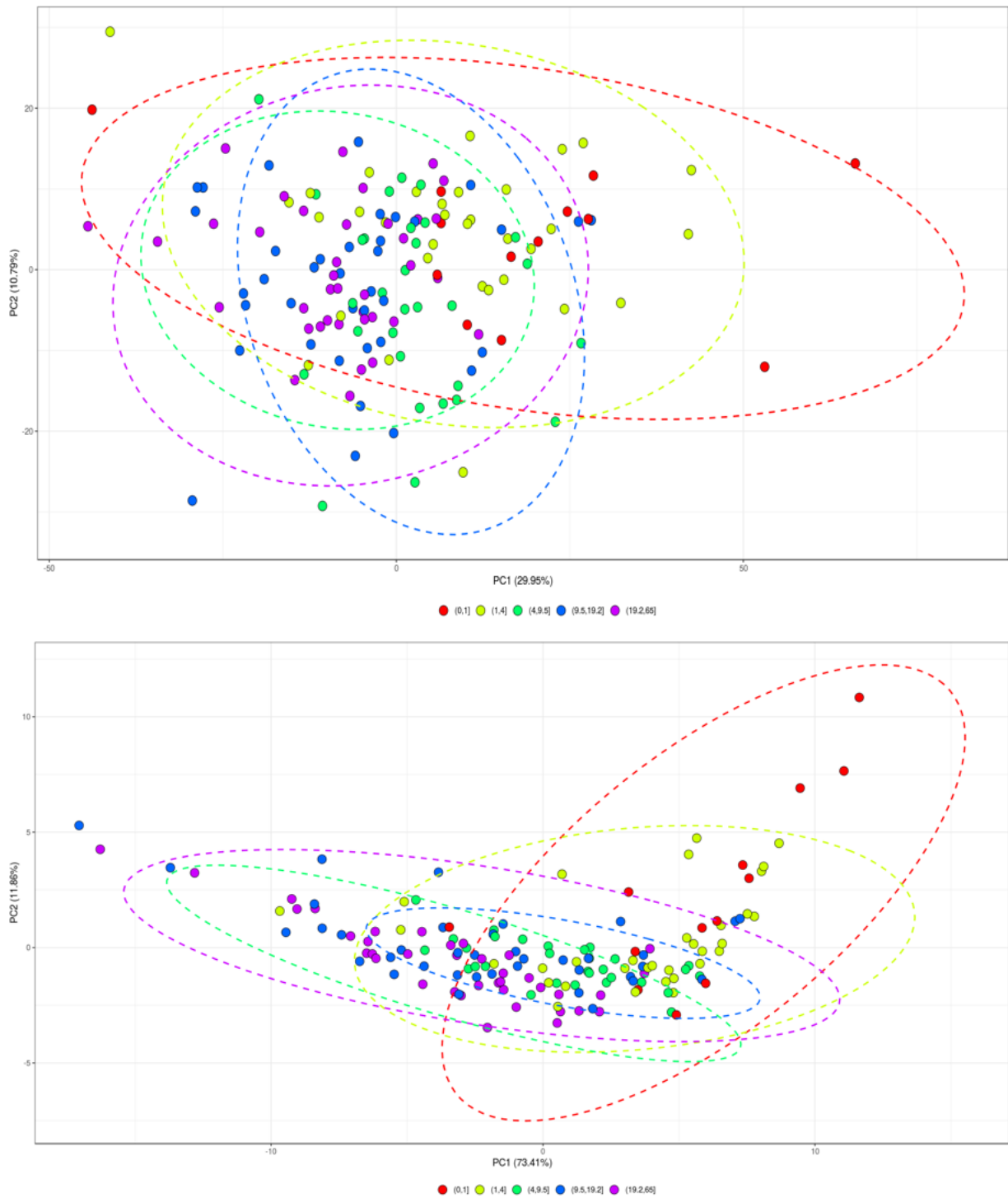
Fig 3.22 (a) PCA plot of the data coloured by tumour size (20%, 40%, 60%, 80% and 100%; coloured red, yellow, green, blue and purple respectively). It would appear that the smaller tumours tend to cluster and do not overlap completely with the larger ones. **(b)** Following spearman correlation calculation between each feature and tumour size, those features that had absolute correlations >0.5 were selected and the PCA redone and displayed here. From this we can deduce that tumour size clearly influences some of the descriptors.

a.

1	log.sigma.1.0.mm.3D_glcm_JointAverage
2	log.sigma.1.0.mm.3D_glcm_SumAverage
3	log.sigma.1.0.mm.3D_glcm_Autocorrelation
4	log.sigma.1.0.mm.3D_gldm_HighGrayLevelEmphasis
5	log.sigma.1.0.mm.3D_glrlm_ShortRunHighGrayLevelEmphasis
6	log.sigma.1.0.mm.3D_glrlm_HighGrayLevelRunEmphasis
7	log.sigma.1.0.mm.3D_glszm_SmallAreaHighGrayLevelEmphasis
8	log.sigma.1.0.mm.3D_glszm_HighGrayLevelZoneEmphasis
9	wavelet.HLL_gldm_DependenceEntropy
10	wavelet.HLL_glszm_ZoneEntropy
11	wavelet.LHL_glszm_ZoneEntropy
12	wavelet.LHH_gldm_DependenceEntropy
13	wavelet.LHH_glszm_ZoneEntropy
14	"wavelet.LLH_gldm_DependenceEntropy
15	wavelet.LLH_glszm_ZoneEntropy
16	wavelet.HLH_glszm_ZoneEntropy
17	wavelet.HHL_gldm_DependenceEntropy
18	wavelet.HHL_glszm_ZoneEntropy

b.
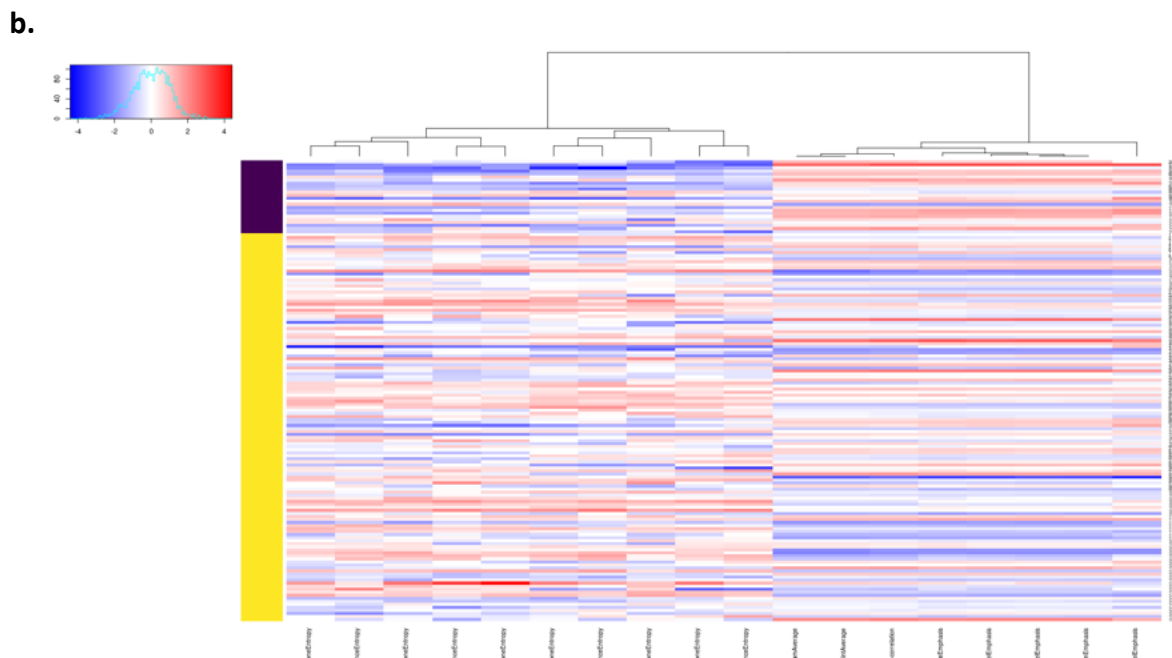
Fig 3.23 (a) A table to show the 18 texture features that were highlighted as important in the classification of IDH Mut from WT following multiple linear model fitting. **(b)** A heatmap of the 18 variables, whereby samples have been ordered by mutational status. There is clear grouping of the variables according to IDH status (blue and red).

Model Accuracy over 20 models

ROC area under the curve over 20 models
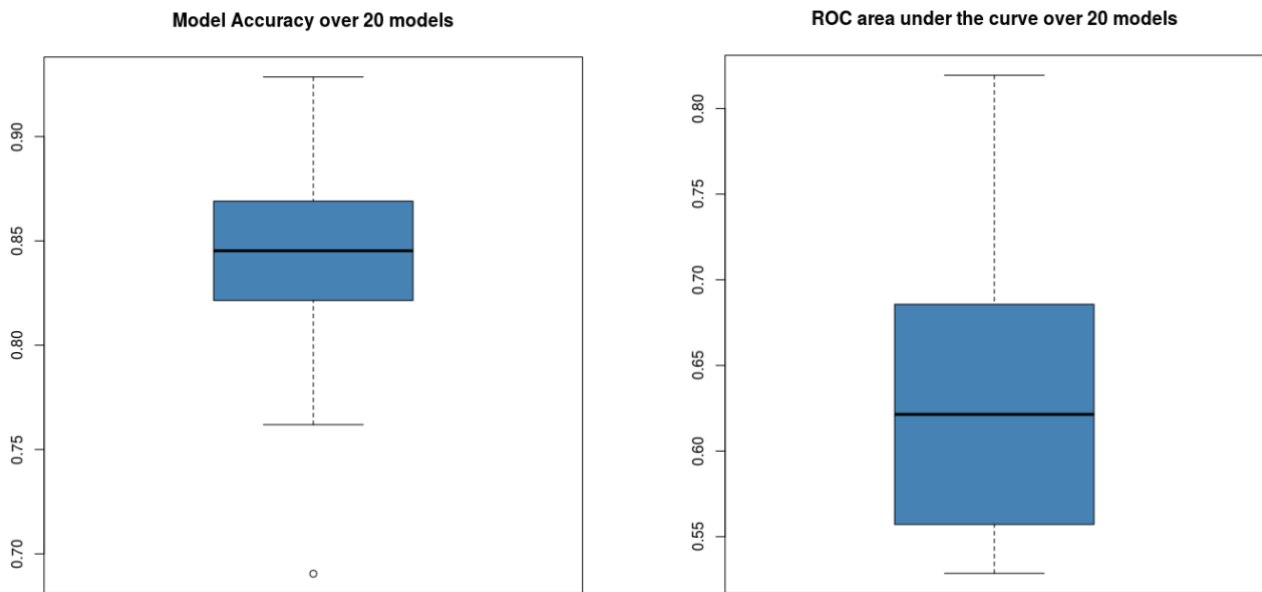
		Real labels	
		MUT	WT
Predicted labels	MUT	4	5
	WT	2	31

Fig 3.24 (a) Box plots to indicate the accuracy, and the AUC of the ROC for the 20 random forest models that were fit using the 18 variables identified. The results show a large variation in the AUC as well as accuracy. This suggests that there are samples within the data that are much harder to predict than others, which is often the case with studies that have low patient numbers. **(b)** Confusion matrix form the first random forest model fitted using the 18 features. This model achieved 83% accuracy, however this was somewhat inflated by the large imbalance within the dataset. The performance was achieved by weighting the importance of the Mut samples by 1000:1.

accounted for. This allowed the selection of variables where the IDH status was important but the other factors were not. From this we were able to identify 18 variables (fig 3.23a) that seemed to be important in the classification of Mut's from WT's. The data was visualised in a heatmap (fig 3.23b)where there is ordering of the variables into groups as we can see clustering of red and blue. Further to this, a random forest model was fit to the 18 features selecting a test set of 6 IDH Mut samples, and 36 WT samples. The model achieved 83% accuracy, however this value was somewhat inflated by the large imbalance within the dataset,

whereby there were many more WT samples than Mut. Nevertheless, those positive examples that did feature within the data were predicted reasonably well. The performance of the model was achieved by weighting the importance of the Mut samples 1000:1 at a cost of more false positives. Following this, 20 random forest models were then fit to random permutations of training/test data split to assess the stability of the predictions. Figure 3.24 shows the values for model accuracy and ROC AUC for each of the models as box plots. The results show a large variation in the AUC of the ROC which suggests that there are samples that are much harder to predict than others.

3.4 Discussion

Here we present results that have explored the use of texture features extracted from MR images of glioma, and the application of machine learning in order to classify images into IDH Mut or WT. Firstly a GA approach was taken for both feature selection and classification, however the resultant models were unsuccessful at classifying the images. We explored the use of random forests, again for both feature selection and classification. Although initial results indicated that this method was not successful at classification, we were able to identify several factors related to image acquisition that introduced unwanted bias into the data. Despite accounting for these extra factors, the final model showed variable performance upon cross validation, however this may be attributed to the low sample size, and an increase in patients within both testing groups may aid in building a representative model.

The paradigm shift towards personalised medicine in oncological treatment has perpetuated the need for detailed diagnostic information at the earliest possible stages of tumour detection. Medical imaging is routinely used as a primary investigation when assessing patients presenting with potential tumours, and it has the capability to provide a wealth of data pertaining to the intra-tumour heterogeneity throughout the lesion volume. Here, we have investigated the use of two methods for feature selection and representative model construction, to determine if radiomic features are able to distinguish IDH Mut from WT patients using standard imaging data.

There have been a number of previous studies that have investigated the role of radiomics for the non-invasive detection of IDH mutation. A study by Lu et al looked at a combination of post-contrast T_1, T_2 (Fluid attenuated inversion recovery)FLAIR, T_2 weighted

and diffusion weighted images (DWI)[116]. They reported that by using cubic support vector machine (SVM) and a quadratic SVM for glioblastomas and low grade gliomas (LGGs) respectively, they were able to achieve 95.7% sensitivity and 100% specificity for GMB IDH classification, and 85.7% sensitivity and 93.0% specificity for LGG IDH classification. These results, however, incorporated textures features calculated from all 4 contrast mechanisms, including DWI, which are not always routinely carried out on GBM patients for primary diagnostics in the UK. Furthermore, this study failed to report which of the texture features from the images were important for classification within the respective models. Further to this a study by Ren et al. also took a multiparametric approach, incorporating 3D arterial spin labelling (3D-ASL), T_2 FLAIR and DWI to carry out classification of both IDH mutation and ATRX expression[117]. They reported 100% sensitivity and 85.7% specificity for IDH mutation classification. Although the incorporation of advanced imaging contrast mechanisms such as ASL and DWI may aid in the discrimination of IDH Mut from WT, the clinical utility of these contrast mechanisms must also be taken into consideration. One of the major limiting factors in the acquisition of ASL and DWI is the time taken for each patient in the magnet. One goal that our investigation set out to achieve was to focus on the 'standard' imaging sequences (T_1 and T_2 weighted images) and determine if texture features calculated from these images are useful in the construction of a classifier for non-invasive IDH detection. During the data collection phase of this study it was observed that there was not a specific imaging protocol that was carried out on newly diagnosed patients. Each patient may receive a different number and combination of contrast mechanisms when scanned. Therefore, not every patient had undergone the same combination of post-contrast T_1, T_2 FLAIR, T_2 weighted imaging. Here we have been investigating the utility of single contrast mechanisms for the radiological classification of IDH mutation as it seems this would have more clinical translatability. With this in mind, one of the big challenges in the use of standard MRI imaging sequences is that these are acquired in arbitrary units that are not comparable between acquisitions, although the discretization step described in the methods section is designed to help normalise the intensities to aid in inter-subject comparison. One of the benefits of utilising functional and quantitative imaging methods such as ASL and DWI are that the units used in the calculation of features are absolute, such as blood flow or mean diffusivity.

There were a number of other limitations associated with this study. As the data has all be previously collected between 2016-2019 by clinical radiologists, we had no control over the

acquisition parameters or the quality of image acquisition. As such, the images that were included encompassed a range of resolutions, dependent on which scanner they originated from. To allow comparison of features generated from images from different cohorts as well as making features rotationally invariant, interpolation was carried out on all images. This equalised all the image resolutions, however the cost of this was that some images had to be up-sampled and other had to be down-sampled, dependent on the original resolution of the image. This, in turn, may cause partial volume effect within the sampled region of the image. Therefore any change in MR signal that may occur as a result of the presence of 2-HG within the tissue of a given voxel, that in turn causes a change in the texture features generated from the image, may be masked if that 2-HG-containing voxel is blended with a neighbouring voxel that has no such signal change. By this method, partial volume effect may suppress the classification power of this technique. One way in which to mitigate this limitation would be to carry out prospective imaging data collection with more standardised sequences that were sampled at the same resolution.

Another potential limitation was the use of such a high number of features for analysis. The approach of extracting 973 features from the images using a combination of filtered and unfiltered images was adopted as we had no clear sign as to what to look for that may indicate IDH mutation. It therefore stood to reason that extracting a large number of features in an attempt to fish for potential variables that could be discriminative would be a sensible approach. Retrospectively, working with such highly dimensional data, however, may have negatively impacted the discriminative power of the experiments, as potentially useful variable could have been masked by the large number of uninformative ones. One way in which to circumvent this limitation would be to generate less features from the images in the first instance. For example, reducing the number of filters applied, as the results here show that only features from the LoG filter with a sigma value of 1, and the wavelet bandpass filters produced features that were relevant to IDH classification.

Finally, computing power required in order to attempt to extract different types of features from the images also acted to limit the scope of this investigation. Texture features are relatively simple to extract, using third party software, and they are tractable, as they are based on established mathematical definitions. However, the rigid way in which these imaging features are calculated may exclude inter-relationships between greyscales that could be present due to IDH mutation. One way to probe for patterns within the images that are

undefined is to use deep learning methods, such as convoluted neural networks (CNNs).There have been a number of studies that have shown the benefits to using hybrid deep learning approaches that carry out both tumour segmentation, along with feature extraction and image classification[118–120]. The value of these methods is rooted in several niches. Firstly, they are end-to-end, in that the resultant models are able to take normalised images, and carry out all radiomics steps, without user input, and produce a classified image. Further to this, the features upon which classification is based are not ridgid, and they are therefore able to search for patterns within the images that may have previously been overlooked. One of the major drawbacks to the study presented here, and why such methods were not pursued, was in part due to the low sample size, as deep learning requires very large datasets in order to train and optimise the model. Additionally, these models require very large amounts of computing power, often clusters, in order to run the algorithms, an infrastructure that our lab did not have in place. One way in which to overcome this and strengthen this investigation in the future would be to form a collaboration with another centre to pursue a deep learning approach. Already having the collaboration with the Walton Centre NHS Foundation trust is a great way to access a large amount of clinical data, a prospect that may serve as very attractive to a mathematics department looking to collaborate.

A study by Zhang et al also looked at the ability of texture features calculated from multimodal MR imaging for IDH classification, the combination of T_1, T_2 FLAIR, T_2 weighted and diffusion weighted images (DWI) contrast mechanisms, however they failed to report a number of crucial image standardisation steps, including grey level intensity discretisation and image interpolation[121]. In this study we have strictly adhered to the processing guidelines set out by the IBSI, we have also endeavoured to report fully all steps and parameters chosen to facilitate the reproducibility and replicability of the study[81]. Other studies have reported good classification ability of models but have failed in the full and proper reporting of how images were handled before and during the calculation of features, hampering the reproducibility and therefore translational capacity of such models. These include a study carried out by Yu et al[122] whereby all image pre-processing steps such as interpolation and discretisation were not discussed or reported. Similarly, the study by Lui et al[123] reported the use of images with non-isotropic voxels, but failed to report the use of interpolation, which may confound the overall results as calculation of features from non-isotropic voxels renders them no longer rotationally invariant. One particular criticism of radiomics is the distinct lack of reproducibility. The

complex methods that underpin the practice of radiomics necessitates the comprehensive reporting of methodological parameters used, as well as external validation.

3.4.1 GLAGO for feature selection and classification

Here we have used a GA approach to select radiomic features derived from T_1 post contrast whole brain and tumour segmented images, as well as T_2 whole brain images. We wanted to explore the possibility of the computation of discriminative features being calculated from the whole brain as well as just the tumour, as it has been shown that diffuse microstructural damage occurs across the cerebrum in the presence of infiltrative glial cancer cells[124,125]. These microstructural changes, that have previously been demonstrated by diffusion imaging, may induce changes that alter the relaxation of the spins to facilitate alterations within the distribution of grey scale intensities within standard T_1 and T_2 imaging that remain undetectable by the human eye.

Figure 3.3 illustrates the process of the GA approach for feature selection prior to classifier construction. First, we assessed the ability of radiomic features derived from contrast enhanced whole brain T_1 analysis to discriminate IDH Mut from IDH WT patients. Following the GA feature selection process, it is possible to plot the evolution of the fitness function of the constructed chromosomes. Although the fitness function, which is a derivative of the classification error of the variables within the chromosome at distinguishing IDH Mut from WT, showed that we were able to surpass the pre-defined classification error value of 0.8. Although the initial fitness function plot indicates that the classification error of the chromosomes was greater than the minimum expected value following only 69 generations, this metric supplied limited information regarding the ability for the variables within the chromosomes to be able to classify the images accurately. To inspect this further, the mean sensitivity and specificity of the chromosomes must be calculated. When the sensitivity and specificity of both WT and Mut classification was examined, it was clear that the performance of the models in distinguishing a Mut glioma patient was poor (fig 3.7). We postulated that this imbalance in classification specificity may be attributed to the imbalance in class sample size, as there are far fewer Mut cases when compared to WT cases. To this end, we applied a subsampling scheme to artificially balance the contribution from both classes. This approach did not make much difference to Mut classification, as there was a negligible increase in the sensitivity. Moreover, there is evidence to suggest that the K-nearest neighbour (KNN) algorithm is unaffected by an

imbalance in classes, as it takes into account the class distribution of a wide region of probabilities to predict instance labels[126]. Despite the poor sensitivity described by the confusion matrix plot, the forward selection procedure was followed to evaluate if a representative, or close to representative model, may be constructed from the features that have been shown to be most discriminative following the initial GA inspection. Figure 3.8 illustrates the disparity between Mut and WT classification. From these results we can determine that we were unable to build a reliable and representative model for IDH classification based on these features derived from whole brain analysis.

Further to this, we used the same approach to assess if we are able to classify the patients using data derived from whole brain T_2 images. In these experiments the sensitivity of both WT and Mut classification for the chromosomes generated by the GA, calculated following the bootstrapping method was below the minimum expected value of 0.8. We wanted to understand the variability in the chromosome classification ability further, and continued with the forward selection procedure. From this we were able to deduce that not only did the best preforming model fail to surpass the minimum classification error boundary, but there was a huge amount of variation in the models that were proposed by the GA.

Although calculating features over the whole brain was implemented in an attempt to capture some of the global changes that have been reported to occur in the presence of glioma, this inclusion of such a broad area from which to calculate features may have contributed to the poor classification. Whilst we understand that microstructural changes that occur during gliomagenesis and disease progression can be measured using diffusion MRI, such diffuse changes are minor when compared to the remaining amount of healthy brain tissue[127,128]. Therefore, since it is in fact the metabolic implications of the presence of IDH mutations within malignant cells that may facilitate the alterations in spin environments that in turn alter the distribution and relationship of signal intensities following Fourier transform of the MR signal, the inclusion of such a large amount of healthy tissue in which these metabolic changes are absent, may have introduced a large amount of noise/redundancy into the data. The presence of abundant healthy tissue may be masking the effect of IDH mutation on grey level distribution. There may be a number of other possible reasons as to why this approach failed to classify the samples with high enough accuracy for consideration of clinical translation. These data are highly dimensional, whereby the number of observations far exceeds the number of samples within the dataset. This curse of dimensionality leads to large

amounts of redundancy and overlap within the data itself. The great number of variables leads to a large search area for any algorithm, however the GA feature selection is optimised for such studies, as the multivariate method tests all variables in combination with each other by arranging such variables into smaller, more manageable groups of chromosomes.

To further investigate a potentially more robust method of detection, we generated the same texture features from within the tumour. More precisely, we segmented the tumours such that only the enhancing regions within the tumours were included for feature generation. We set out to include only the enhancing region of the tumour for a number of reasons. Firstly, this is the most active part of the tumour, whereby an abundance of proliferating and infiltrating glioma cells can be found. The other regions within the centre of the tumour that do not enhance may include a varying amount of active and proliferating cells, alongside necrotic regions, cystic components and blood vessels. By including only the enhancing region, we are reducing the heterogeneity of the data, and therefore reducing the potential effect of masking features that may arise due to the presence of IDH mutation. Although radiomics is heralded for its ability to assess heterogeneity, as we have seen from the whole brain analysis, since this investigation is focused on finding features that correlate with a specific genetic mutation, inclusion of heterogenous regions of the tumour and/or healthy brain that are known not to incorporate active cells with IDH mutations may, indeed, mask potentially discriminating features that could be important in classification. Further to this, one practical advantage to inclusion of only the enhancing region of the tumour is accuracy and reproducibility of tumour segmentation. Due to lack of radiology training it was postulated that the best method of segmentation would be to include only the enhancing region, so as to eliminate the potentiality of inclusion of non-tumour regions. The results of this line of investigation showed that following GA feature selection, the classification accuracy for Mut classification showed a sensitivity of only 0.181, compared to the WT classification sensitivity of 0.951 (fig 3.11). The disparity between Mut and WT classification using this approach was inflated when compared to that of the whole brain analysis. In addition, when the forward selection procedure was followed to identify the model that performed best at classification, the plot in figure 3.11b further demonstrates the disproportion between WT and Mut classification. The forward selection procedure also demonstrated the importance of understanding the calculation of mean classification errors when evaluating model performance, as at first glance the best performing model appeared to achieve a high

classification accuracy of 0.88, however this is calculated by dividing the number of misclassified samples by the total number of samples. As we have noted that there is a much larger number of samples in the WT class, and the sensitivity of classification was very good in this group, this will introduce a large amount of bias into the overall error calculation. The inflation of the difference between Mut and WT classification when compared to the whole brain analysis may be attributed, not to the imbalance in sample numbers within each class, but rather to the low number of samples within Mut class itself. Due to the low number of samples, the model simply doesn't have the opportunity to learn any dependency of feature presence for class labelling. It has been previously suggested that whilst K-nearest neighbour (KNN) is less sensitive to imbalanced datasets, the results of classification can vary amongst datasets with lower sample sizes[129]. The lower number of samples of the segmented tumours was due to the required elimination of certain patient data due to difficulties during segmentation. Here we had to discount 26.5% of the T_1 post-contrast images. The difficulties include the cases when the enhancing region of the tumour was too small and when there was no clear distinction between the tumour and other enhancing tissue such as meninges and blood vessels.

Taken together, utilising a GA for feature selection and model construction in this capacity did not yield results that represent a clinically translatable model. The failure of T_1 and T_2 whole brain imaging features to reach overall classification accuracy surpassing the minimum expected value, and the disparity between Mut and WT classification errors across all analyses, but in particular the T_1 intra-tumoural features, has warranted that alternative routes of investigation are to be pursued.

3.4.2 Random Forests for feature selection and classification

The previous experiments explored the capacity of GA approach to feature selection to find the optimal feature subset for model construction. The GA approach belongs to the wrapper method family of feature selection algorithms, and whilst this group of techniques often execute well in providing the best performing feature set for a particular type of model, they are inherently computationally expensive, as for each subset of variables a new model must be trained and evaluated. An alternative multivariate approach to feature selection and eventual classification would be to investigate the capabilities of an embedded method. In embedded techniques, the feature selection algorithm is integrated as part of the learning

process[130]. One of the most widely used embedded methods is the decision tree algorithm, and here we have used the popular random forest, implemented in the 'mlr' package in R[131].

The random forest classifier is an ensemble method that works by constructing a vast number of uncorrelated decision trees based on averaging a random selection of predictor variables. One of the key concepts of random forests in performing as an embedded method for feature selection as well as a robust classifier, is that it is able to construct many uncorrelated models. The critical concept here is that the models are uncorrelated, the more independent observations we are able to make, the higher the chances of making correct predictions. Each individual decision tree error is protected by all the other trees in the forest, whilst some trees may make the wrong prediction, many other will be right, and so the predictive power of the population is able to advance in the correct direction.

Prior to the initial stage of this investigation, we plotted the raw texture feature data as a heatmap, and carried out PCA to quickly analyse if there was any correlation between the vast amount of features and IDH mutation. Unsurprisingly, and most likely attributed to the high dimensionality and presence of redundancy within the unfiltered dataset, there was no pre-existing correlation between these features and mutational status.

An initial random forest model was trained on a 2/3 train 1/3 test split of the data, and the ROC for the model was plotted in figure 3.17. Upon inspection of the ROC, the AUC is calculated at 0.796 only just below out minimum accuracy threshold of 0.8. There is, however, a requirement to cross validate this result, to ensure that the model is detecting correct patterns within the data and it is not picking up too much of the noise or overfitting. The rationale behind the choice of K-Fold cross validation as a method of evaluating the model performance is due to the low sample size of our dataset. If we were to remove large parts of the data for validation purposes, then we risk losing potentially important patterns within the data, and therefore increasing error induced by bias. This method divides the data into k subsets, otherwise known as folds. From this, each one of the k subsets is used as the validation set, and the other k-1 subsets are amalgamated to form a training set. The plot in figure 3.18 illustrates the high variability between all the models that were fitted using this method of validation. This variance in our model outcomes is likely due to overfitting. It is a reflection on the lack of generalizability of the model, as it illustrates how dramatically different the model would perform if we were to train it on different datasets. This overfitting is due to the model identifying patterns and complexity within random noise within the data, when what we are

actually looking for is a real trend. To further illustrate this, performance metrics from each of the cross validation iterations were plotted in figure 3.19, and it is obvious that the variation in model performance between each split of the data is large.

As it would appear that random forests seemed to be overfitting, potentially finding patterns within uninformative noise, or even patterns within the data that may be attributed to some other form of variance that may be present within the data due to factors other than IDH status, we decided to investigate if there was any correlation with the metadata. These metadata included pathological origin (oligo, astro, GBM or secondary GBM), scanner manufacture and size of the tumour. When visualised in heatmaps and following PCA calculation, it was evident that both tumour size and scanner manufacture were having an impact on the distribution of the data. The correlation between tumour size and a number of the features is logical, as although all the texture matrices may be both rotationally and translationally invariant, none of the matrices are scale invariant. It therefore stands to reason that there would be association of tumour size with texture matrix features[81].

In addition to tumour size, scanner manufacturer was also found to have an impact on the distribution of the data, whereby the most pronounced difference was seen between GE and Siemens. Since the recent surge in popularity of radiomics studies employing MR images for correlation with clinical endpoints, molecular and genetic information, there have been a number of studies that have examined the impact of varying acquisition parameters on the calculation of radiomics features. Studies such as those by Yag et al[132], Saha et al[133] and Ford et al[134] have shown that a varying number of features may be influenced by acquisition parameters. Between these studies there was no consensus on a definite list of parameters that are influenced by acquisition parameters and by what degree, however this variability that has been observed in the calculation, calls for caution to be taken when carrying out the machine learning stages of any radiomics workflow, particularly those ones that concern MR images. With this in mind, and following the observation that when we group the samples according to scanner manufacturer following PCA we are able to observe clustering, we decided to investigate each variable and evaluate in terms of the variance that each factor is accounted for. This allowed us to curate a list of 18 features that may contain discriminative information to classify IDH Mut from WT. Despite the more robust feature selection approach, and the dramatically reduced dimensionality of the data following assessment via multiple linear models, the AUC of the ROCs from 20 validation iterations had a large variability, ranging

from 0.47 to 0.82. This may be attributed to the low number of cases included in the fitting of the model. As opposed to the likely overfitting of random noise within the vast number of features that we experienced in the first random forest model, here there are simply not enough examples for the model to learn clear dependencies between the features and the outcome, and some cases that are rare within the data end up being learned as well. This model still seemed to do better than random, which suggests that there may be information within the features that could be useful for determining mutational status, however it will require more data to achieve useful performance.

3.5 Concluding Remarks

There were a number of limitations that were experienced throughout this study that were unavoidable, but must be considered when drawing conclusions. Foremost is that the imaging data that were used to generate features were retrospective. As we had no control over the acquisition of the images, 30.4% of T_1 post-contrast and 25% of T_2 images were immediately disregarded from the original cohort due to poor image quality. Furthermore, as we have already seen, all images were acquired across three different magnet vendors over a number of years. This is inherent of clinical data, as it is often the case that each hospital will house magnets from a number of manufacturers that are of different ages. We have noted that this may serve as a limitation to a radiomics approach, however, if we are to move the field forward towards building clinically translatable models that may be used in patient diagnostics, then we must take into account that each individual hospital will have a number of different scanner types. To this end, the relationship between scanner manufacture and acquisition parameters should be further explored. If we are able to understand more about the dependencies of particular imaging features within these factors, then perhaps these can be taken into account when attempting to construct models that can be used in multicentre investigations. Moreover, to fully assess the potential of this technology, first, our own investigation must be expanded, to validate if the trend that we have observed here, that there may be some information contained within the features that allows discrimination of the Mut and WT images, can be confirmed as a true classification. Next a multicentre, prospective study would be followed, whereby acquisition parameters are standardised as much as the practical limits of individual clinicians' approaches will allow, and data is collected from a number of sites.

4. NMR Metabolomics to Identify a Metabolic Phenotype in IDH Mutant Glioma

4.0 Synopsis

NMR spectroscopy in combination with a metabolomics analytical approach has the potential to provide a wealth of data pertaining to the metabolic state of a sample of tissue. This study has explored the use of both 1 dimensional (1D) and 2 dimensional (2D) spectroscopy to extract and collate a body of information that can allow us to further our understanding of how the mutation contributes towards the development and progression of the disease, as well uncovering consistent metabolic shifts that can serve as non-invasive markers for the detection of the mutation *in-vivo*.

For this study glioma tissue samples were collected during surgical resection and immediately snap-frozen in order to preserve the metabolic profile. Methods for the extraction of metabolites, as well as the analysis of spectra had to be optimised, as this was the first study to incorporate this type of tissue to be conducted in our centre. Combining metabolomics information produced by both 1D homonuclear and 2D heteronuclear NMR is something that has seldom been explored in the context of IDH mutation in glioma, however here we have demonstrated that merging results from these two investigatory avenues can strengthen the accuracy and precision of the observations. Each approach may help guide the other and vice versa, allowing one to build a strong body of evidence as to which metabolic shifts are present as a result, or in combination with the mutation.

The following results suggest a consistent metabolic phenotype with respect to a discrete group of molecules. They provide rationale for further investigation into how these shifts facilitate the overall clinical phenotype observed in IDH Mut glioma patients, which will allow is to deepen our knowledge of the underlying mechanisms of the pathology. This metabolic pattern may also provide metabolic alterations to look for on *in-vivo* spectra, potentially providing the grounds for investigation into non-invasive biomarkers for the detection of the mutation using MRS.

4.1 Introduction

4.1.1 NMR Metabolomics

Metabolomics is a relatively new addition to the '-omics' field of study, and focuses primarily on the downstream effects of the combined environmental, genetic and proteomic distinctions between cell/tissue types. By assessing which metabolites are present and their abundance, metabolomics studies can provide a myriad of information pertaining to the intracellular status

of a given tissue type, and facilitate the discovery of both biochemical signatures and biomarkers for specific disease states[135]. In addition to biomarker discovery, metabolomics provides a global view of cellular metabolism, and therefore can offer further insights into the pathophysiology of diseases such as glioma. Significant alterations in metabolite levels that can be observed in distinct tissue types may be indicative of modifications in a plethora of biochemical pathways, the detection of which may provide information on the pathogenesis of the disease[136,137]. Changes in gene expression, genetic mutations and protein dysregulation can have influences on metabolite flux, which in turn will alter the metabolic environment of the cell and dictate phenotypic expression[138]. Assessment at the metabolic level therefore seems a logical analytical pathway, as it reflects potential aberrations at all levels. The study of metabolomics provides both qualitative and quantitative information that is able to describe the biochemical profile of a given tissue[139].

Nuclear magnetic resonance (NMR) spectroscopy may be applied for the identification and quantification of chemicals within complex mixtures. This technique has been a popular method of probing metabolism since the early pivotal study that saw its application in assessing ethanol metabolism using isotope-tracer analysis in the early 1970's[140]. Since this point, NMR spectroscopy has been used in countless studies probing metabolism of all manner of living creatures, *in vivo* and *in vitro*, using ^{1}H, ^{13}C and ^{31}P nuclei[141]. The technique is widely popular in metabolomic studies due to it being quantitative, reproducible and the spectra contain a wealth of biological information[142–144].

The technical details of both 1D and 2D spectral acquisition have been covered in chapter 2, however this section will discuss the differences, advantages and disadvantages of using 1D and 2D NMR spectroscopy to carry out metabolic profiling. As with all techniques that exploit the NMR phenomenon, acquisition of optimal signal to noise ratio in order to generate a spectrum following may be obtained in a matter of minutes, however in order to extract meaningful information from spectra to answer complex biological questions, we must obtain a dataset with a high signal-to-noise ratio, which involves increasing the number of averages and therefore increasing acquisition time. A typical 1D *in vitro* NMR experiment takes approximately 30 minutes, and due to the advent of automated sample changers for liquid state NMR tubes, it is possible to collect data on hundreds of samples in a single experiment[141]. Metabolite concentrations are inherently dynamic throughout biological samples. Additionally, there are often a wide range of concentrations that may be considered physiologically normal

for a specific metabolite within a given tissue type[145]. These factors provide a source of bias when carrying out statistical analysis of 1D NMR data, as the resonant peaks produced by the more abundant metabolites often overlap and hide the NMR peaks from the less concentrated metabolites. A number of methods have been employed in an attempt to circumvent this spectral overcrowding, including 'targeted' metabolic profiling by fitting spectra to a library of known metabolites in a 1D NMR database[146]; on- or off-line sample fractionation and the use of isotopically enriched metabolites[147,148]. All of these techniques present with distinct limitations when considering quantitative interpretation. Targeted metabolite fitting requires *a priori* knowledge of the system in question, and isotope labelling is not widely used in human studies.

Multidimensional approaches to NMR acquisition have been utilised in methods such as chromatography, electrophoresis and NMR spectroscopy in order to increase resolution[149]. Both homo- and heteronuclear NMR experiments , including total correlation spectroscopy (TOCSY), heteronuclear multiple bond correlation (HMBC) and heteronuclear single quantum coherence spectroscopy (HSQC), until recently were performed exclusively on selective proteomic samples for structural elucidation[150]. Popularity in the use of these techniques for high resolution NMR began to gain traction in the early 2000's when groups such as Dumas et al. carried out 2D HMBC NMR on bovine urine samples that had been exposed to steroids[151]. When considering the challenge of increasing resolution, heteronuclear experiments exploiting the spin property of ^{13}C may provide an attractive method of dealing with spectral overcrowding. The chemical shift range of ^{1}H NMR is relatively narrow at 0-12 ppm, whereas the ^{13}C chemical shift range observed in natural products is over 0-200 ppm, therefore acquiring signal in the second dimension using carbon can be of great benefit in inducing peak dispersion. One significant obstacle in the acquisition of ^{13}C spectra is the very low natural abundance of ^{13}C at around 1.1%. This makes HMBC an especially insensitive experiment with a long acquisition time. To overcome the issue of reduced signal and long acquisition when acquiring on the carbon, here we have used a HSQC pulse sequence, where magnetization is transferred from the proton to the carbon via an insensitive nuclei enhanced by polarization transfer (INEPT) step and then back onto the proton via a retro-INEPT step following time t_1 (as discussed in further detail in chapter 2).

4.1.2 Metabolic Phenotype in IDH Mutant glioma

As discussed in previous chapters, gliomas present a significant proportion of malignant intracranial tumours, and the presence or absence of a mutation in the IDH1/2 gene is deterministic of both diagnosis and prognosis[7]. The gain-of-function neomorphic enzymatic activity is well known to facilitate an aberrant accumulation of the oncometabolite 2-hydroxygluarate (2-HG) within the cytoplasm. Although the presence of the oncometabolite has been suggested to contribute towards the malignant progression of the tumour, the mechanisms by which this process occurs are still widely debated[152]. Furthermore, the means by which the build-up of 2-HG confers enhanced prognosis and increased response to chemoradiotherapy whilst at the same time driving malignant progression, remain to be fully elucidated. Evaluation of the global metabolome of IDH mutated tumours has the power to provide us with information pertaining to the upstream affects that facilitate this juxtaposition of phenotypic manifestations. If we are able to understand the consequence of the mutation and subsequent intracellular build-up of 2-HG on processes such as gene expression, transcription and intracellular trafficking, then we may begin to understand the mechanisms behind disease progression, and the formulation of targeted therapies becomes a much more realistic prospect.

4.1.3 Metabolism as a Biomarker

Mutations in IDH1/2 catalyse the production of the oncometabolite 2-HG via reduction of αKG. The subsequent accumulation of the oncometabolite is facilitated by the heterozygous manner of the mutation, and the dominant effect that the gain-of-function mutation imparts over the remaining WT allele, which serves to provide the mutant enzyme with a continuous supply of αKG to be aberrantly reduced into 2-HG[25] (Fig 1.1). It is well reported that these mutations are present in approximately 70% of astrocytomas and oligodendrogliomas of WHO grade II-III, and ~12% of GBM of which are mostly secondary GBMs[7,153]. There is no other known cause for the production and accumulation of 2-HG in the brain, and it may be considered the only non-invasively detectable biomarker that is a direct consequence of genetic alteration known in humans.

Here we have assessed the metabolic profile of glioma samples both with and without IDH1/2 mutations, in order to determine if there is an overall metabolic signature that may be

linked to the mutation and the accumulation of oncometabolite. The acquisition and interpretation of *in vivo* spectra for the detection of 2-HG is technically challenging, due to the large amount of spectral overlap of the 2-HG peaks by metabolites found in great abundance in the healthy and diseased brain. If we are able to identify other metabolic aberrations that are detectable via NMR, then it may be possible to form a metabolic fingerprint in order to non-invasively probe for the mutation, as well as to further study the pathophysiology of the disease.

4.1.4 Scope of Study

This study explores the use of a metabolomics approach to 1D and 2D NMR spectroscopy in order to identify a metabolic phenotype in IDH Mut glioma. The observation of consistent metabolic shifts in relation to IDH mutation can aid in furthering our understanding of how the mutation contributes towards pathological development and progression, as well as providing a potential metabolic fingerprint that may translate into a non-invasive biomarker to the mutation.

4.2 Methods

4.2.1 Sample Collection

Patients for this study were identified at clinical multi-disciplinary tumour meetings at the Walton Centre NHS foundation trust. The study was approved by the Walton Centre Research Tissue Bank research ethics committee (REC) approval (Reference number 15/WA/0385). Patients were included if they were a suspected of a newly diagnosed glioma from radiological studies and were to undergo surgical resection. Inclusion criterion was that the patients must not have had any previous surgery or cancer treatment. Samples were collected during surgical resection and snap frozen in liquid nitrogen immediately following removal from the brain. Samples were kept at -80°c until metabolite extraction. All patients had provided informed consent and the use of the tissue for the study was approved by the NHS foundation trust and sponsored by the University of Liverpool. A total of 22 patients, 18 IDH Mut and 9 IDH WT, were included (table 4.1).

4.2.2 Metabolite Extraction

Metabolite extraction method was optimised by carrying out both a perchloric acid extraction and acetonitrile extraction on mouse brain samples in order to determine the method with the least intra-observer variability. Snap frozen mouse brain samples were divided into three approximately equal size sections of wet weights 400mg-900mg. For percholoric acid extraction, ice cold 4% percholoric (1:4, w/v) acid was added to each sample. Each sample was sonicated under ice cold conditions for 3x30 second bursts to prevent heating. The liberated samples where then vortexed for 1 minute, and centrifuged for 10 minutes at 4°c. The supernatant was collected into a second Eppendorf and lyophilised overnight. To the lyophilised material 178µl 100% D_2O, 20µl 1mM sodium phosphate pH7.4 in 100% D_2O, 0.2µl 100mM TSP in 100% D_2O and 0.2µl 1.2mM NaN_3 in 100% D_2O was added. The mixture was then vortexed for 1 minute. NaOH was then added dropwise to each sample to neutralise the pH (7.0-7.5). The neutralised mixture was then vortexed again and centrifuged at room temperature x120000g for 2 minutes.

For acetonitrile extraction, ice cold 50% HPLC grade acetonitrile in double distilled water was prepared and added to each sample (1:4, w/v). each sample was sonicated under ice cold conditions for 3x30 second bursts to prevent heating. The liberated samples we then vortexed for 1 minute followed by centrifugation at 4°c for 10 minutes. The supernatant was collected into a second Eppendorf and lyophilised overnight. To the lyophilised material, 178µl 100% D_2O, 20µl 1mM sodium phosphate pH7.4 in 100% D_2O, 0.2µl 100mMTSP in 100% D_2O and 0.2µl 1.2mM NaN_3 in 100% D_2O was added. The mixture was then vortexed for 1 minute. 190µl each from all samples from both extraction techniques were pipetted into separate 3mm NMR tubes for NMR analysis.

4.2.3 NMR Spectroscopy

All NMR experiments were conducted on Bruker 700MHz Avance IIIHD spectrometer equipped with $^1H/^{13}C/^{15}N$ Cryoprobe with Z-gradient. NMR setup was performed by calibrating the temperature to 25°C using the deuterated methanol thermometer method[154]. Tuning and matching was conducted followed by the automated optimisation of the acquisition parameters in order to adjust the residual water suppression. Automatic shimming routine was performed on each sample. All chemical shift values were set relative to TSP at 0.00 ppm for

Neuropathological Diagnosis							
Astrocytoma II							4
Astrocytoma III							5
Oligodendroglioma II							4
GBM IV							9
IDH		MGMT Methylation		ATRX Loss/Retention		1p/19q Co-deletion	
Mut	13	Methylated	7	Lost	9	Co-deleted	3
WT	9	Unmethylated	2	Retained	12	Not deleted	9
	22		9		21		12

Table 4.1 A table to show the distribution of glioma samples included within this study. A total of 22 samples were included in the investigation, from mixed pathological origin. The MGMT, ATRX and 1p/19q status was not available for all samples however here we have reported all available information.

both ^{1}H and ^{13}C. 1D acquisition was performed using a Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (cmpgpr1d) with 4 dummy scans, 256 scans, dwell time of 41.6μs, 9.6ms echo time, 32,000 points and 3.1s acquisition time. The 2D acquisition used a ^{1}H-^{13}C heteronuclear single quantum coherence coupling (HSQC) sequence (hsqcetgpsisp.2) with 16 dummy scans, 4 scans, 47.33μs dwell time with 768 increments and a spectral width in the indirect dimension of 120ppm (offset at 50ppm). The spectra were Fourier transformed with vendor supplied standard processing routines (apk0.noe for 1D and 2dinv for 2D) comprising sine bell window function, zero filling and automated phase correction with 2D spectra subjected to additional manual phase adjustment.

4.2.4 Metabolite Identification

Metabolite identification was optimised via three methods. Initially, spectra were bucketed at regular intervals using the Bruker software AMIX. To create the bucket tables, spectra were imported into AMIX and regular rectangle buckets were selected. Integration mode was set to sum of the intensities and the scaling was set to the total intensity of the spectrum. The region of 4.6-4.8 ppm was excluded to negate the peak produced by residual water signal. Buckets were set at intervals of 0.04 ppm. The bucket table was exported for metabolomic analysis in Metaboanalyst. The integrals of the peaks within each bucket were used for metabolomics.

The second method of metabolite identification was to create a pattern file containing the ppm range values for each metabolite in order to bucket all the spectra in a targeted manner. A list of metabolites contained in the pattern file can be found in appendix 2. For this, a publicly available pattern file derived from a neuroblastoma cell line containing 186 metabolites was used as a base to be modified with the addition and deletion of relevant metabolites. The resultant pattern file contained 92 metabolites to be analysed. All spectra were uploaded onto the TAME NMR server, and the pattern file was overlaid. The ppm values were adjusted within the .csv pattern file, so as that each metabolite range would capture the majority of the relevant peaks. The spectra were then bucketed, using AMIX. To bucket the spectra into intervals of a pre-defined width based on the pattern file input, the 'variable sized buckets' option was selected and the path to the pattern file was given. The integration mode was set to sum of the intensities and the scaling was set to total intensity of the spectrum. The region of 4.6-4.8 ppm was excluded to negate the water peak, and the resultant bucket table was exported for metabolomics analysis in metaboanalyst.

The final method of metabolite identification was to process the spectra using Chenomx NMR Suite 8.2 Software (Chenomx, Edmonton, Canada). Chemical shifts were referenced in relation to the TSP signal at 0.0 ppm. A semi-automated approach was used towards spectral profiling. The software provides a library of 350 metabolites that may be fitted in human samples. We edited this library so as to only contain metabolites that are known to be found in both healthy brain and brain tumour tissue. This edited library consisted of 78 metabolites in total. The reference library provides a Lorentzian peak shape model for each automatically fitted peak that is superimposed on top of the spectrum being analysed. Following automated fitting using this model, each peak was checked, and the optimal fit was achieved by manually adjusting the peak height and by making alterations to cluster locations. This was a necessary step in the metabolite identification and quantification in order to achieve the optimal result, as we observed a significant amount of spectral shift between samples. This step was also necessary in order to eliminate any false positives that may have arisen from the fitting of peaks that are heavily overlapped. The linear combination of all modelled metabolites gave rise to total spectral fit, which can be evaluated within the software as a summation line superimposed onto both the analysed spectrum and the estimated fit. An estimated concentration was generated following peak integration and the concentration table was then

exported for each peak and concatenated for statistical metabolomic analysis in Metaboanalyst.

For metabolite identification and statistical metabolomic analysis of the 2D NMR spectroscopy data, the data were phase adjusted in TopSpin and analysed in CCPN NMR[155]. Once the spectra had been manually zero and first order phase corrected, the resultant spectra were converted to UCSF format to be imported into the software CCPN. Once spectra from all samples had been imported, contour levels were adjusted for each spectrum, so that a similar level of noise was present in all. To account for any spectral shift, all spectra were aligned to the lactate peak from one representative sample. To conduct metabolite identification, standard spectra of metabolites of interest that had been acquired in-house were imported and overlaid. Peaks corresponding to the standards were assigned. All unassigned peaks that then featured in all spectra across all samples were given numerical assignments prior to metabolomics analysis. Concentrations calculated from contour intensities were then exported for statistical metabolomics using metaboanalyst. Following statistical metabolomics, those peaks that were previously unidentified metabolites and assigned only a numerical identifier, and had shown to be significantly altered between the two groups, were queried using the biological magnetic resonance data bank[156].

4.2.5 Statistical Analysis

Samples were divided into groups dependant on their IDH1/2 mutational status. Statistical analysis was carried out using in-house R scripts and Metaboanalyst. Both univariate and multivariate analyses were carried out on the data. Univariate testing is applied when individual variables are compared one at a time between the different testing groups. Here we applied Wilcoxon rank test. Wilcoxon rank sum test was chosen as the statistical test because it is commonly used to compare two groups of non-parametric data. As the data here is of unknown distribution, this stricter approach was appropriate. When carrying out multiple significance tests such as in the context of any 'omics' study, there is a requirement to correct for multiple tests to decrease the probability of detecting false positives[157]. This correction procedure can be considered as a fundamental part of univariate testing within an 'omics' framework, and as such must be taken into account when interpreting the results, as application of different methods of multiple comparison correction may facilitate the

production of different results. Metaboanalyst applies the Bonferroni correction for multiple comparisons, as well as providing the false discovery rate (FDR) adjusted P value.

Multivariate data arises from experiments whereby two or more variables are measured. When employing multivariate tests on high-dimensional data, it is important to understand that these methods make use of all variables simultaneously and deal with the simultaneous relationship amongst such variables. Whereas a univariate analysis considers only the mean of a single variable and its variance, multivariate analysis looks at covariances and correlations within the variables that may be able to describe the extent of the relationship between them. Here we applied two methods of multivariate analysis, principal component analysis (PCA) and partial least squares discriminant analysis (PLS-DA). PCA is a useful tool, particularly in metabolomic studies, as it allows the projection of data into a lower dimensional space whilst capturing a large amount of original information describing the variation. In this way, quick and convenient plots can be produced to visualise separation between testing groups. Principal components (PCs) involve the linear transformation of the data, and conversion of a set of correlated variables into sets of new, uncorrelated principal components. Principal components are derived in such a way that the first PC accounts for the maximum amount of variation in the original data. The second PC accounts for the maximum amount of variation that the first PC was unable to explain. This procedure produces a list of PCs that are ordered by the variance that they each account for.

PLS, much like PCA, aims at capturing a large amount of information from the original data that describes the variance, and plotting it in a lower dimensional space, often on smaller sets of individuals or samples, and in a more condensed way than PCA. This is realised by also considering the relationship between the independent and dependant variables within the analysis. As the dependant variables that we deal with in metabolomics are categorical, the PLS variant that is able to achieve this is PLS-DA. One of the primary caveats in PLS-DA model generation is the tendency to overfit the data. Therefore, cross-validation strategies are a requirement to circumvent this issue[158]. One other issue that can arise from the application of PLS-DA models in the search for biomarkers and biological signatures is interpretation, which out of the many variables tested and analysed contribute to the discrimination between the two testing groups.

4.2.6 Normalisation Calibration

Normalisation of the spectral data acquired from an NMR experiment, such as the data presented in this chapter, requires some form of normalisation strategy prior to quantitative statistical analysis so that the data from all the samples are comparable with each other. There are a number of reasons as to why unwanted variance within the data may be present. Dilution variability is a process that influences concentrations of all metabolites by the same factor within each sample, and thus all peak intensities will be affected by the same magnitude, which may also be referred to as unspecific alterations of metabolites. In contrast, we are assessing for differences that are present due to the presence/absence of IDH mutation, which will affect only a number of selected peaks by varying degrees of magnitude. There may also be variation within the data that is present due to technical reasons linked with the acquisition of the data. Normalisation allows the suppression of all of these factors that may mask the specific changes in metabolite concentration that we are seeking. Here we assessed the ability of three normalisation strategies to account for the intra-sample and intra-group variability, so that the biologically relevant information can be easily assessed. The normalisation procedures followed were normalisation to the mass of the tumour, to the TSP peak and to the total area underneath the spectrum all within metaboanalyst. The data that were used to assess the ability of these methods to account for the unwanted variability whilst preserving the biological importance were acquired from neuroblastoma tumours cultured in chick embryos for the study by Al-Mutawa et al[154].

4.2.7 Pathway Analysis

Metabolites identified as significantly altered between IDH Mut and WT glioma were then used in pathway analysis in order to asses for relevant pathways that may have been altered in the presence of the mutation, providing information that can be used to form future research questions in order to probe the pathophysiology of IDH mutations. For this the pathway analysis tool in Metaboanalyst was used, which combines the results from pathway enrichment analysis and pathway topology analysis in order to help identify aberrant pathways. Data was inputted as metabolite names with one compound per row. The Homo sapiens library was selected, and the over representation analysis method selected was hypergeometric test.

4.3 Results

4.3.1 Quality Control and Optimisation Experiments

4.3.1.1 Extraction Calibration

The PCA of the two different extraction methods, acetonitrile and perchloric acid, have been displayed in figures 4.1 and 4.2. During the extraction process, it is necessary to neutralise the pH of the sample with a basic solution, and in this case we used NaOH. Due to the steep gradient of the pH curve when using a strong acid and base such as $HClO_4$ and NaOH, the window of neutralisation is very small. As such, three samples were not neutralised to a pH in the range of 7.0-7.5, and were labelled as bad pH for analysis. This led to the formation of three testing groups.

The PCA score plot demonstrates the extent of variability seen between the different extraction methods. A three component model as seen in figure 4.1 facilitates 72.6% of the overall variance. In addition, the 4th and 5th components account for 7.9% and 7.6% of the variance respectively. In the PCA plot (figure 4.1) for all three components it is clear that the acetonitrile samples are clearly separated from both perchloric acid pH groups.

Consideration of the PCA score plot (figure 4.2) allows us to capture the maximum amount of varition within the data, in decreasing order. Here the variables correspond to spectral bins, whereby all aligned spectra were divided up into regular intervals and the value for testing is the integral of the peak within the limit of the bin. PCA evaluation shows a clear grouping of the acetonitrile extracted samples (red) with very little variance on either the first or second PCs. Conversly, both the percholric acid-extracted groups display a much larger amount of variance. These results formed the rationale for choosing acetonitrile extraction here.

4.3.1.2 Normalisation Calibration

Figure 4.3 shows the PCA score plots with the first two PCs for each of the normalisation methods explored. 4.3 (a) illustrates data that has been normalised using the mass of each of the tumour samples. As is evident, some of the data points skew the distribution, and there is a loss of biologically relevant information as a result of this. The remaining presence of unwanted variation following normalisation may be due to tumour heterogeneity. Tumours are composed of varying amounts of blood vessels, necrotic tissue, and active tumour. As the active region of the tumour is the only region that is contributing to the variation in peak
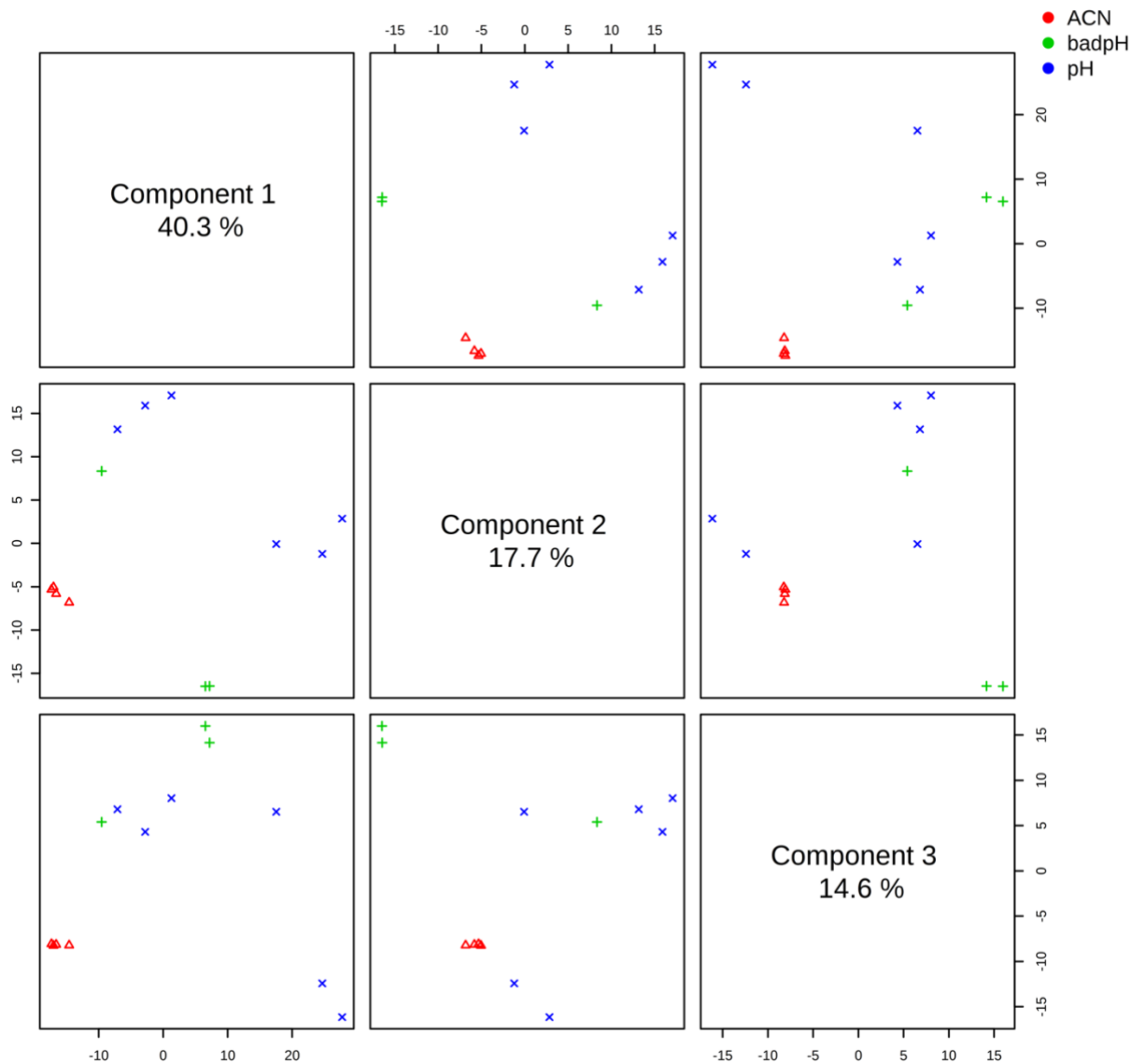
Fig 4.1 Three component pairwise plot representing the variance accounted for by each PC in the corresponding diagonal box. ACN = acetonitrile extraction (red). badpH = perchloric acid extraction with a pH outside the range of 7.0-7.5 (green). pH = perchloric acid extracted with a pH neutralised within the range of 7.0-7.5 (blue). There is a clear grouping of all samples that underwent acetonitrile extraction when all combinations of the first three PCs are taken into account.
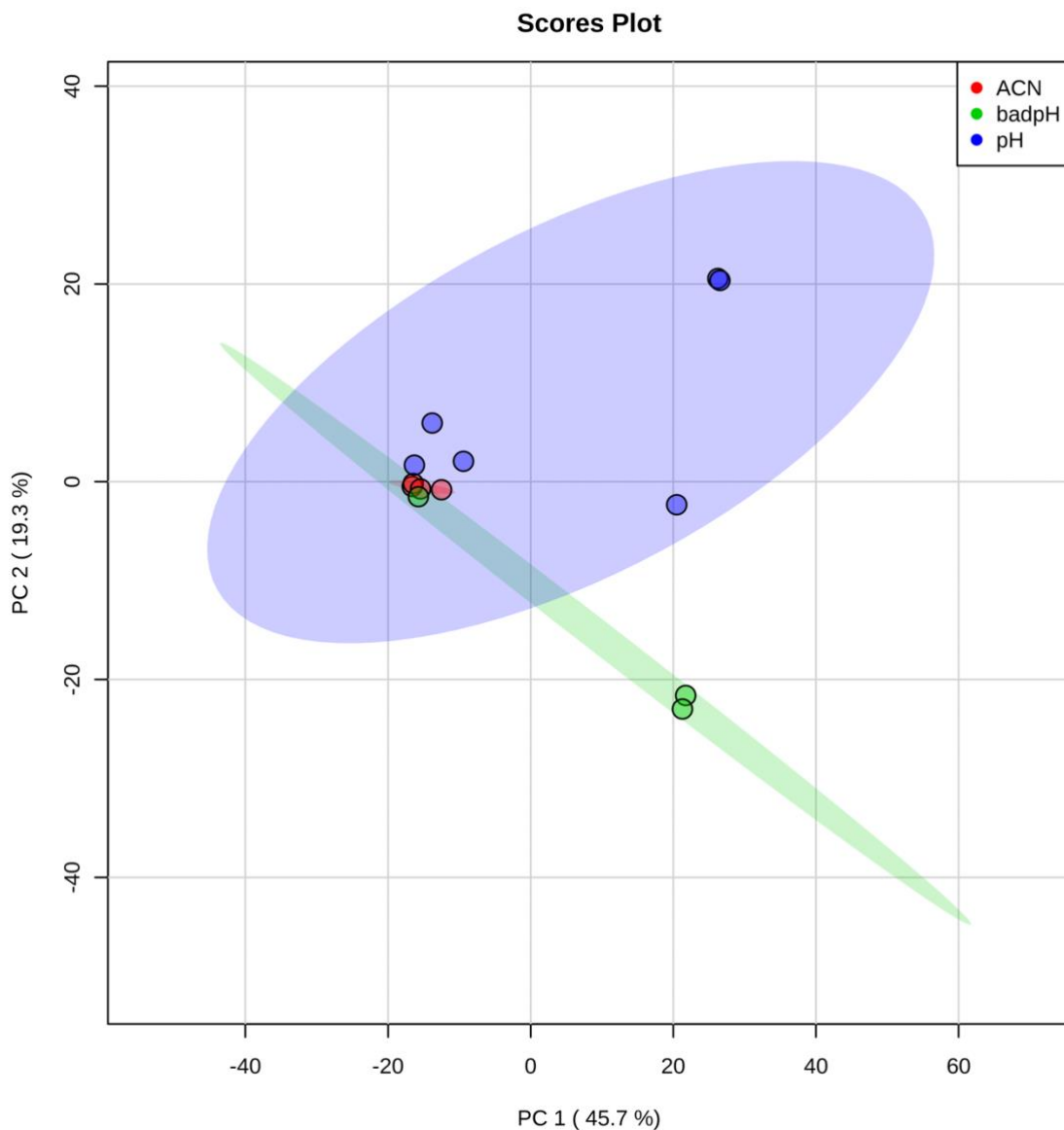
Fig 4.2 Principal component analysis 2D score plot for all samples in the three groups pertaining to the two different extraction techniques. The samples extracted using the acetonitrile approach display the least amount of variance on both the first and second PCs. Both the groups with the perchloric acid extracted samples displayed a large amount of variation between each individual sample. Interestingly, a much larger variation was observed within the samples that had the pH appropriately neutralised to within the range of 7.0-7.5.
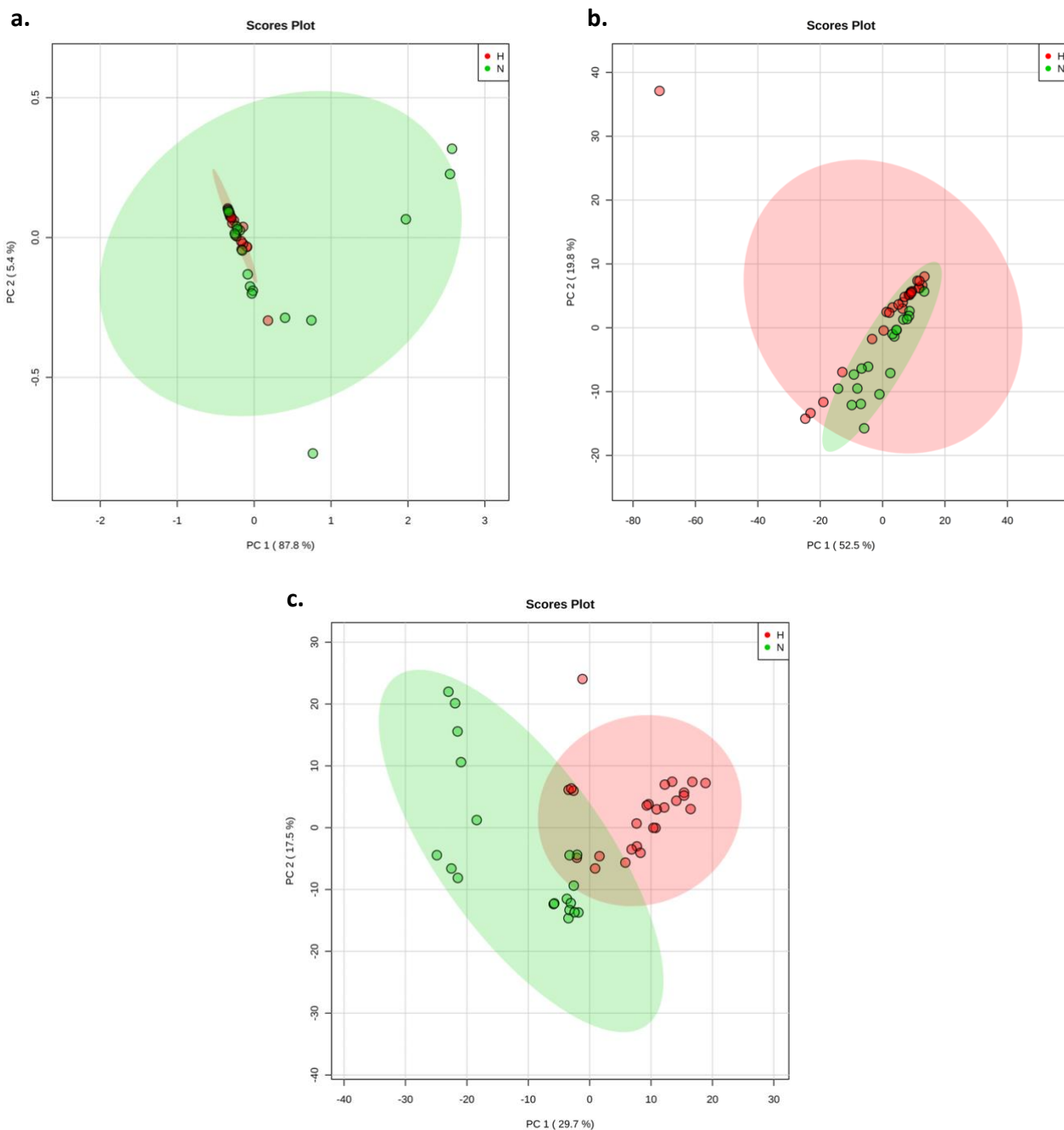
Figure 4.3 PCA score plots of NMR data acquired from neuroblastoma tumour samples. Three methods of normalisation were applied here in order to determine the optimal protocol for reducing variance whilst preserving biological importance. **(a)** samples were normalised to the mass of each tumour sample. Here the data has been vastly skewed and much of the biologically important information has been lost. **(b)** samples were normalised to the TSP peak. The data here is skewed slightly, and there is little separation between the two testing groups. **(c)** samples normalised to total area under the spectrum. Here we are able to see separation between the two groups, with little skewness.

intensity that we are interested in, the variability introduced by the differing levels of other tissue components may be masking the biologically relevant data.

Figure 4.3 (b) shows the PCA score plot for data that was normalised to the TSP peak from each spectrum. Here the data is much less skewed when compared to the normalisation by mass protocol data, however, there seems to be a clear overlap of both the testing groups. Due to the lack of separation between the two groups, relevant biological information may be lost during the analysis. One reason for the suboptimal performance of TSP normalisation could be that it has been shown the protein can bind to TSP[138]. Although metabolites extraction was carried out, the complex nature of the tissues being analysed may have resulted in some residual protein present within the samples. This will in turn cause a variation in the final TSP concentration within each sample, introducing variability and skewing the data following normalisation using this method.

Figure 4.3 (c) shows the PCA score plot for the normalisation to the total integral of the spectrum for each sample. Here we can see a clear separation between the two testing groups, along with good clustering of the datapoints within each group. From this we are able to observe the apparent differences between the two groups that may be attributed to specific alterations in metabolite concentrations. Taken together, these results suggested that the most appropriate method for normalisation of the NMR data was to use the total area of the spectrum.

4.3.1.3 Metabolite Annotation and Quantification – Spectral Binning

Figure 4.4 shows a representative spectrum for IDH Mut and WT samples. The PCA plot and PCA score plot for the initial cohort of 17 glioma samples can be seen in figure 4.5. The two testing groups refer to the IDH mutational status. There was a large overlap between the two testing groups when considering PCs 1 and 2. In fact, when looking at the first 3 PCs, this only accounts for 68.3% of the overall variance seen throughout all samples. PCA involves the transformation of variables in one coordinate system into PCs within a new coordinate system, and it is therefore useful to estimate how much each of the old variables (in this case spectral bins) contribute to the formation of each of the new PCs. This can be displayed by way of a loading plot (figure 4.6). The annotated spectral bins (1.32500005, 1.33500004, 2.0250001, 3.03500009, 3.2250014, 3.56500006, 3.78500009 and 3.93499994) are the main contributors to separation along the x-axis, which indicates biological significance, whereas the points
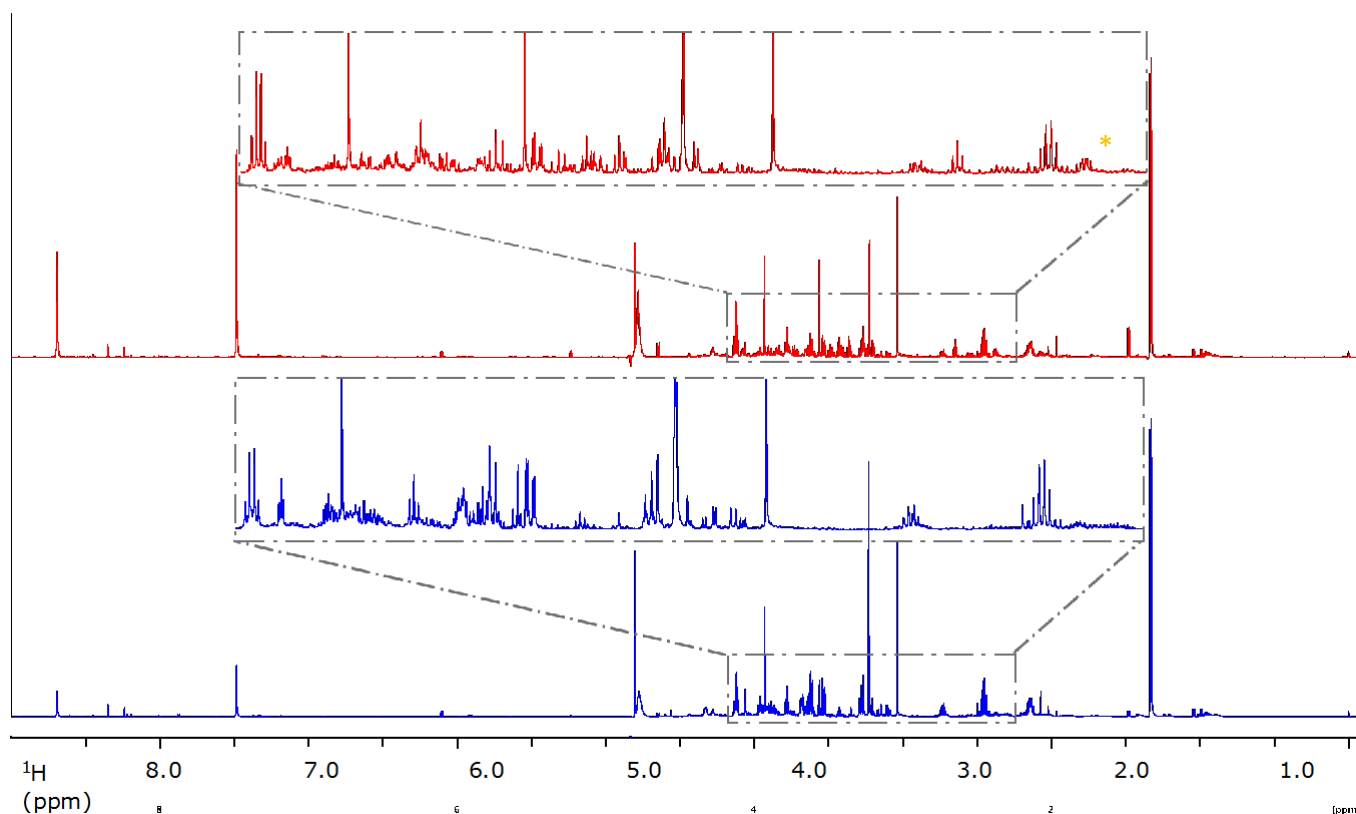
Fig 4.4 Two sample spectra acquired from IDH Mut (red) and WT (blue) patient samples. The insert shows a zoom of the crowded aliphatic region. The 2-HG peak at 2.25 is denoted by the yellow Asterix.

scattered along the y-axis are more likely to contribute to the separation of technical replicates. The integrals of these spectral bins are at much lower values than others, and therefore are more likely to be affected by the presence of noise or any residual baseline distortion.

Although there was not a clear separation in the initial PCA analysis, there is utility in assessing the spectral bin that the loading plot has shown to contribute to the slight degree of separation between groups that we are able to observe. To do this the spectra were overlaid on top of each other and the bins plotted over. Figure 4.7 illustrates sections of the spectra that correspond to the bins shown to contribute to the separation from the PCA loading plot. Upon inspection, it was evident that each bin did not contain the same integral area of each peak from each spectrum. This is due to a varying amount of spectral shift seen between samples. For this reason, a more tailored method for metabolite identification was followed, whereby a custom pattern file was produced in order to bin the spectra at variable intervals that should include the same peaks from all spectra.

a.

b.
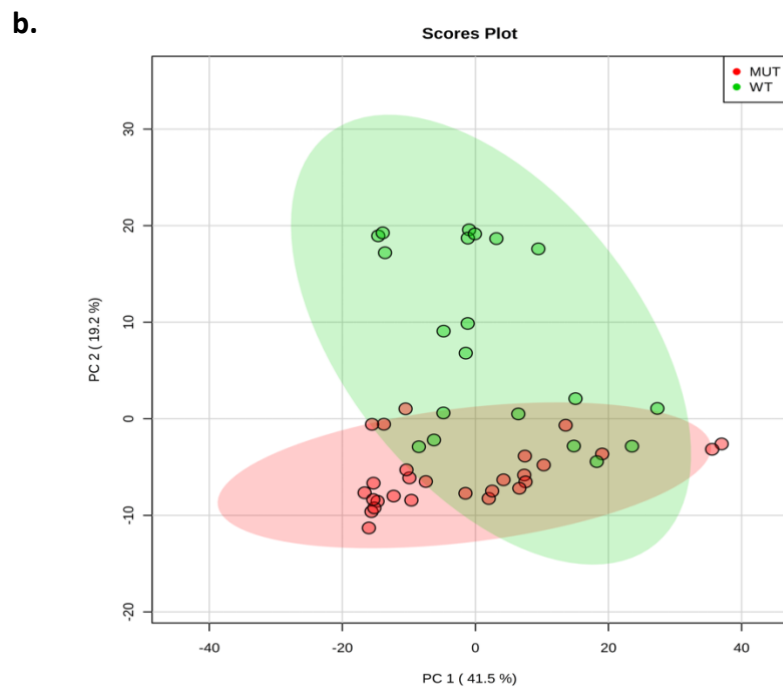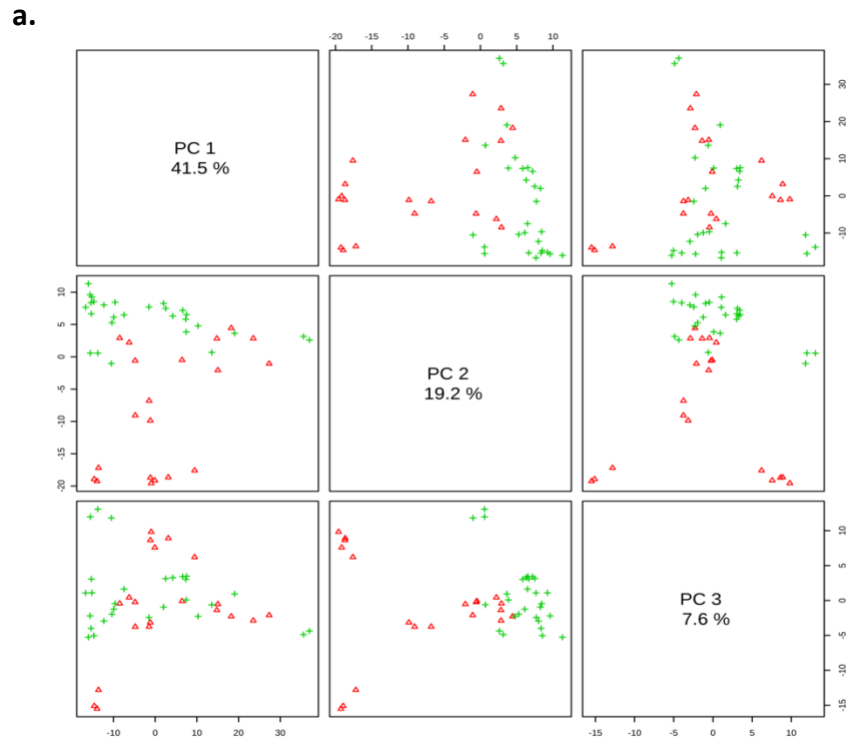
Fig 4.5 (a) PCA plot for NMR spectra that had undergone spectral binning at regular intervals of 0.01ppm. the first three components only account for 68.3% of the variability and there is no obvious separation of the groups when plotted in any of the dimensions. (b) PCA score plot for the same dataset. There is a large amount of overlap between the WT and Mut groups, although there does seem to be a subgroup within the WT samples, demonstrating variability within the testing group that is not attributed to the mutational status.
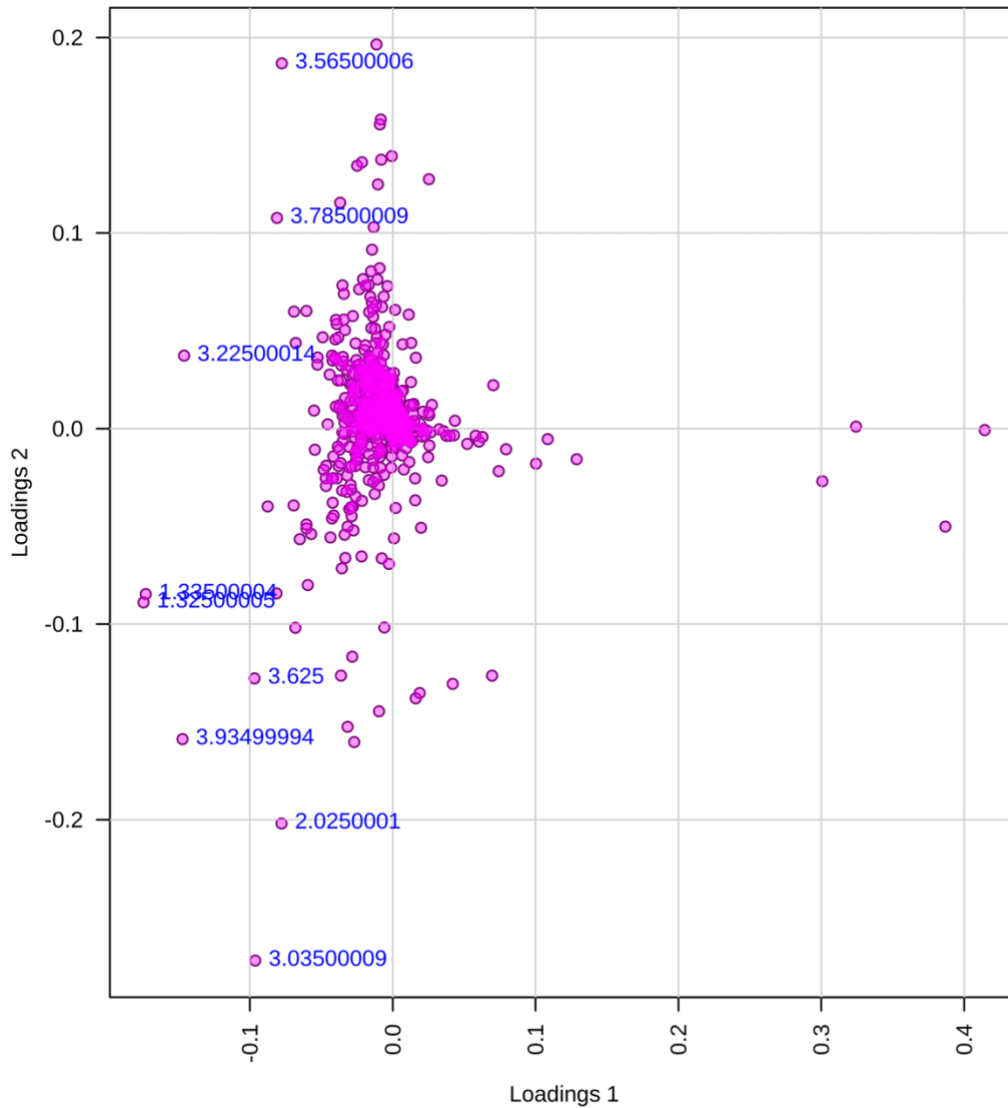
Fig 4.6 PCA loading plot. When the data is transformed from one coordinate system (spectral bins) to another (PCs) the contribution of each of the original variables to the new PCs can be assessed by way of a loading plot. The bins that contribute to the variation seen on the PCA score plots can be further assessed by evaluating the spectra. The labelled points grouped along the lower end of the x axis are most likely attributed to biological significance.
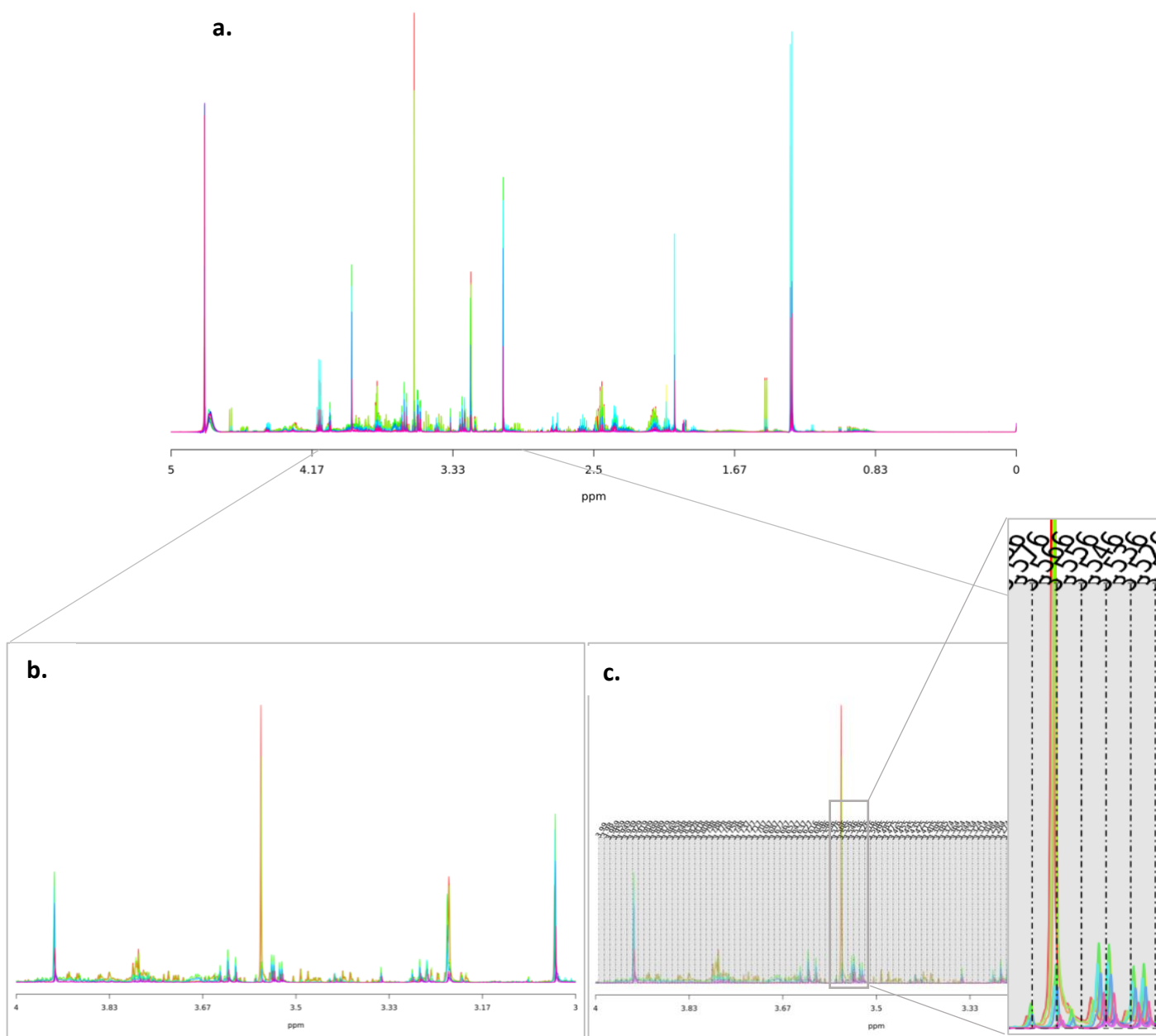
Fig 4.7 (a) Raw spectra from all samples overlaid (both IDH WT and Mut). It is evident from this aspect that there is some spectral shift between spectra. **(b)** Zoomed view of the same spectra, plotted between 3.0 - 4.0-ppm. This was the region that a number of spectral highlighted in the PCA loading plot originate from. **(c)** The same section of spectra as in (b), but with the spectral bin ID's plotted. Upon closer inspection, it is clear that this method of spectral binning at regular intervals is not appropriate as different aspects of peaks from individual spectra are contributing to each bin in varying amounts, which introduces bias in the statistical analysis. The bin labels are arbitrary numbers assigned consecutively to each bin.

4.3.1.4 Metabolite Annotation and Quantification – Custom Pattern File

As highlighted by the method of binning at regular intervals, a significant level of spectral shift was observed between samples from different patients, which may be attributed to differences in pH and varying salt concentrations, the reasons for which will be discussed later in this chapter. Due to this inconsistency within the data, a different approach was taken toward metabolite identification. A pattern file was created that consisted of a spreadsheet containing the upper and lower ppm values for each bin, which would be determined peak-by-peak, so that each bin was to contain the full integral from each peak. Figure 4.8 shows the PCA and score plot for metabolomics carried out using the custom edited pattern file to determine the bin limits for each peak. Initial inspection of this multivariate analysis shows that there is little-to-no separation of the IDH Mut when compared to the WT using this method of metabolite identification. When considering the PCA score plot, both the IDH groups are severely overlapped, with only 3 Mut samples lying outside the WT region.

It is sometimes the case that differences may be observed univariately but not via multivariate analysis, and vice versa, and the reasons for this imbalance are explored further in this chapter. Here we carried out Wilcoxon rank test as a univariate analysis to conduct pairwise comparisons between each identified peak. Using this approach, we were able to observe significant differences in 35 metabolites when comparing IDH Mut tumour samples to their WT counterparts. Figure 4.9 shows a scatter plot of all identified peaks (both assigned to metabolites and unassigned) plotted against the p-value generated from the univariate test, in addition to box and whisker plots of metabolites identified to be significantly altered. As expected, 2-HG was significantly altered, although unexpectedly the concentration in the WT appears to be >0.0mol. Upon closer inspection of the spectrum, the bins dedicated to 2-HG peaks contain varying levels of noise and baseline distortion, and the differences in spectral shift between samples caused the contributions from these artefacts to present at different amounts in each bin. This led to the false detection of 2-HG within the WT samples. Extrapolation of this observation throughout the rest of the bins covering the spectra gives a clear indication that this phenomenon has occurred in a number of spectral locations, leading to varying contributions of each integral to each bin, and therefore varying levels of bias in the statistical analysis. This varying contribution of misaligned peaks and noise formed the rationale for the individual metabolite fitting approach that was later adopted.
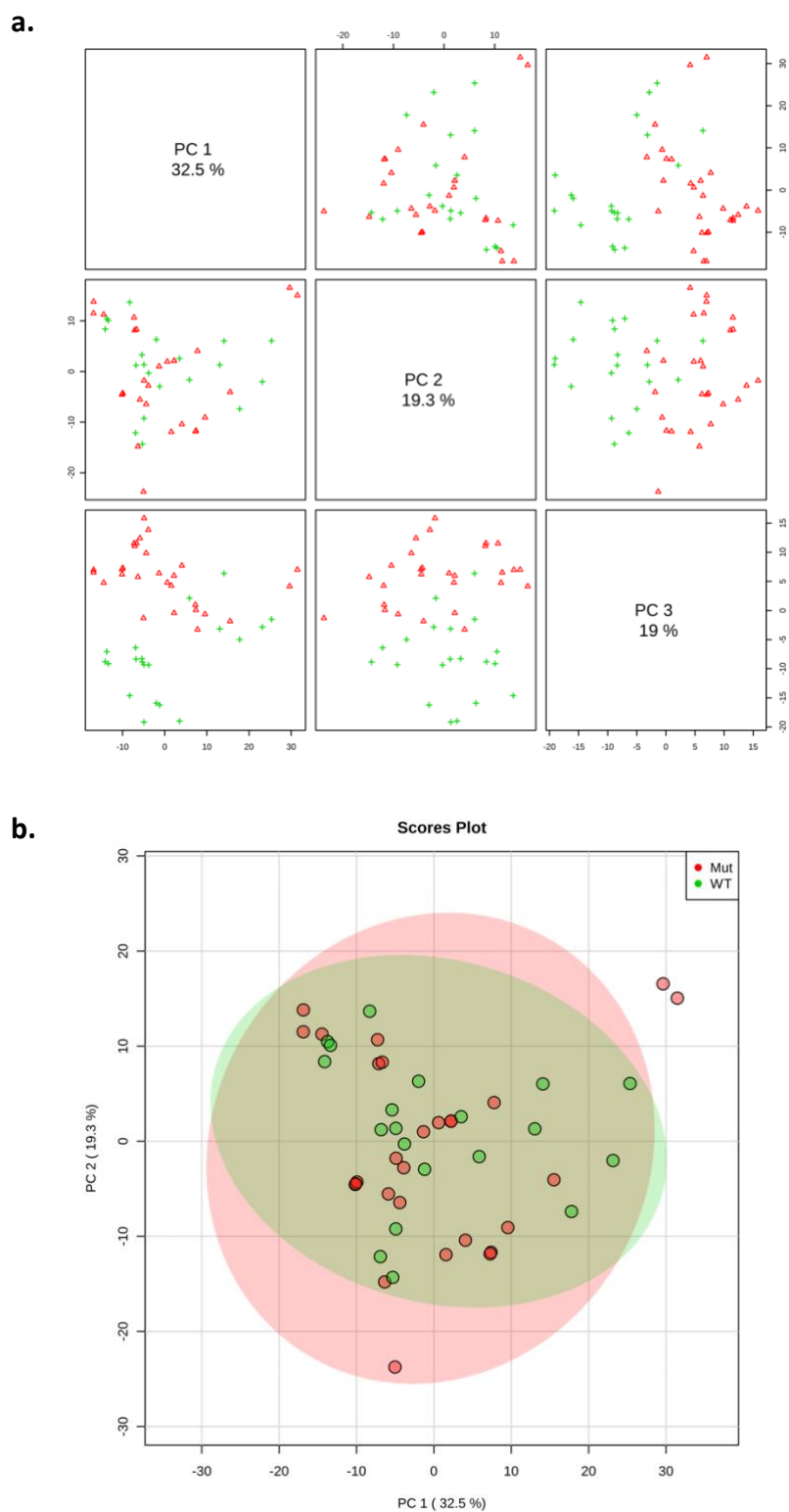
Fig 4.8 results from the 1D NMR custom pattern file approach. **(a)** A PCA plot displaying the first three PCs. Note that there is a great deal of overlap between the WT and Mut groups on all PC plots. **(b)** PCA score plot, there is a large amount of overlap between both the groups, with two outliers belonging to the Mut group.
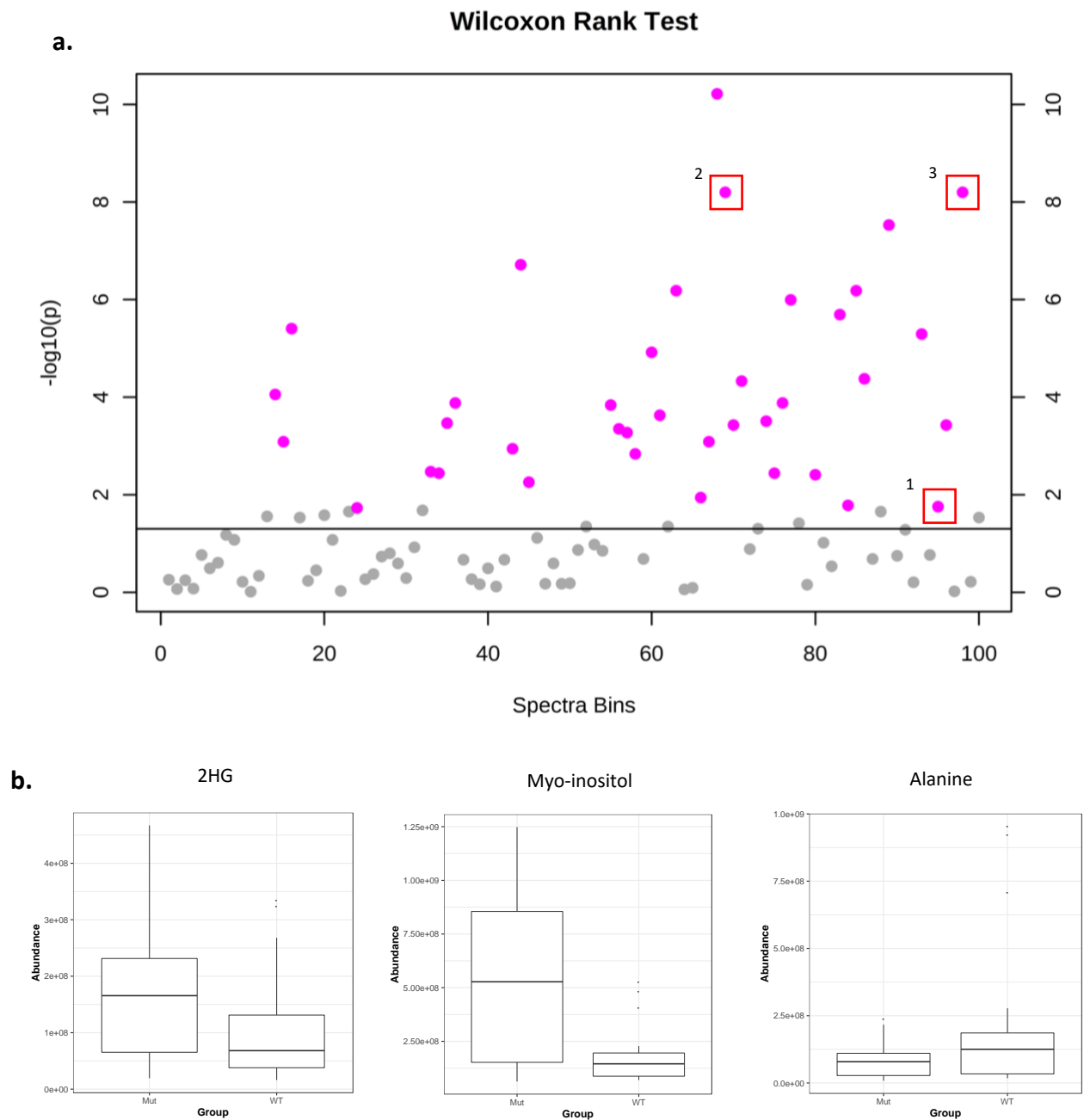
Fig 4.9 (a) A scatter plot of the adjusted -log(p values) when comparing IDH Mut and WT gliomas using a Wilcoxon Mann Whitney univariate test. There are 35 pink points that lie above the significance threshold where p<0.05. In the boxes are a selection of metabolite of interest 1. 2-HG. 2. Myo-inositol. 3. Alanine. **(b)** Box plots showing the normalised concentrations estimated from the integrals of the peaks within each bin.

4.3.2 Experimental Results

4.3.2.1 Metabolite Annotation and Quantification – Spectral Fitting

Using a custom pattern file order to bucket the spectra for quantification of individual peaks is a convenient approach when all the spectra are fully aligned and contain uniform, minimal noise and residual baseline distortion. This was not the case with the glioma samples used in this study, and elucidation as to why the effect occurred will be covered in the discussion. Due to this factor, an alternative approach was taken in order to quantify the metabolites measured by metabolomics.

Fitting of individual peaks on each spectrum using the Chenomx metabolites library allowed for accurate assignment of every peak of interest in all spectra, and therefore a reliable concentration to be calculated from the integral of each peak. Figure 4.10 shows scatter and box plots from the univariate analysis of both the 1D spectra. From the 1D analysis six metabolites were found to be significantly different when comparing the Mut relative to the WT tumours, namely an increase in myo-inositol ($p<0.0005$), an increase in 2-HG ($p<0.0005$), a lower choline ($p<0.0005$), an increase in 3-hydroxybutyrate ($p<0.05$), a decrease in alanine ($p<0.05$) and a decrease in methyl-succinate ($p<0.05$).

The metabolite concentrations exported from Chenomx following the fitting routine were then also analysed using multivariate approaches, both PCA and PLS-DA. The PCA and PCA score plots can be seen in figure 4.11. The PCA analysis shows that the first three PCs only account for 49.3% of the overall variance, and when the first two PCs are selected to plot against there is a large amount of overlap between the Mut and WT groups, with 2 outliers belonging to the WT group and 4 outliers belonging to the Mut group. From these multivariate analyses, there was no significant difference between the Mut and WT groups based on their metabolic profile.

Figure 4.12 and 4.13 shows the results from the supervised multivariate analysis of the 1D. A PLS-DA model was built for the 1D spectra with 2 components (fig 4.12). The figure shows the score plot of the first 2 components and the variance in projection (VIP) scores (fig 4.13) for the 20 metabolites that were found to have a VIP score of greater than 1 and therefore most influential in discriminating between the groups. The area under the curve (AUC) for the receiver operator curve (ROC) for this model was 1 (figure 4.14). As expected, the six metabolites that were found to be significantly altered *via* univariate analysis also featured in the top 20 metabolites that were most influential in the PLS-DA model. A further 14
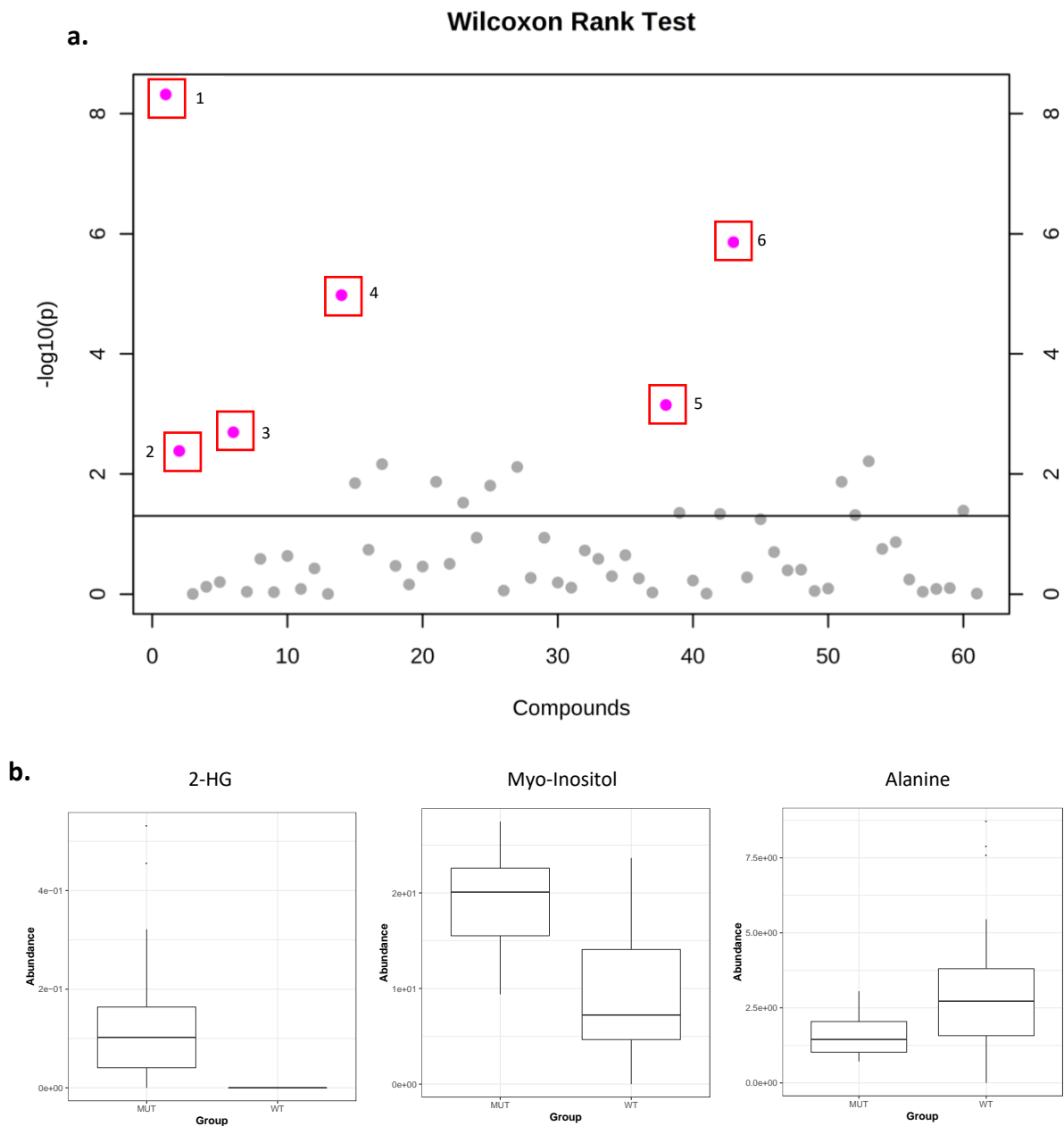
Fig 4.10 (a) Scatter plot depicting the -log10(p) values from the univariate Wilcoxon Mann-Whiney test 1) 2-HG. 2) 3-Hydroxybutyrate. 3) Alanine. 4) Choline. 5) Methyl-succinate. 6) Myo-inositol. **(b)** Box and whisker plots to show the concentration change of 2-HG, Myo-inositol and Alanine
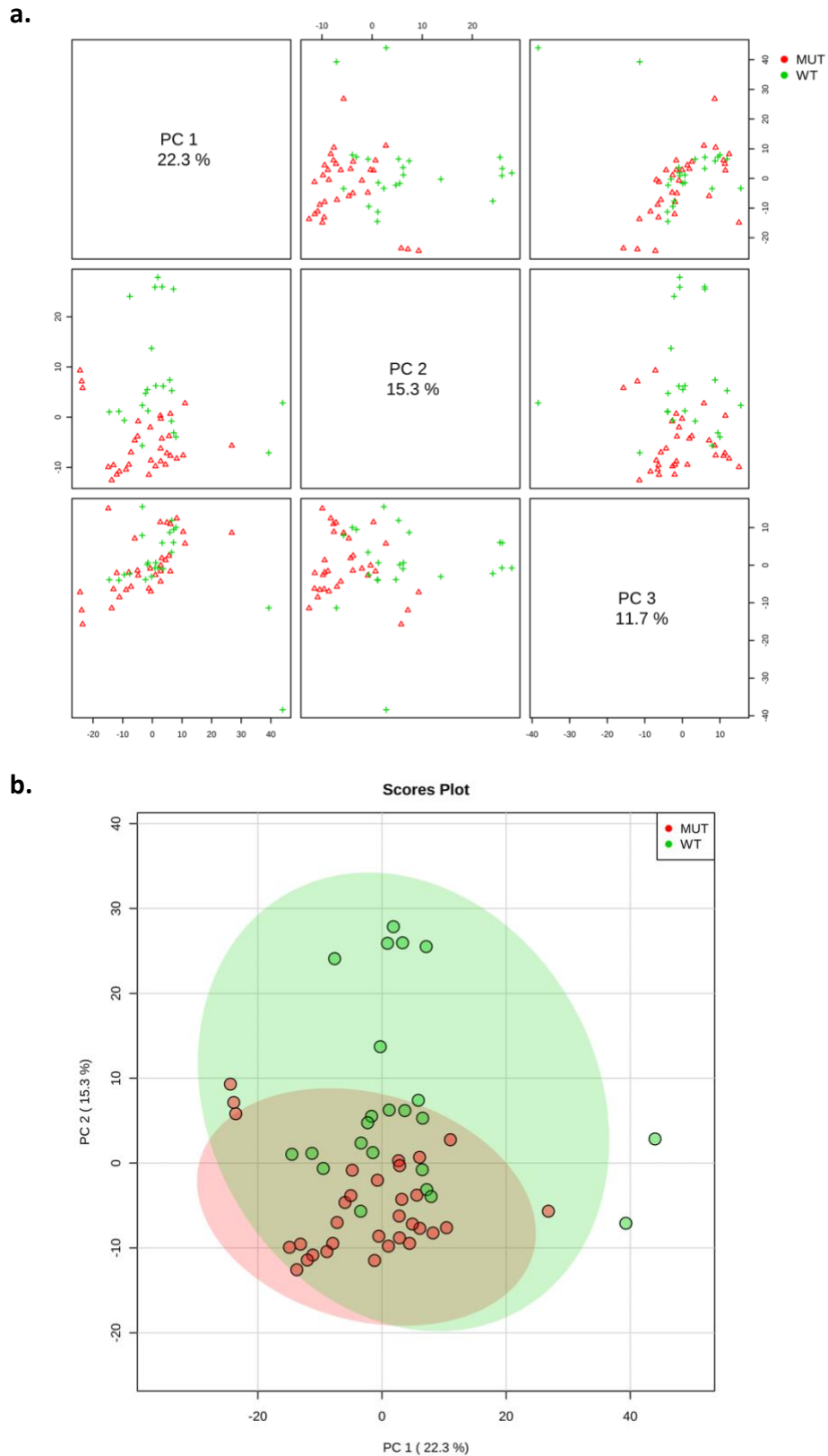
Fig 4.11 Unsupervised multivariate analysis from the data generated from spectral fitting. **(a)** PCA plot for the first three PCs. Taking into account the first three PCs, these only account for 49.3% of the overall variance, and both Mut and WT groups have poor separation in each of the scatter plots for all PCs. **(b)** PCA score plot for the first two PCs. There is a large degree of overlap between both groups and two distinct outliers belonging to the WT group.
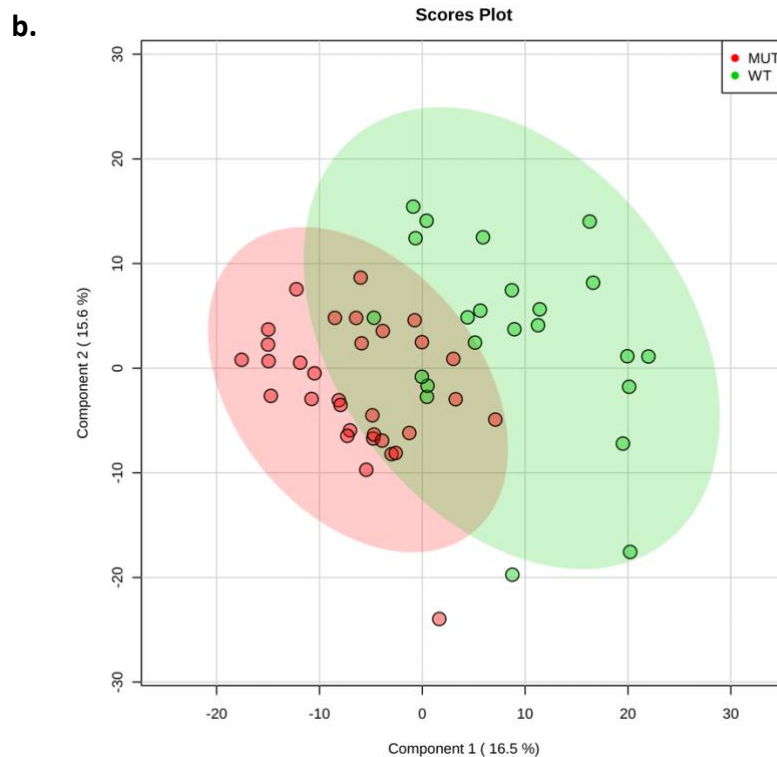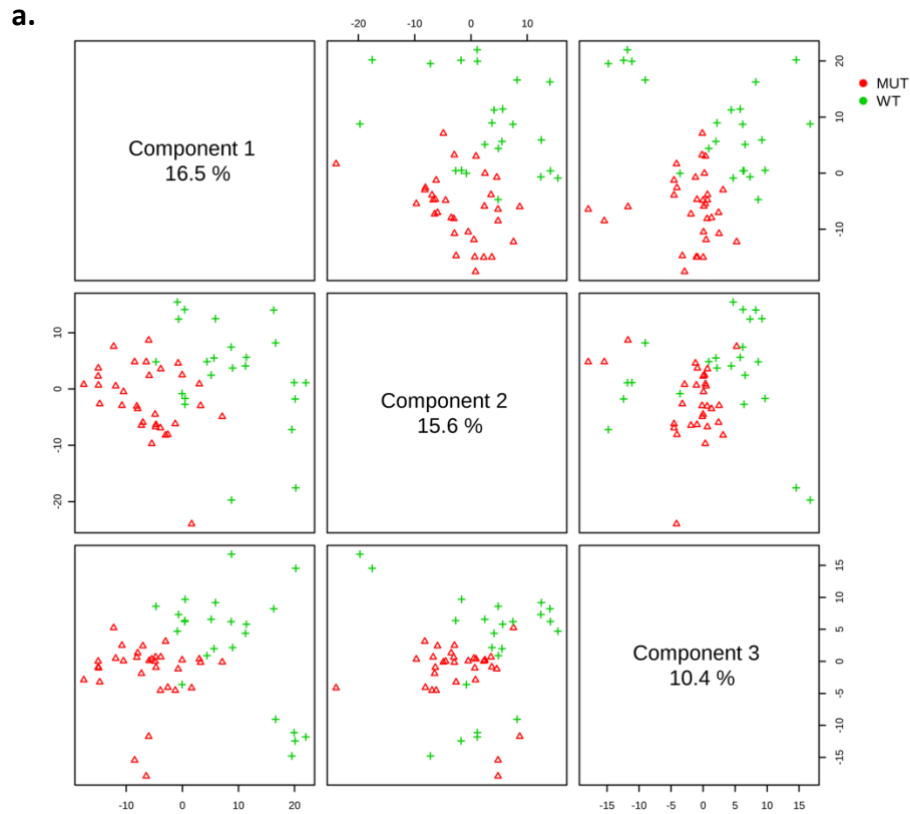
Fig 4.12 (a) PLS-DA plot of the first three components, these account for only 42.5% of the variability, less than reported by PCA. Like the PCA analysis, there is no obvious grouping on the scatter plots for all components. **(b)** PLS-DA score plot for PC 1 and 2. The pattern observed, including the degree of overlap and outlier position is very similar to that from the PCA analysis. Additionally, there is an outlier belonging to the Mut group. The R_2 and Q_2 for this model are 0.41 and 0.78 respectively.

VIP scores for 2-comp. PLS-DA model
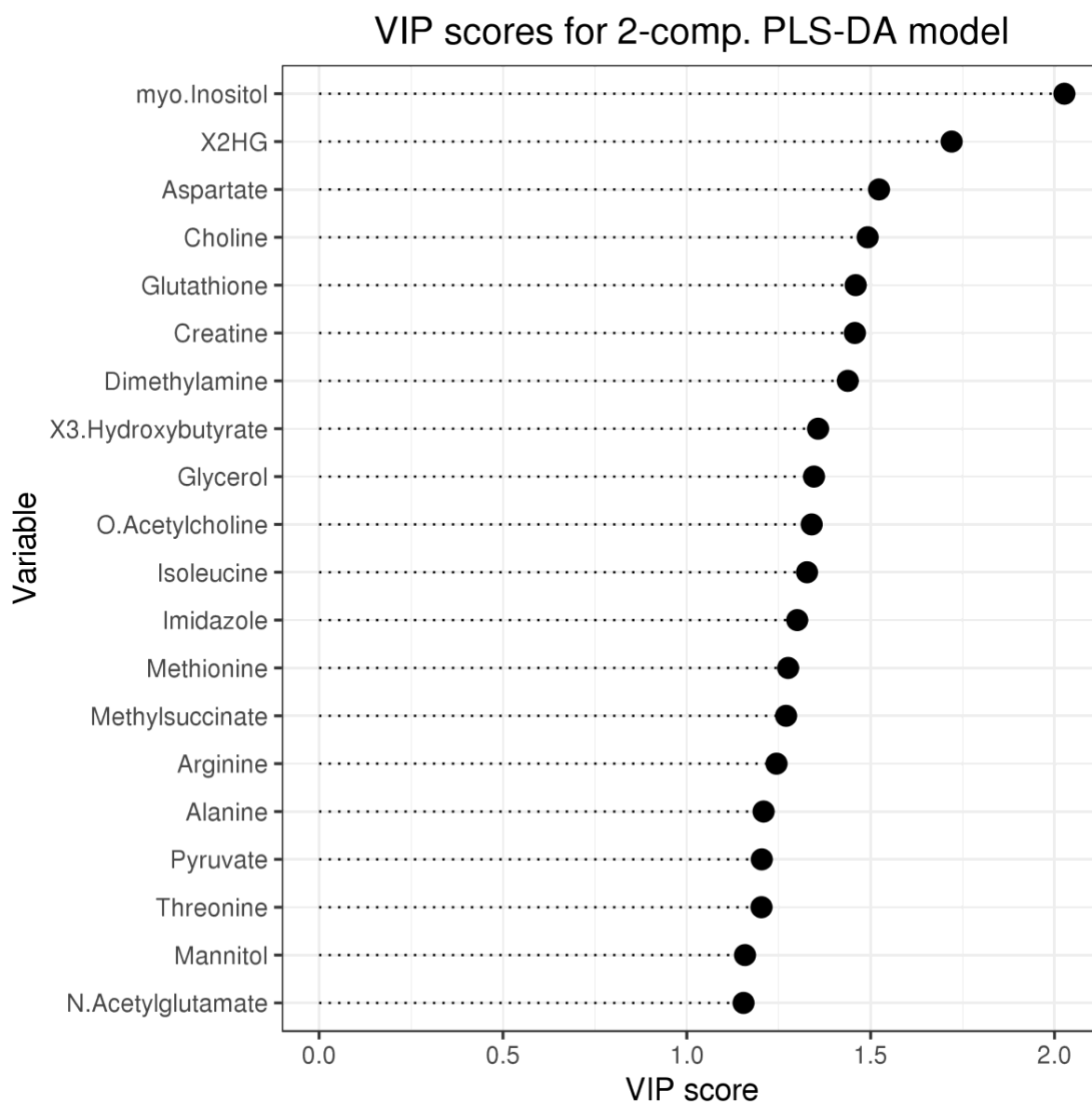
Fig 4.13 Variable importance in projection (VIP) plot, showing the contribution or importance of a variable to the model. Since VIP scores are calculated for each component, when more than one component is used to calculate the feature importance, such as here, the average of VIP scores is used. From this we can deduce that myo-inositol contributes to the most amount of variation, on average, across all components.
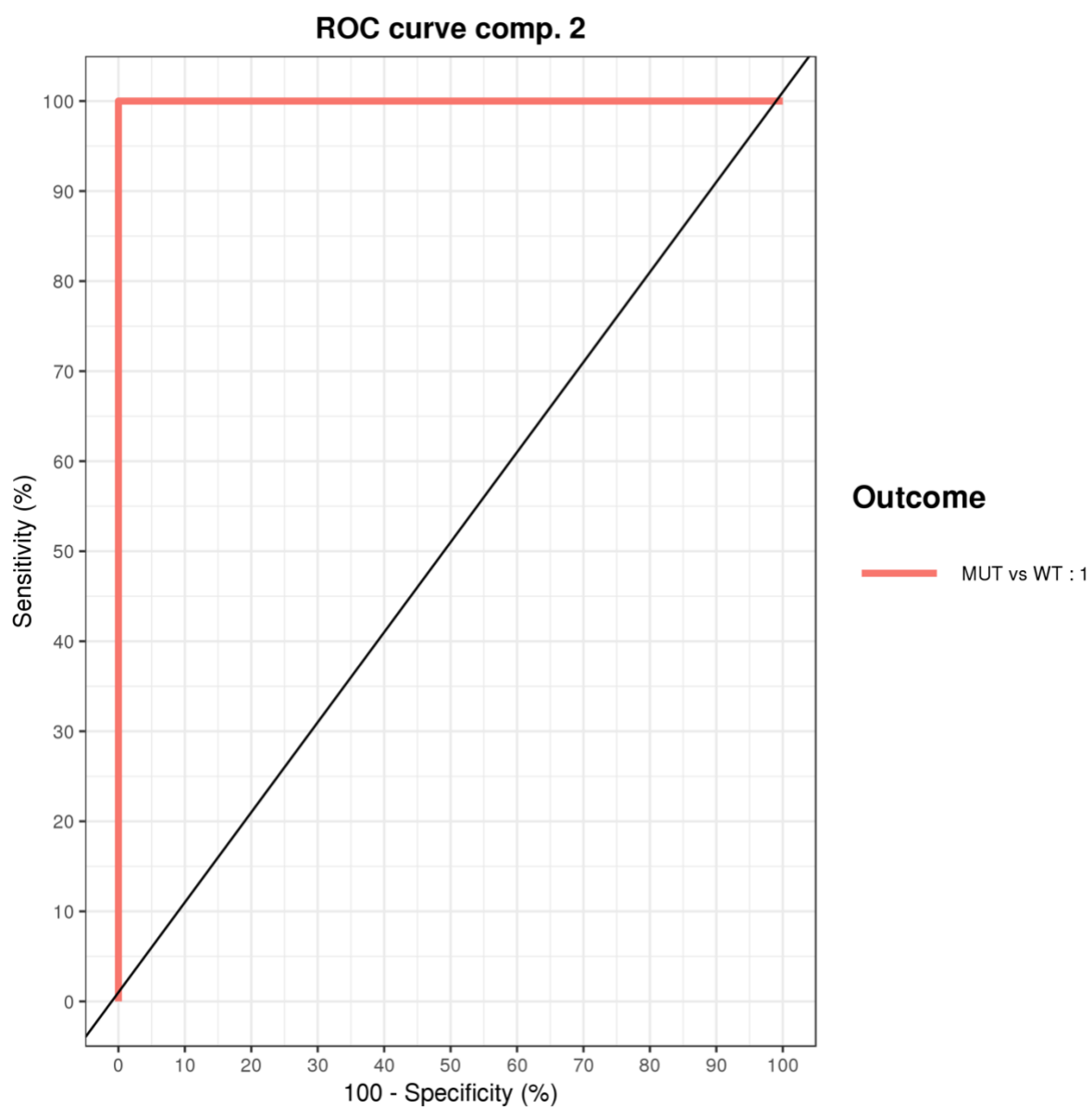
Fig 4.14 ROC curve for the 2 compartment supervised multivariate PLS-DA model calculated from the 1D NMR data. Here these data produced a model with a sensitivity and specificity of 100% resulting in an AUC of 1.

Fig 4.15 A 2D HSQC ^{1}H – ^{13}C spectrum of the aliphatic region of IDH WT(black) and IDH Mut (pink) overlaid. The yellow box indicates the location of the 2-HG peak. Here the ^{1}H resonance can be seen on the x-axis and the ^{13}C resonance on the y-axis.

metabolites were shown to be influential in discriminating between IDH Mut and WT tumours. All metabolites with a VIP score >1 were used in the metabolite enrichment pathway analysis.

4.3.2.2 2-Dimensional HSQC NMR for metabolite annotation and quantification

Although there is a wealth of information that may be gathered from 1D NMR spectra, resolution and subsequent quantification of certain metabolites, especially those residing in the crowded 0.0-3.0 ppm range, can often be problematic. 2D NMR helps to circumvent overcrowding of peaks. An example 2D HSQC spectrum can be seen in figure 4.15. The presence of 2-HG in the IDH Mut tumours can unambiguously be confirmed via 2D NMR
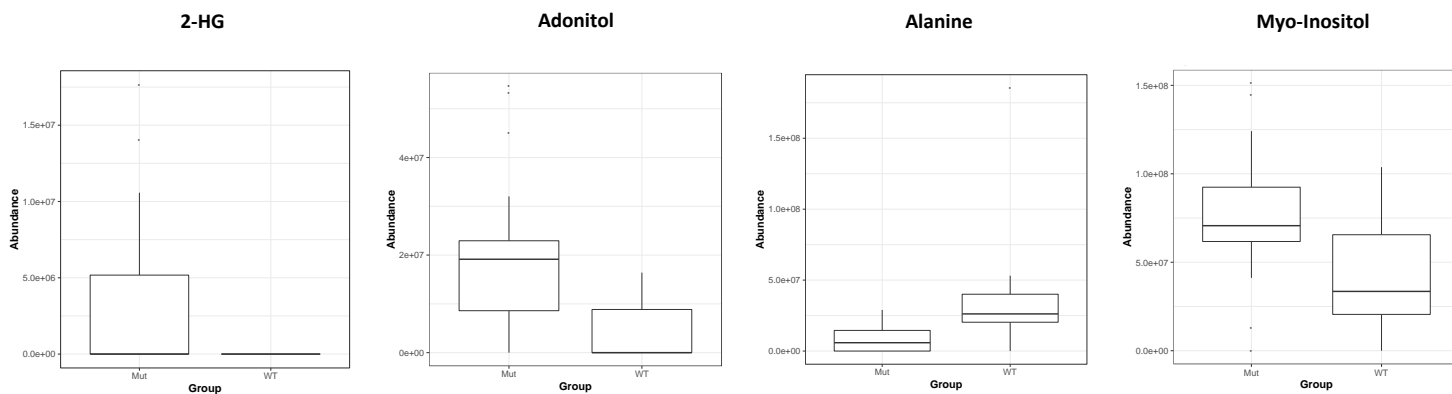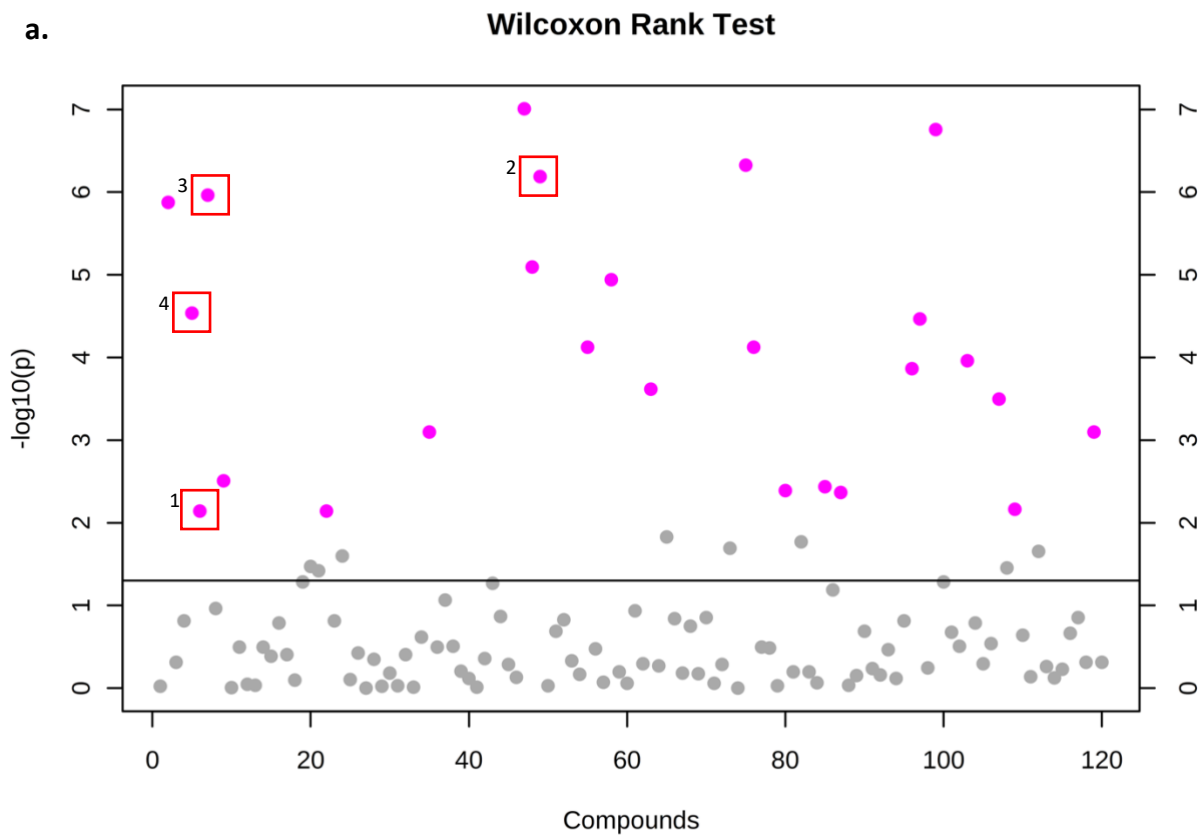
Fig 4.16 (a) A scatter plot to display the -log10(p) values from the Wilcoxon rank test for the normalised concentrations calculated from 2D NMR following multiple comparisons correction. A total of 22 metabolites were found to be significantly altered when comparing WT and Mut tumour groups. 1) 2-HG p<0.05. 2) Adonitol p<0.0005. 3) Alanine p<0.0005. 4) Myoinositol p<0.0005. **(b)** Box and whisker pots to demonstrate the alterations in metabolite concentration between Mut and WT groups. 2-HG, Adonitol and myo-inositol were all increased in the Mut samples, whereas Alanine was found to be decreased.
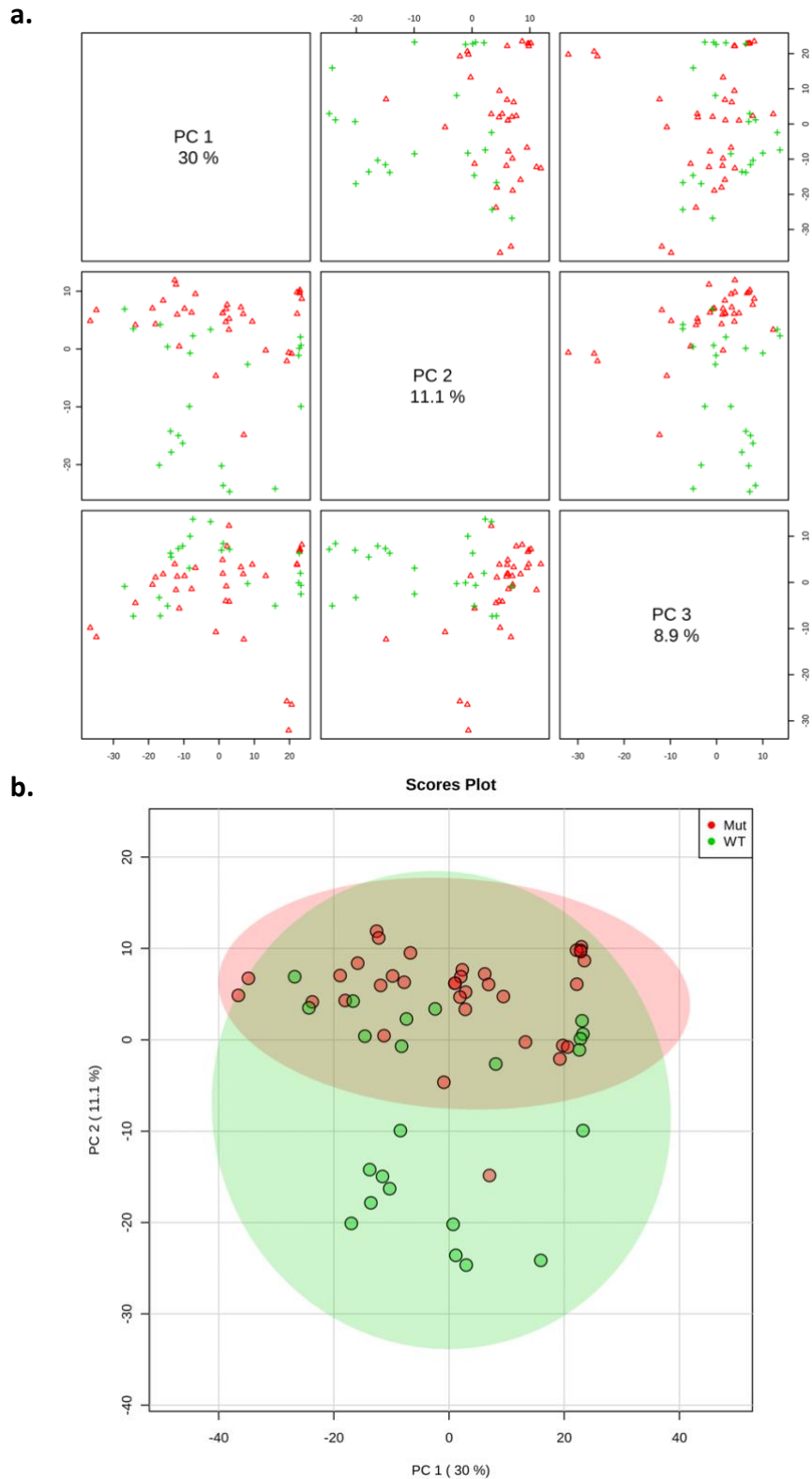
Fig 4.17 (a) PCA to show the first 3 PCs. Together these PCs only account for 50% of the variation seen throughout the data. There is a large amount of overlap observed when each of the first 3 PCs is plotted against each other. **(b)** PCA score plot for the first two PCs. There is a large amount of overlap observed between both groups here. No clear pattern in being able to discern between WT and Mut based on metabolites.

a.

Scores Plot

Component 2 (26.9 %)

Component 1 (11.6 %)

b.

VIP scores for 2-comp. PLS-DA model
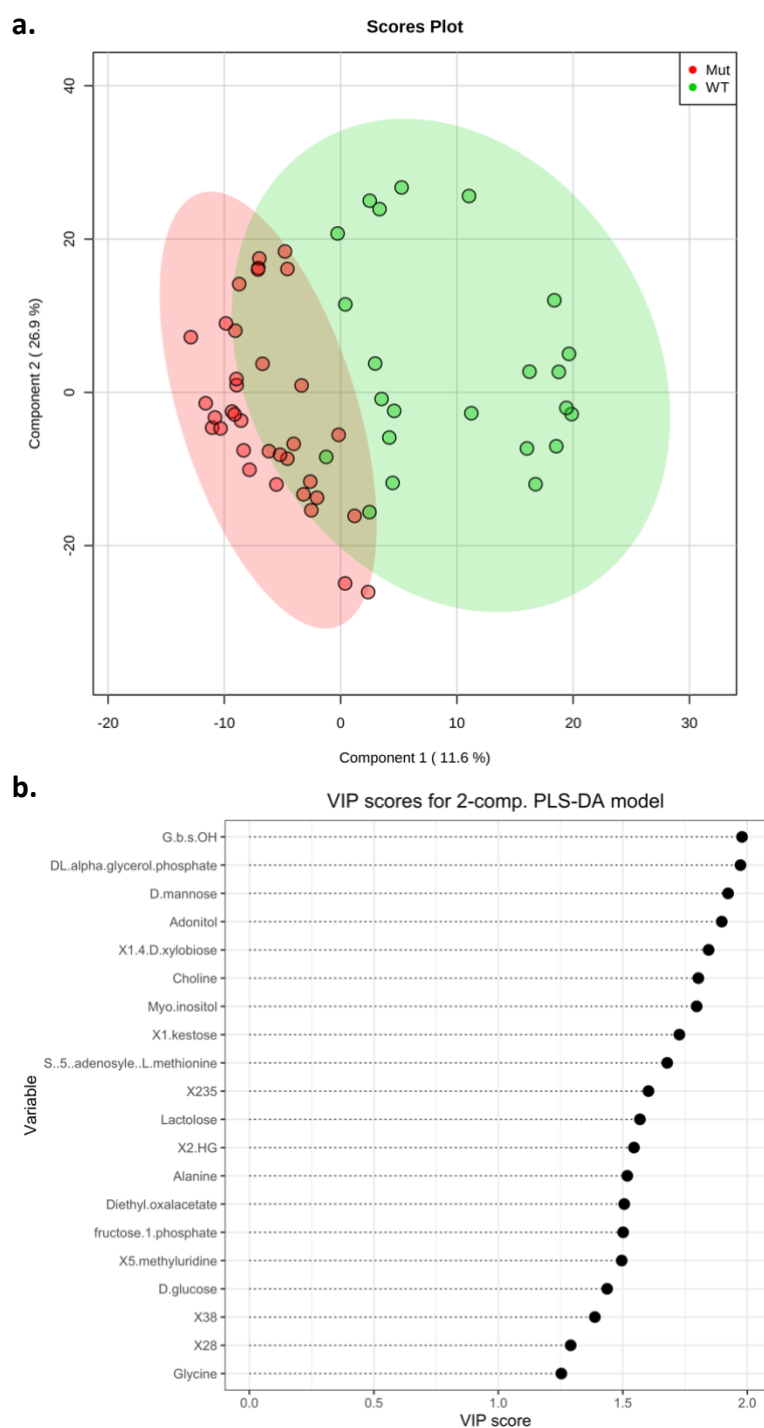
Fig 4.18 (a) 2D score plot of the first two components. There is a degree of separation between the WT and Mut groups based on the metabolite concentrations calculated from 2D NMR. Although there is some residual overlap between the two groups with this model, certain metabolites appear to be significantly altered between the two disease states. (b) VIP scores for the first component of the PLS-DA analysis. Here we can see a number of metabolites that were shown to be important ion the septation of IDH Mut and WT from the 1D multivariate analysis, alongside a number of other metabolites.

spectroscopy (fig 4.15). The results from the univariate analysis can be seen in figure 4.16. Of the 6 metabolites that were shown to be significantly altered when comparing IDH Mut and WT gliomas from the 1D NMR spectra, 4 metabolites were shown to be significantly altered between the two groups using the 2D quantification method, along with a further 18 metabolites. From the 2D quantification and univariate analysis 22 metabolites were significantly altered. The scatter plot and selected box plots can be seen in figure 4.16. As expected, 2-HG was found to be significantly altered, however it was not the most significantly different metabolite.

The results from the multivariate PCA analysis can be seen in figure 4.17. The first three PCs only account for 50% of the variability within these samples, and when the first two PCs are plotted on the 2D score plot, there is a significant overlap between the two groups. The results from the PLS-DA analysis provided more separation between the two groups (figure 4.18). Although the first three PCs only accounted for 46.5% of the overall variability, when considering the first two PCs plotted on the 2D score plot, there is a much clearer separation between the two groups. The VIP scores can be seen in figure 4.18, which shows that the choline peak contributed most to the separation seen in this model. When comparing this to the VIP scores calculated from the 1D analysis, myo-inositol again played a significant role to the separation, complementing the findings from 1D analysis.

4.3.2.3 Pathway Analysis

Following identification of metabolites that were significantly altered between IDH WT and Mut gliomas using both a 1D and 2D NMR spectroscopic approach, one can use this information in order to estimate upstream cellular pathways that have been subject to alteration in the presence of the mutation.

The pathway results from the metabolites identified as significantly different from the 1D NMR analysis can be seen in table 4.2 and in graphical form in figure 4.19. As the same pathways are being tested many times, the statistical p values needed to be adjusted for multiple comparisons. The total column in table 4.2 refers to the total number of metabolites within the pathway, and the hits is the number of metabolites from the input list that actually matched with metabolites known to be part of the pathway. From the 12 pathways that were found to have been affected by alterations in the metabolites highlighted by metabolomics statistics, only one metabolite from the input list was associated with each pathway.

	Total	Hits	Raw p	-log(p)	Holm-adjust	FDR	Impact
Synthesis and degradation of ketone bodies	6	1	0.012412	4.3891	0.99295	0.7671	0
Taurine and hypotaurine metabolism	20	1	0.040894	3.1968	1	0.7671	0.03237
Selenoamino acid metabolism	22	1	0.044909	3.1031	1	0.7671	0
Alanine, aspartate and glutamate metabolism	24	1	0.04891	3.0178	1	0.7671	0.05698
Glycerophospholipid metabolism	39	1	0.078494	2.5447	1	0.7671	0.0212
Inositol phosphate metabolism	39	1	0.078494	2.5447	1	0.7671	0.13703
Butanoate metabolism	40	1	0.080439	2.5202	1	0.7671	0.0048
Galactose metabolism	41	1	0.082382	2.4964	1	0.7671	0
Ascorbate and aldarate metabolism	45	1	0.090119	2.4066	1	0.7671	0
Glycine, serine and threonine metabolism	48	1	0.095887	2.3446	1	0.7671	0
Cysteine and methionine metabolism	56	1	0.11113	2.1971	1	0.8082	0
Aminoacyl-tRNA biosynthesis	75	1	0.1465	1.9207	1	0.97666	0

Table 4.2 Results from pathway analysis. Each of the pathways that have been identified as significantly altered between IDH WT and Mut only contain 1 hit from the input list of metabolites ascertained from 1D NMR metabolomics. Following multiple comparison adjustment, no pathways were found to be significantly altered between the two disease states from this analysis.
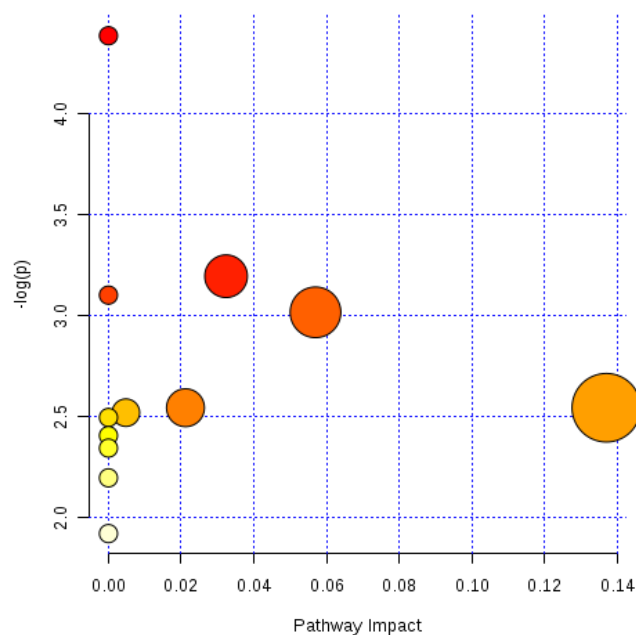
Fig 4.19 Graphical depiction of pathway result. X axis corresponds to pathway impact which is determined by the number and placing of nodes within the pathway. Y axis corresponds to -log10(p) determining the statistical significance of the alteration within the pathway. The size of the node corresponds to the total (number of nodes in pathway) and the colour reflects the significance level of the about metabolites within the highlighted pathway (red – higher significance, yellow – lower).

	Total	Hits	Raw p	-log(p)	Holm adjust	FDR	Impact
Galactose metabolism	41	4	0.00018678	8.5856	0.014942	0.014942	0.01992
Glycine, serine and threonine metabolism	48	3	0.0049228	5.3139	0.3889	0.19691	0.28435
Aminoacyl-tRNA biosynthesis	75	3	0.016933	4.0785	1	0.45154	0
Amino sugar and nucleotide sugar metabolism	88	3	0.02589	3.6539	1	0.51781	0.01122
Glycerophospholipid metabolism	39	2	0.033241	3.404	1	0.53186	0.05375
Starch and sucrose metabolism	50	2	0.05236	2.9496	1	0.69813	0.0765
Cyanoamino acid metabolism	16	1	0.1135	2.1759	1	0.97347	0
Taurine and hypotaurine metabolism	20	1	0.13991	1.9667	1	0.97347	0.03237
Riboflavin metabolism	21	1	0.1464	1.9214	1	0.97347	0
Selenoamino acid metabolism	22	1	0.15284	1.8784	1	0.97347	0
Thiamine metabolism	24	1	0.16558	1.7983	1	0.97347	0
Alanine, aspartate and glutamate metabolism	24	1	0.16558	1.7983	1	0.97347	0.05698
Porphyrin and chlorophyll metabolism	104	2	0.18101	1.7092	1	0.97347	0
Valine, leucine and isoleucine biosynthesis	27	1	0.18436	1.6909	1	0.97347	0
Glycolysis or Gluconeogenesis	31	1	0.20877	1.5665	1	0.97347	0
Pentose phosphate pathway	32	1	0.21476	1.5382	1	0.97347	0
Glycerolipid metabolism	32	1	0.21476	1.5382	1	0.97347	0.00714
Methane metabolism	34	1	0.22662	1.4845	1	0.97347	0
Glutathione metabolism	38	1	0.24984	1.387	1	0.97347	0
Nitrogen metabolism	39	1	0.25554	1.3644	1	0.97347	0
Inositol phosphate metabolism	39	1	0.25554	1.3644	1	0.97347	0.13703
Ascorbate and aldarate metabolism	45	1	0.28888	1.2417	1	0.97608	0
Lysine degradation	47	1	0.29968	1.205	1	0.97608	0
Primary bile acid biosynthesis	47	1	0.29968	1.205	1	0.97608	0.00822
Fructose and mannose metabolism	48	1	0.30503	1.1874	1	0.97608	0.02948
Pentose and glucuronate interconversions	53	1	0.33116	1.1052	1	1	0
Cysteine and methionine metabolism	56	1	0.34639	1.0602	1	1	0
Purine metabolism	92	1	0.50541	0.68239	1	1	0

Table 4.3 Results from the pathways analysis carried out using the list of 24 metabolites found to be significantly altered between IDH Mut and WT gliomas from 2D NMR metabolomics. As was seen in the pathway analysis using the 1D NMR identified metabolites, 2-HG has not been identified in any of the above pathways. Galactose metabolism was shown to be the alteration with the highest significance and incorporates 4 of the input metabolites from metabolomics analysis (Sucrose, D-glucose, D-mannose, myoinositol). Galactose metabolism was the only pathway that was significantly altered following multiple comparison adjustment

Furthermore, 2-HG was not highlighted as a significant metabolite in any pathways highlighted, and following correction for multiple comparison, no pathways were found to be significantly altered from the 1D analysis.

The results of the pathway analysis carried out on metabolites shown to be significantly altered between IDH Mut and WT obtained from 2D NMR metabolomics can been seen in table 4.3. The only pathway that appeared to be significantly altered following multiple comparison correction was galactose metabolism, which incorporated 4 of the inputed metabolites within the network of 41 metabolites that comprise the pathway. Once again, 2-HG was not implicated in any of the pathways found to be significantly altered from the input list of metabolites.

4.4 Discussion

4.4.1 NMR Metabolomics to Study IDH Mutation in Glioma

Here we present results that have demonstrated a metabolic phenotype in IDH Mut glioma tissue samples using 1D and 2D NMR spectroscopy. In total 6 metabolites were found to be consistently altered across all experimental and analytical methods, and these results can suggest why IDH tumours behave in the way that they do, and may facilitate the formation of new hypotheses for future investigation into the pathophysiology of the disease.

The importance of IDH mutation in glioma has been widely acknowledged and accepted for a number of years, which is reflected in its inclusion into the WHO diagnostic criteria for glioma in 2016. Although it is clear that the mutation is integral to diagnosis and prognostication, there is an unmet need to 1) find a reliable and reproducible way to non-invasively probe for the mutation 2) understand the pathophysiology of the mutation on progression of the disease. Metabolomics provides an ideal avenue of investigation for both these purposes. The systematic profiling of metabolite concentrations provides an indication of global cellular function and has the ability to reflect aberrations in a myriad of cellular processes such as genetic stability, gene transcription, translation and protein homeostasis. If we are to understand more about the influence of IDH mutation within glioma, then a global view of metabolic modifications is an ideal place to start. Moreover, the utility of non-invasive detection *in-vivo* via MRS may be aided by assessing for metabolites other than 2-HG that are significantly altered in the presence of the mutation, in order to form a detectable metabolic

'fingerprint'. The concept of combined metabolic biomarkers may benefit the acceleration of clinical translation of MRS diagnostic techniques, as currently the reliability and reproducibility of 2-HG detection methods is widely debated[25].

A number of previous studies that have looked into metabolic profiles of gliomas using a variety of 1 and/or 2D NMR spectroscopic methods at a range of magnetic field strengths[37,159,160]. Although it is well known that IDH mutation facilitates the production and subsequent accumulation of the oncometabolite 2-HG, quantitative information pertaining to overall metabolome alterations and phenotypic aberrations that expedite the progression of malignancy in the presence of the mutation is sparse. This is particularly true when considering information that has been ascertained from resected tumour samples. There are a number of studies concerned with the metabolic profiling of cultured glioma cell lines[37,161,162], however 1D cell culture does not reflect the true heterogenous tumour environment. It has been demonstrated that the heterogeneity of tumours, including gliomas, gives rise to discrete sub-populations of tumour cells, and that this is a significant contributor towards malignant progression[163–165]. It is therefore reasonable to address the question as to what influence IDH mutation may have on the development and advancement of the disease, and which metabolic shifts are consistent with the mutation for *in-vivo* diagnostic detection. Thus, performing the metabolomic studies on extracted tumour tissue samples would enable better understanding of the factors contributing to the pathophysiology.

Previous studies such as the investigation conducted by Hyung et al[159] included a number of tumour samples from a range of cellular phenotypic origins, including ganglioma and neurocytoma, which are of mixed glial and neuronal cell progenitor origin. As mutations in the IDH gene have only been implicated in gliomas and acute myeloid leukaemia, inclusion of neurocytoma or ganglioma is not relevant to address the question about the impact of IDH mutation in NMR metabolomic profile. Here we have included only those tumours that are classed as glioma by the WHO guidelines, encompassing astrocytoma, oligodendroglioma and glioblastoma, and in which, IDH mutations are commonly observed. Although these samples may have different glial origins, and therefore different phenotypic manifestations, we are scrutinising the role of IDH and inspecting for potential biological signatures that may contribute towards non-invasive detection. The involvement of IDH mutation in the prognosis and diagnosis of all gliomas is analogous throughout the various subgroups. In addition, when considering the search for a spectroscopic signature combining additional metabolite

alterations with 2-HG for *in-vivo* detection, it is desirable to assess for metabolic changes that occur independent of tumour subtype.

The initial phase of this investigation was carried out on a cohort of 17 tumour samples. The spectra that were acquired were analysed first using the spectral binning method as previously described. The results from the PCA showed there was some degree of separation between the Mut and WT groups, however there was still a significant amount of overlap between groups based on metabolite concentration. To investigate which peaks, if any, contributed to the degree of separation the PCA loading plot was examined. Loading plots allowed us to examine which of the original variables, which here are the metabolites, contribute most to the variation seen in the new PCs. The variation seen along the x axis is likely to have arisen due to biologically relevant factors, whereas the variation seen along the y axis is likely to be due to technical replicates. From inspection of this plot (fig 4.6) it was possible to determine which of the spectral bins had contributed to the separation seen in the PCA. Upon close inspection of the bins and the sections of each of the spectra that the bin boundaries bisected, a number of the spectral bins highlighted by the loading plot were not over a specific peak of interest, rather they contained varying amounts of residual baseline noise. Moreover, those bins that were flagged by the loading plot that were positioned over potential peaks of interest contained varying proportions of integrals of each of the peaks, as the spectra were not all aligned with each other. This indicates that the variation seen in the PCA that allowed for separation of the Mut and WT groups does not originate from biologically relevant distinction within the spectra. From this relatively rapid and simple analysis we were able to determine that the numerical data being used for metabolomics analysis was not appropriate, as it did not reflect metabolic changes within the samples, rather it was a reflection of spectral shift between each sample.

In order to ascertain data that did reflect alterations in the metabolome of these samples, we constructed a custom pattern file to determine the ppm boundaries for each of the peaks of interest. For this process, a publicly available template of metabolites of interest found in neuroblastoma cells was used. Using literature searches, this list of metabolites was edited via the addition and subtraction of a number of molecules, so as to include only those metabolites that are relevant to glioma. All 22 subjects included in the study were plotted over each other and the upper and lower ppm bounds for each of the identified peaks of interest were used to determine the spectral bin limits. When metabolite concentration data calculated
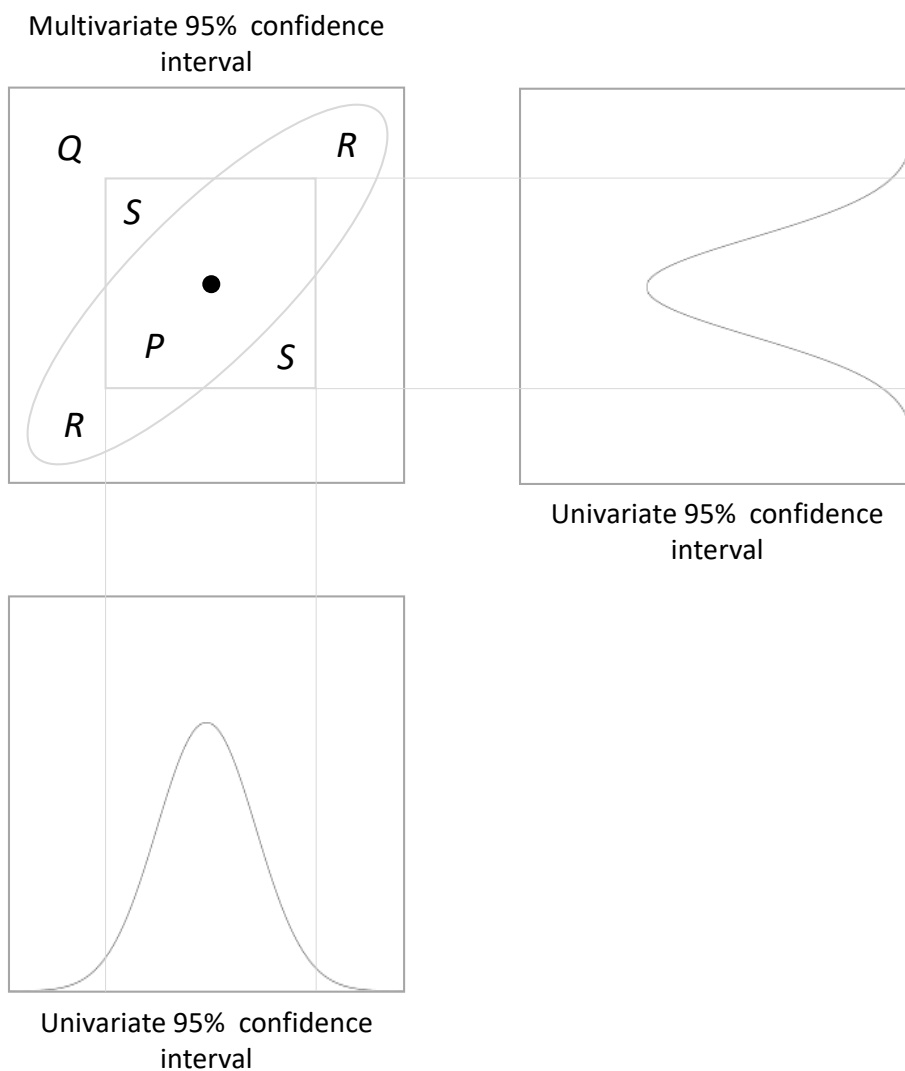
Fig 4.20 A geometric depiction of the relationship between univariate and multivariate hypothesis testing confidence intervals. The ellipse represents the 95% confidence interval of the multivariate distribution of the two metabolites A and B. The inner square represents the two independent 95% confidence intervals calculated from normal distribution of the two (uncorrelated) metabolites with the same variance[166].

from spectra using this method were used in an initial PCA analysis, the 2D plot showed almost complete overlap with no separation. When these data were analysed using a univariate approach, we were able to distinguish 35 metabolites that were significantly altered between Mut and WT groups. There are a number of reasons as to why significant differences between variables may be detectable univariately but not via multivariate testing and vice versa.

As multivariate analysis methods must consider all variables at the same time, there is often a large amount of information to handle. Information may therefore easily be masked by uninformative variables present in the data. Biologically relevant information is most often

attributed solely to a small subset of variables within large datasets consisting of hundreds of metabolites. If this subpopulation of biologically relevant variables is not previously known, then the requirement to measure as many variables as possible is irrefutable. Despite the measurement of a large number of variables increasing the probability of measuring something with biological significance, the burden of such an approach is that one must at the same time consider a large number of uninformative variables that may impede the power of the multivariate analysis[166]. In addition to the potential masking of biological data by uninformative variables, the difficulty in estimation of correlations and covariances may impact on the ability to detect significant differences when applying multivariate analysis when compared to univariate methods. This is a factor in the performance of multivariate methods, particularly when a large number of variables are being measured over a relatively small number of samples[167]. The use of multivariate methods requires that the true covariance matrix from the population from which the samples have been drawn are known. However, as this is not known, some estimate must be used. Estimating a covariance matrix across only a few samples is a known difficulty in statistics, and it has been shown that in cases where there are a large number of variables from a small sample size the usual estimation methods fail to generate reliable matrices[166]. An additional factor that may contribute towards the failure of multivariate testing to uncover significant differences where univariate testing performs well is that testing procedures from both methods may not overlap. To illustrate this let us consider two metabolites, A and B, belonging to two groups 1 and 2. In a univariate approach each of the metabolites A and B will be assessed by performing independent tests, such as a *t*-test. For this example, we will test the null hypothesis at the 95% confidence interval, whereby H_o is the difference between the means of metabolites A and B is 0. We accept H_0 if the mean (o) lies within the 95% confidence interval. In figure 4.20 the rectangular area represents the univariate 95% confidence interval. Using a multivariate method, the means of both metabolites A and B are assessed simultaneously. We are now assessing whether the means of both metabolites A and B (o, o) are within the simultaneous 95% confidence interval, which is represented by the ellipse in figure 4.20. At this point, it is clear from figure 4.20 that the 95% confidence intervals for both the univariate and multivariate methods do not completely overlap, and therefore there may be a difference in acceptance or rejection of H_0 dependant on location of the point (o, o). Figure 4.20 allows us to geometrically explain 4 instances that may arise when applying both univariate and multivariate methods:

1. The point zero lies within both the univariate and multivariate 95% confidence intervals (region *P*), and therefore application of either of these testing methods will both lead to acceptance of the H_0 and there is no distinction between the two testing groups.

2. The point zero lies outside both the univariate and multivariate 95% confidence intervals (region *Q*), indicating a significant difference at using either approach.

3. The point zero is inside the ellipse, but resides outside the rectangle (region *R*). Here application of a multivariate methods will not detect a significant difference, and we will accept H_0, however use of a univariate approach will lead to the detection of a significant difference in at least one of the metabolites, A or B, at the 95 % confidence interval.

4. The point zero is outside the ellipse but can be seen inside the rectangle (region *S*, figure 4.22). In this instance the multivariate test will detect a significant difference at the 0.05 level, whereas the univariate test will fail to recognise any distinction between groups 1 and 2 based on metabolites A and B.

The third point coupled with figure 4.20 describes why it may not be possible to detect significant differences using a multivariate approach, but it is possible to see discrimination between two testing groups using univariate methods. It is worth noting that in practice this is a more complex problem, as univariate testing requires correction for multiple testing, but put simply if we were to apply this to figure 4.20 it would result in broader univariate confidence intervals. Taking into account the aforementioned differences that can arise in the application of both statistical testing strategies, in addition to the fact that metabolomics is the practice of attempting to characterise and interpret biological phenomena based on the interrelation of metabolites, it may be argued that the most appropriate method for detecting biologically relevant information pertaining to differences between two testing groups is the employment of both univariate and multivariate analyses. As can be seen in our results, both methods can produce complimentary information, as well as additional information regarding statistical differences. Although the results from univariate and multivariate analysis of the data produced by application of a custom pattern file to set bucket limits for spectral binning contradicted each other, the complimentary combination of statistical approaches is nicely illustrated when considering the results generated from individually fitted spectra.

The results generated from the statistical analysis of the data ascertained from fitting each peak of interest in each spectrum show significant differences in 6 metabolites when using a univariate approach, and these results were complimented by the multivariate tests. Moreover, not only did the metabolites that were shown to be significantly altered in the univariate method yield significant results in the multivariate approach, a number of other potential metabolites were highlighted as contributing factors in the separation seen in the PLS-DA analysis. Considering the metabolites that have been shown to be significantly altered between the Mut and WT tumours in addition to 2-HG, both myo-inositol and alanine were consistent in both statistical analyses, as well as choline and 2-hydroxybutarate. In addition to the 1D analyses, we carried out quantitative metabolomics on 2D HSQC data and our findings complement the 1D results and provide additional information pertaining to metabolic alterations that have been observed in the presence of the mutation, whereby 22 metabolites were found to be significantly changed. Utilising heteronuclear 2D NMR allowed us to unequivocally identify the coupled metabolites, which is generally a problem in the over-crowded 1D spectrum. Decoupling the peaks using ^{13}C chemical shifts along the second dimension allowed us to observe alterations in metabolites that may have previously been overlapped by those metabolites found in greater abundance in both the healthy and diseased brain tissue. It is important here to note, however, that this method of 2D spectroscopy is not viable *in vivo*, especially not at clinical field strengths.

One drawback to the inclusion of 2D NMR in quantitative studies is that peak volumes are dependent on a number of peak-dependant parameters such as heteronuclear coupling constants, pulse delays and off-resonance effects. These confounding factors are much more intricate in 2D when compared to 1D quantitative NMR, where the peaks are solely dependent on the number of equivalent spins and the concentration of the analyte, given that full spin relaxation has been achieved[168]. The multi-pulse nature of 2D NMR has led to speculation within the spectroscopy community that such sequences may not be used to determine absolute concentrations and cannot be employed in a quantitative manner. In this metabolomics approach we were not concerned with absolute concentrations, rather we were interested in whether a given metabolite (or set of metabolites) is more abundant in one group of samples when compared to the other. For this, normalised values to the total intensity allow for inter-subject comparison in order to determine if there is fold-change in metabolite levels between WT and Mut samples.

Another limitation of multidimensional quantitative NMR is the relatively long acquisition time due to the requirement to obtain an adequate number of t_1 increments to sample the indirect F_1 dimension in order to yield a sufficient resolution. In addition to the obvious impact on usage of the spectrometer, these long acquisition times can impact on the quality of the data collected, as spectrometer instabilities cumulate over time. These instabilities can include variation in receiver gain, phase, B_0 inhomogeneity, lock instabilities and variations in pulse angle[169,170]. These perturbations, whist negligible in the direct F_2 dimension where points are acquired over a short space of time, can build up as additional noise within the indirect F_1 dimension where acquisition requires long TR values. Consideration of these limitations is important when incorporating 2D quantitative approaches within metabolomics. Here we have combined both 1D and 2D to provide complementary information pertaining to metabolic aberrations, rather than relying on the quantitative capacity of just one of the methods of investigation. Further to this, with respect to clinical investigation, this type of spectroscopy is not possible in humans at clinical magnetic field strengths. The incredibly low natural abundance of ^{13}C would require scan times of many hours in order to acquire enough signal to produce a spectrum that is interpretable. The spectra in this experiment were acquired on a 700 MHz magnet, which is equivalent to 16.5 tesla and required 8 hours of acquisition time, which already far surpasses clinical scan times. At 3 tesla to obtain spectra with reasonable SNR for clinical interpretation, this would be unfeasibly long.

The results of the 2D analysis showed that univariately 22 metabolites were indicated as significantly altered between IDH Mut and WT tumours, including 2-HG, alanine, myo-inositol and choline, all of which had been highlighted in the 1D univariate analysis. Multivariate testing showed separation between the two testing groups, where the PLS-DA performed better than the PCA, which is to be expected when considering reduced sample size of a heterogeneous nature, such as here. Whilst the unsupervised PCA method is excellent at providing an unbiased way of reducing data dimensionality, it can only perform optimally when the intra-group variation is sufficiently less than the inter-group variation. The supervised approach of PLS-DA relies on class membership of each variable in order to reduce dimensionality and determine separation. The VIP scores generated from the PLS-DA provided additional complementary information implementing myo-inositol, alanine and choline as principal metabolites in the separation between the two testing groups.

Our results showed that using a combination of both 1D and 2D NMR spectroscopy, in addition to both multivariate and univariate analysis, we have been able to identify six metabolites that are consistently altered in the presence of IDH mutation across all analyses. Furthermore, we have shown that by combining the various analyses of the different spectra, we were able to distinguish a number of other metabolites that may contribute towards our understanding of the clinical phenotype that could have been overlooked should only a single approach have been adopted. In addition to the expected 2-HG, we observed increases in level of myo-inositol and 3-hydroxybutyrate as well as decreases in choline, alanine and methyl-succinate.

Myo-inositol and its phosphate derivative-namely inositol hexasixphosphate (InsP6)-have been implicated in a wide variety of diseases and are gaining increasing momentum within the field of oncology[171]. Myo-inositol may be attributed to a multiplicity of functions throughout the human body and has been shown to have regulatory influence over processes such as chromatin remodelling, p53 activation, mRNA transcription and cytoskeletal configuration to name but a few. The significant increase in myo-inositol in IDH Mut tumours may contribute towards the increased radio-chemosensitivity and overall survival rate *via* a number of molecular mechanisms. First, *via* its influence over the p53 network within the apoptosis pathway. The selective activation of p53 is a crucial step in the commitment of cellular fate to apoptosis/proliferation inhibition[171]. It has been shown that an increase in Ins6/myo-inositol increases levels of p53 several fold, and that activation of p53 in turn, increases the endogenous production of further myo-inositol, re-enforcing apoptotic commitment *via* a positive feedback loop[172]. In addition, the observed increase in intracellular myo-inositol is known to support growth inhibition, and may contribute towards the prevention of the development of chemo-resistant cells[172]. Further to dysregulation of p53 by excess myo-inositol build-up in Mut tumours, increased survival may also be conferred via inhibition of the PI3K/Akt pathway. PI3K induces the phosphorylation-dependant activation of Akt, which in turn functions as a regulator of cell cycle progression, dictating the aggressive and invasive nature of cancer cells[173,174]. Myo-inositol and InsP6 have been shown to significantly reduce expression of both the protein and mRNA of PI3K expression, as well as inhibiting Akt phosphorylation[173–175]. With this in mind, the increase in myo-inositol that we observed within the Mut tumours may play a role in the protective phenotype that is observed clinically, however the absolute cause of myo-inositol increase within the Mut tumours is yet

to be elucidated, as our pathway analysis failed to identify a significantly altered pathway with its inclusion. Further investigation is required to uncover the mechanisms by which this particular metabolite functions as both a barrier to malignant progression and facilitator for increased radio-chemosensitivity.

In addition to the increase in myo-inositol within Mut tumours, we observed a significant decrease in alanine concentration. Cancer cells characteristically exhibit non-oxidative D-glucose metabolism (glycolysis), whereby mitochondrial respiration is impaired, which according to Warburg's hypothesis, is a driving force behind malignancy[176]. Although there is speculation as to whether the Warburg effect actually causes malignancy, it has been demonstrated that impairing D-glucose metabolism in malignant cells induces oxidative metabolism, which has been implicated in the slowing of malignant growth[177,178]. Further to this, it has been shown that the conversion of L-pyruvate into L-alanine via the enzyme L-alanine aminotransferase (ALAT) is a crucial step in glycolysis, and therefore a factor in the development and progression of the malignancy[179]. Moreover, the administration of ALAT inhibitors can shift the metabolism of malignant cells towards a more oxidative phenotype[179]. Although here we were unable to determine the mechanism by which IDH mutation causes a decrease in alanine, the reduction in the metabolite concentration may reflect a shift away from glycolysis towards oxidative metabolism, whereby less pyruvate is undergoing conversion into L-alanine. This possible shift in metabolism may facilitate the decrease in the invasive nature of the cancer cell, slowing tumour growth *in vivo*.

The observed decrease in choline that we observed in the Mut tumours when compared to the WT may be coherent with the slower growth rate and increased overall survival of those patients harbouring the mutation, as it has previously been shown that increased choline correlates with increased cell density in glioma[180]. Furthermore, it has been demonstrated that choline is a crucial step in the synthesis of the phospholipid component of cell membranes, and that increased phosphocholine is intrinsically linked to increased cellularity and proliferation index in glioma[181,182].

In addition to the above alterations, we also observed an increase in the ketone body 3-hydroxybutyrate. It has been demonstrated that glioma cells that are under glucose restriction are unable to utilise ketone bodies for metabolism, suggesting a survival disadvantage[183]. This has led to the proposed benefits of a ketogenic diet in conjunction with chemo/radiotherapy for the treatment of glioma. Here, due to limited meta data collection,

we are unable to determine whether the increase in 3-hydroxybutyrate is a cause of a change in diet of the patient, or whether this increase is due to an intrinsic shift in intracellular metabolism in the presence of the mutation. Either due to external or internal factors, this metabolic observation is likely to contribute to the increased overall survival and obtaining detailed participant information would be of major benefit to future metabolomics studies of this nature.

The consistent observation of a reduction in methyl succinate may also fall in line with the favourable clinical phenotype of IDH Mut patients, as it has been shown that the accumulation of succinate, which is a product of hydrolysation of methyl-succinate, can become oncogenic[184]. Increased levels of the metabolite have been demonstrated to increase genome-wide histone methylation and accumulation of hypoxia-inducible factor-α, both of which serve as oncogenic factors to push cells towards a malignant state[184]. Therefore, the reduction of this metabolite in the Mut tumours may aid in the acquisition of a favourable phenotype.

In addition to our statistical analysis of the spectra, we also carried out pathway enrichment analysis. This novel statistical approach uses online databases of metabolic pathway, available through metaboanalyst. It assesses which of these pathways are statistically enriched by any of the metabolites identified as significantly altered in the presence of IDH mutation, in the NMR experiments[185]. One of the major limitations to this approach with regards to this experiment, is that 2-HG, the major metabolic characteristic of IDH Mut tumours, does not feature in any of the pathways within the database. Furthermore, this tool has been constructed as a complimentary tool to guide further investigation. The method employs the use of statistics in order to suggest metabolic pathway shifts, rather than biological data itself. In conclusion, this approach did not contribute much benefit towards this study, as the metabolic pathways concerned are unknown.

A number of limitations influenced the potential scope of this study. Firstly, the sample population was both small and unbalanced. We had originally set out to include a total of 40 patient samples, with approximately 20 from each testing group. The tissue collection strategy, working alongside clinicians, took time to establish and identifying suitable patients for inclusion before they underwent surgical resection was difficult. As such, only 22 patients were included in the analysis. In addition to the small sample size, the heterogeneity of the tumour samples, in size and types of tumour tissue that were included (blood vessels, lipids such as

myelin, necrotic regions) may have introduced unwanted variability into the data. Finally, the previously described 2D NMR experimentation and analysis was carried out using the same metabolomics approach as the 1D. 2D NMR, however, inherently presents a larger number of variables for analysis, and consequently will additionally contain a larger number of uninformative variables for statistical analysis when looking at all the peaks identified. The 2D NMR sequence could have been optimised and tuned to the peaks that were previously identified in the 1D experiment, in order to further ratify the results.

4.5 Concluding Remarks

Mutations in the IDH1/2 gene in glioma have been shown to facilitate the aberrant production of the oncometabolite 2-HG, which has been suggested as a putative biomarker for the non-invasive detection of the mutation *in vivo*. Furthermore, those tumours that possess the mutation have increased survival rates and have been shown to respond more favourably to radio- and chemotherapy. Here we have assessed the cellular metabolome in glioma samples in order to further understand the impact of the mutation on the development and progression of the disease, whilst additionally searching for other spectroscopic markers alongside 2-HG that may be used as a biomarker in the detection of the mutation. We have demonstrated that using high resolution 1D and 2D high-resolution NMR spectroscopy, a number of metabolites, including myo-inositol and alanine, were significantly altered between IDH Mut and WT tumours in addition to 2-HG. The viability of the use of these alterations as biomarkers needs to be further explored in order to determine reproducibility, moreover investigations into the recapitulation of these observations *in vivo* is required to assess translation potential. However, the shifts in metabolic phenotype that have been discerned may be used to form hypotheses in the investigation as to how IDH Mut tumours display increased survival and treatment response. A deeper understanding of the molecular mechanism driving the progression of the disease and the influence of individual molecular changes may guide the development of targeted strategies for the treatment of glioma, as we move further towards the age of personalised cancer intervention.

5 *In-vivo* semi-LASER MRS for Detection of 2-HG in Glioma

5.0 Synopsis

The production and subsequent accumulation of the oncometabolite 2-HG, by IDH Mut glioma cells has been suggested to be a putative biomarker for the non-invasive detection of the mutation *via* MRS. This study set out to explore the clinical utility and translatability of a novel semi-LASER MRS pulse sequence with a TE of 110ms, which has been optimised for the *in vivo* detection of 2-HG, and therefore IDH mutation.

In order to carry out this investigation we originally planned to recruit a total of 40 patients who would be scanned using a novel semi-LASER sequence prior to surgery. The mutational status of the patient would then be confirmed by a pathologist. The MRS sequence in question was developed by Uzay Emir at the fMRIB group at Oxford University[186]. As such it was shared with our group for the purpose of this study and had to be installed on a clinical system at the Walton Centre NHS Foundation Trust, where the work was carried out. Although delays in ethical approval and technical issues with the installation and running of the sequence stalled the commencement of the study, a number of spectra of varying quality, from a range of anatomical locations, where acquired from glioma patients. The spectra were analysed in the commercial software LCModel, using a custom basis set containing 2-HG that was also shared with us by the fMRIB group led by Uzay Emir[186,187]. This work focuses on the clinical translation of novel MRS advancements in order to strengthen radiological diagnostic practices.

5.1 Introduction

5.1.1 Detection of 2-HG *via* MRS

There are few known biomarkers that are as reflective of the underlying molecular changes seen in tumour pathology as 2-HG. In fact, accumulation of this metabolite is the only known direct consequence of a genetic mutation that can be detected and quantified using non-invasive imaging techniques. The markedly increased concentration of 2-HG in IDH mutant tumours has been demonstrated to correlate with tumour cellularity, and its accumulation may also be used to guide tissue sampling during surgery[16]. Although 2-HG represents as an attractive marker for diagnosis and monitoring of disease progression, unambiguous detection via MRS has proven difficult to establish. Complex spectral overlap by a number of metabolites, such as glutamate, glutamine and gamma-aminobutyric acid (GABA) (Figure 5.1), found in

abundance within healthy brain tissue, often confound the identification and detection of 2-HG as well as compromise accurate quantification of metabolite concentration.

2-HG consists of five non-exchangeable scalar coupled protons comprising a complex 5-spin system (Figure 5.1)[188]. In addition to the 5-spin system, the J-coupling pattern of 2-HG leads to several multiplets with peaks identifiable at three spectral locations at a field strength of 3T centred around 4.02 ppm (H2), 2.25 ppm (H4 and H4') and 1.9 ppm (H3 and H3')[189]. The largest identifiable peak is located at 2.25 ppm due to the H4 and H4' protons that resonate proximally to each other[190]. However, there is a severe spectral overlap from a number of other physiologically occurring metabolites at this location, including glutamate (Glu), glutamine (Gln) and N-acetlyaspartylglutamate (NAAG), among others[16,191]. The influence of these overlapping resonances becomes more prevalent at clinically used field strengths (1.5-3T), as lower fields and shorter acquisition times lead to inadequate spectral resolution, resulting in complex spectra that are challenging to resolve[189,192].

Several methods for detection of 2-HG *in-vivo* have been proposed to optimise MRS for this application[16,18,19,186,193–195]. These encompass a range of acquisition and post-processing protocols
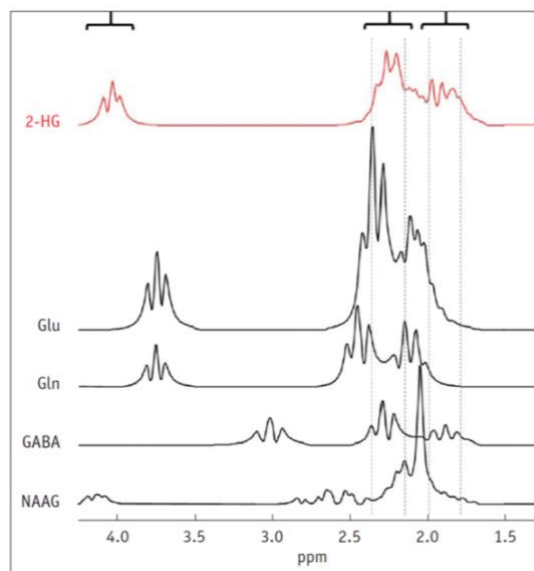

Fig 5.1. Simulated MR spectrum of 2-HG and overlapping metabolites. The largest 2-HG peak at 2.3 ppm is overlapped by a number of other metabolites found in abundance in brain tissue. 2-HG = 2-hydroxygluarate, Glu = Glutamate, Gln = Glutamine, GABA = gamma-aminobutyric acid, NAAG = N-acetlyaspartylglutamate. Adapted from Kim et al.[196]
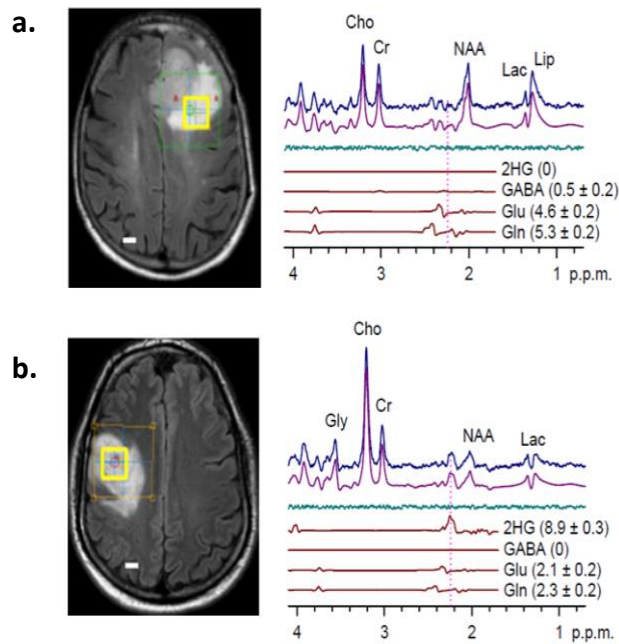
Fig 5.2. In vivo MR spectra from glioma patients acquired using single-voxel Echo Time (TE) 97 ms PRESS sequence at 3T, fitted using an LCModel. **(a)** grade IV Glioblastoma (GBM) IDH Wild Type (WT), notice the absence of the 2-HG peak at 2.25 ppm; **(b)** grade III Oligodendroglioma IDH1R132H, 2-HG peak at 2.25ppm. Taken from Choi et al.[16]

designed to eliminate the confounding spectral overlap, and reliably quantify the 2-HG concentration in patients with IHD mutant glioma. The spectral overlap observed at the multiplets generated by the H4 and H3 peaks may induce a high rate of false positives when attempting to identify 2-HG if spectra are acquired via conventional ^{1}H-MRS methods, and fitted using standard post-processing techniques.

The feasibility of detection of 2-HG *in-vivo* at clinical field strengths, using a standard single-voxel point-resolved spectroscopy (PRESS) sequence with a TE of 30 ms has been reported by Pope et al.[19] The acquired spectra were fitted with the LCModel software (http://s-provencher.com/lcmodel.shtml) using standard spectral fitting algorithms[187]. Using this method, the authors[19] reported a 26% false positive detection rate in IDH-WT tumours, due to severe overlap of the 2-HG signal with Glu and Gln at 2.25 ppm. Although short echo PRESS sequences are widely available in most clinical scanners, the high false positive detection rate limits the practical potential of this method for detection of 2-HG. Ex vivo quantitation via liquid chromatography-mass spectrometry (LC-MS) did, however, show good correlation between

MRS 2-HG measurements and true alterations in 2-HG concentration[19]. Additionally, although no changes in Glu or Gln could be observed in the IDH mutated subjects, there were additional

a.

b.
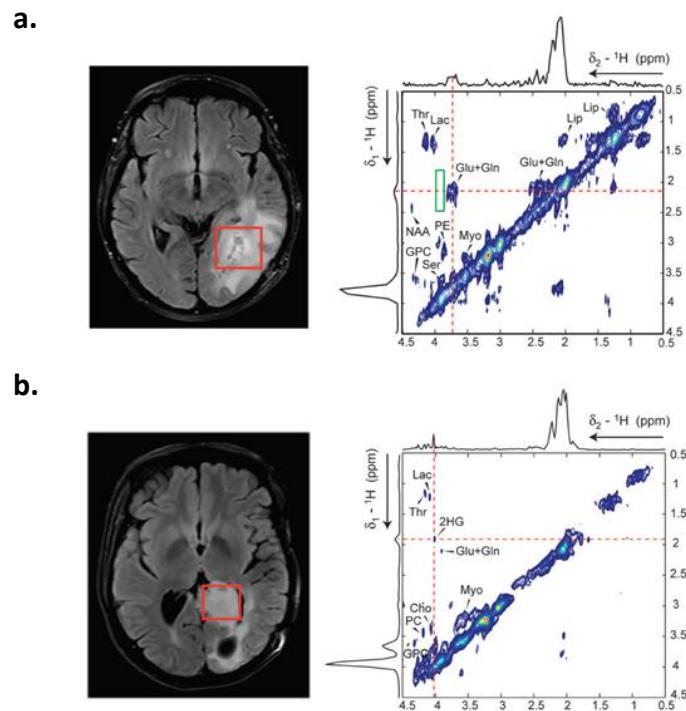
Fig 5.3. 2D Correlation Spectroscopy (COSY) of human patients acquired at 3T, where the $3 \times 3 \times 3$ cm^3 voxel was placed over the tumour region, shown by the red box. **(a)** primary GBM patient IDH WT, there is a clear absence of cross peak in the region of 2-HG (demarked by green rectangle); **(b)** grade III anaplastic astrocytoma patient IDH1^{R132H}; the 2-HG signal is unambiguously identifiable at the centre of crosshair intersection. Adapted from Andronesi et al.[191]

metabolic alterations such as increased Choline (Cho). Cho levels were found to be further elevated in WT when compared to Mut tumours, a measurement validated via LC-MS, and a finding in line with previous studies that have examined metabolome-wide alterations in IDH mutated tumours[19,197]. Such observations via MRS may be applied to furthering our understanding of the underlying mechanisms of the pathology, as elevated Cho levels have previously been correlated with tumour cell density, due to its involvement in cell membrane formation and proliferation[198,199]. Although the clinical utility of the short TE PRESS acquisition method is limited by the high false positive detection rate of 2-HG, it may be applicable within a research setting, aimed at evaluating the role of IDH mutation in development and progression of glioma pathology.

Improved spectral resolution and reduced overlap can be achieved when applying a longer TE to the standard PRESS sequences in order to take advantage of the difference in J-

evolution seen in coupled spin systems[200]. The differential variation between coupled-spin system signal has been shown to allow adequate suppression of overlapping signals from background metabolites such as Glu and Gln, whilst maintaining sufficient signal-to-noise for accurate quantification of 2-HG (Figure 5.2)[201]. To take advantage of this method of spectral editing, an optimized PRESS sequence with a TE of 97 ms has been proposed[16]. Although spectral editing of uncoupled spins requires implementation of the longest possible TE to eliminate baseline noise, the complex 5-spin system of 2-HG requires optimization of interpulse durations and a specific TE in order to minimise signal loss from target metabolites whilst eliminating background noise[202]. To determine the optimal TE for isolation of the 2-HG signal, Choi et al.[16] conducted quantum mechanical simulations. It was concluded that a shorter first subecho time and longer second subecho yielded the greatest H4 resonances, cumulating to a TE of 97 ms (Figure 5.2)[16]. The study reported 100% specificity and sensitivity for the detection of 2-HG. Furthermore, various voxel sizes were tested to ascertain optimal parameters, whereby six out of the six voxels >8 mL in volume belonging to WT tumours did not show any 2-HG signal[16].

The potential role of difference editing in MRS to untangle the 2-HG signal was also investigated. Difference editing, much like the long TE approach, takes advantage of J-coupling to generate spectral contrast[203]. This technique involves the application of frequency-selective radio-frequency pulses tuned at the H3 spins (1.9 ppm), the weak coupling partner to the H2 spins located at 4.0 ppm[16]. Alternate application of such pulses allows for selective 180° rotation of the H3 spins, producing a number of uneven H2 multiplets in the acquired subspectra that may then be subtracted from the overall signal, resulting in the cancellation of all other resonances unaffected by the 180° rotation. In the study carried out by Choi et al.[16], difference editing was also applied to six subjects and yielded 100% sensitivity. However, employment of this method may not provide absolute metabolic concentrations as it incurs imperatively long total TE, and, as such, may be subject to quantification errors that occur owing to signal loss due to relaxation[204]. Such errors associated with relaxation reduce the potential of the editing technique to track 2-HG for disease progression and appropriately monitor drug response.

In addition to the TE 97 ms PRESS sequence at 3T, the same group have recently proposed an optimized extended TE sequence at 7T[205]. The suggested method incurs a TE of 78 ms, taking into account the altered relaxation times at higher field strengths. The authors

directly compared both the 3T and 7T methods, whereby five patients were scanned at both magnetic fields within the same week. It was apparent that application of the optimized PRESS sequence at higher field strengths conferred increased signal – noise ratio (SNR), a point that was demonstrated by the reduction in mean 2-HG Cramer–Rao Lower Bounds (CRLB), a measure of spectral fitting, from 8% at 3T to 5% at 7T[205]. Importantly, it was established that the mutual dependence of 2-HG and GABA signals were greatly reduced at 7T compared to 3T. The peak resonating at 2.25 ppm experiences a significant amount of spectral overlap from GABA resonances, and application of an extended TE PRESS sequence at 7T induced the opposite polarity and narrowing of signals from both 2-HG and GABA, allowing for enhanced signal separation when compared to the TE 97 ms PRESS at 3T[16,205].

To further understand the clinical applicability of the TE 97 ms PRESS sequence, de la Fuente et al.[200] demonstrated the limitations of the MRS sequence in detecting 2-HG, and the ability of MRS to non-invasively monitor response to cytoreductive treatments. They indicated that increased sensitivity for 2-HG-detection when applied to tumours with a volume >3.4 mL. In addition, they investigated the potential role of this sequence in monitoring response to glioma treatment, demonstrating that 2-HG levels correlated with tumour cellularity as measured by increased apparent diffusion coefficient and histopathology[200].

Unambiguous detection of 2-HG in glioma patients has been reported by using a 2D correlation spectroscopy (COSY) sequence at a field strength of 3T, to counteract the known cofounding spectral overlap observed in standard 1D MRS spectra (Figure 5.3)[191]. 2D COSY allows for separation of the chemical shifts of spins that are scalar coupled to each other, and exploits the unlikely possibility that two metabolites would share identical chemical shifts in two dimensions. 2D COSY is particularly well suited for identifying 2-HG, as the cross peak generated by the metabolite appears in a spectral location shared by no other resonance. It was shown that 2D COSY was capable of unambiguous detection of 2-HG, providing full spectral information for accurate and reliable quantification. The abundance of spectroscopic information supplied by 2D methods represents a compromise between accuracy of measurements and complexity of acquisition. Although the most sensitive and specific detection method for identifying 2-HG, the clinical applicability of such a technique is limited by the extended acquisition time. This is an important factor when considering magnetic resonance imaging (MRI) scan time for patients, as current 2D COSY sequences are unlikely to be accommodated alongside clinical MRI brain tumour protocols within standard practice.

2D L-COSY MRS at 7T has been used to not only identify the resonant peaks of 2-HG, but to further assess and resolve the spectra of other brain metabolites that may experience a change in concentration in the presence of IDH mutation (Figure 5.4)[195,206]. The 2D L-COSY method
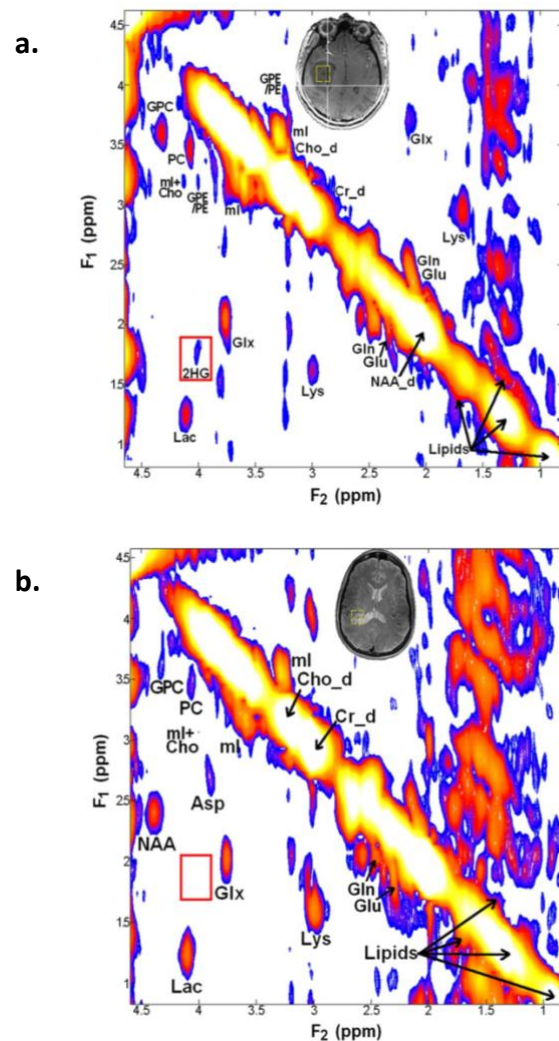
Fig 5.4. 2D L-COSY of human glioma patients at 7T. **(a)** spectrum acquired from an 11.0 mL voxel from a IDH mutant grade III anaplastic astrocytoma patient; **(b)** spectrum acquired from an 8.8 mL voxel form a patient with a WT grade I ganglioglioma patient. The increased field strength not only benefitted the spectral separation of the 2-HG cross peak, but also allowed for increased spectral resolution and reliable quantification of other cancer associated metabolites such as Lactate. Taken from Biomed Central[195].

provided unambiguous detection of the metabolite in question; moreover, the higher field strength of 7T allowed for a reduction in voxel size and proportionally higher spectral separation, facilitating increased specificity of detection and quantification of 2-HG[195]. Whilst the clinical translational potential of the suggested 7T method is limited due to the high field strength, the 2D L-COSY proposed by Verma et al. may prove an excellent tool to further

appreciate and investigate the shift in the metabolic profile of IDH mutated tumours. For example, the study demonstrated reliable quantification of increased levels of lactate, a useful tool in grading tumours, as it has been shown to reflect changes in glycolysis and tissue
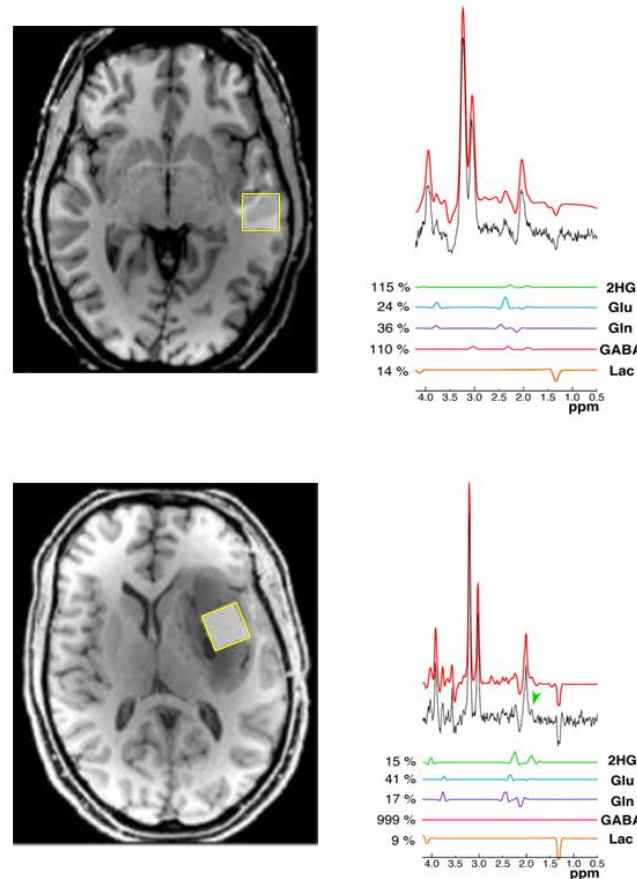
Fig 5.5. In vivo spectra acquired at 3T using an optimized semi-LASER sequence with a TE of 110 ms, the voxel was placed over area demarked by yellow box with LCModel fits and Cramer–Rao lower bounds (%) displayed for Glutamte, Glutamine, Gamma-anminobutyric acid and Lactate shown. **(a)** IDH WT glioma patient, there is an unambiguous absence of 2-HG peak; **(b)** IDH1R132H glioma patient where the 2-HG peak is easily identifiable. Adapted from tomography[186]

perfusion[206]. Reliable quantification of lactate concentration has previously proven problematic when applied to standard MRS methods, due to intense overlapping resonances from lipids often found in brain neoplasms that may be circumvented by the use of 2D L-COSY[195,207].

Although 2D MRS presents as an attractive method for accurate, reliable and sensitive detection and quantification of 2-HG, along with other glioma-associated metabolites, the clinical potential of this modality is limited when considering the incorporation of non-invasive metabolic profiling into diagnostic procedures. An alternative non-invasive method for

detection of 2-HG utilizes an optimized semi-localization by an adiabatic selective refocusing (semi-LASER) sequence at 7T capable of providing quantitative measurements adequate enough to differentiate between cytosolic IDH1 mutant and mitochondrial IDH2 mutant[18]. Spectral changes induced by the presence of IDH mutation were characterised, and feature abnormalities were input into Fisher Linear Discriminant Analysis, where the resulting plot categorised two distinct cluster patterns, pertaining to IDH1 and IDH2 mutation. This detailed analysis was possible as a result of vastly improved localisation of the spectral volumes facilitated by employing a semi-LASER pulse sequence in combination with outer volume suppression, to eliminate contaminating signals from the area outside the volume of interest. The resultant spectra were free from artefacts normally generated from areas of surrounding tissue, and exhibited a flat baseline between the spectral locations of 1.6 ppm and 4.2 ppm[18]. In addition to the method for detection and quantification proposed at 7T, the same group has recently established a method for improved localisation of 2-HG detection at 3T by applying an optimized long-TE semi-LASER pulse sequence[186]. The use of broadband adiabatic localisation on strongly coupled spins allows for the acquisition of 2-HG spectra with a peak identifiable at 1.9 ppm that retains many features of the peak expected using adiabatic hard pulses[186,207]. The outlined 3T semi-LASER method has been directly compared to the TE 97 ms PRESS sequence developed by Choi et al.[16]. The loss of the H3/H3′ peak as seen in TE 97 ms PRESS was circumvented by the use of adiabatic refocusing pulses in the semi-LASER sequence. Application of such refocusing pulses result in significantly reduced compartmental artefacts that would normally arise from extended J-evolution times under enhanced chemical shift displacements seen when narrow bandwidth RF pulses are applied[186,208]. The overall effect is total refocusing of the resonance at 1.9 ppm (H3 spins), which, when combined with diminished Glu and Gln peaks, allows for reliable detection of 2-HG (Figure 5.5). The significance of this method is highlighted by its translational potential and clinical applicability. Semi-LASER is a sequence that is increasingly being made available on 3T MRI within the clinical setting, and could be used for non-invasive in vivo assessment of IDH mutational status.

The mutated IDH pathway represents a possible target for novel glioma therapies, and, as such, there has begun a significant shift towards development of imaging strategies for the reliable longitudinal assessment and quantification of 2-HG levels during treatment and throughout clinical trials. Andronesi et al. propose a 3D functional spectroscopic mapping technique for the monitoring of 2-HG during clinical trials of drugs that may target the mutated

pathway[194]. In order to quantify spatiotemporal changes of metabolite levels associated with IDH mutated glioma, a novel analysis metric was proposed, termed functional spectroscopic mapping (fSM)[194]. The resultant 3D analytical framework is able to provide comprehensive metabolic information over time. Standard single voxel or single slice MRS may often incur a sampling bias when probing for 2-HG in longitudinal analysis due to positional errors; however, 3D metabolite mapping is able to reduce the variance of spectral measurements obtained throughout the trial period[194].

In addition to the in vivo detection methods described above, efforts to characterize the metabolic profile of IDH1 mutated gliomas using *ex vivo* spectroscopic techniques have also been correlated with in vivo MRI parameters[209]. In a study conducted by Elkhaled et al.[209], image-guided biopsies were acquired and subsequently evaluated using magic angle spinning NMR spectroscopy. Although they failed to determine a link between tumour grade and relative abundance of 2-HG, a number of metabolites associated with tumour progression were correlated with the oncometabolite levels. Choline-containing species are some of the best spectroscopic markers for tumour evaluation and have been significantly utilized to determine tumour cellularity. In the study conducted by Elkhaled et al.[209], total choline was shown to significantly correlate with 2-HG levels, indicating the role of detection of the metabolite in determining phenotypic characterisation of glioma[209]. Moreover, the histopathology of the tissue samples was assessed, where both immunostaining and histological investigations were used to assess for a number of parameters, including mitotic activity, relative tumour content and cellular density. 2-HG levels were found to correlate with a number of histopathological markers for mitotic activity, elevated tumour score and cellular density[209]. The strong correlations between 2-HG and increased cellularity of IDH mutant tumours fall in line with the hypothesis that there is increased cellularity in tumours where 2-HG is present. These findings may also play an important role in the further development of in vivo MRS for both diagnosis and treatment response monitoring. If 2-HG levels can confidently aid in non-invasively determining histopathological characteristics of glioma, the acceleration of novel treatments and the diagnostic accuracy of pre-surgical MRI could greatly benefit.

Here we have employed the semi-LASER acquisition protocol at 3T described by Berrington et al, in order to determine if MRS is able to reliably and unambiguously detect 2-HG within a clinical setting.

5.1.2 Scope of Study

The aim of this study is to determine whether the MRS sequence was able to unambiguously detect 2-HG, and therefore identify IDH mutation non-invasively. This is achieved by comparing MRS detection results with gold standard diagnostic histopathological results.

5.2 Methods

5.2.1 Patient Recruitment

This study was approved by the Walton Centre Research Tissue Bank research ethics committee (REC) (Reference number 15/WA/0385). Patients for this study were identified via two channels. The first method of recruitment saw patients identified at clinical multi-disciplinary tumour meetings at the Walton Centre NHS foundation trust. The inclusion criteria for patients in this study was that they presented with a suspected glioma and were yet to undergo any treatment or 3T MR imaging. The second method used the vetting stage of the imaging process, whereby radiographers were able to identify qualifying patients for the study. All patients had provided informed consent and the use of the acquired spectroscopic data for the study was approved by the Walton Centre NHS foundation trust and sponsored by the University of Liverpool. A total of 6 patients were included in this study. Table 5.1 shows the IDH status and pathology for each of the patients.

5.2.2 *In-vivo* Magnetic Resonance Spectroscopy Acquisition

We used a semi-LASER sequence as described by Berrington et al[186]. The sequence was kindly shared with us by our collaborator Uzay Emir from the fMRIB group at Oxford University. The sequence uses an optimised TE of 110ms. Patients were imaged on a 3T whole body MR Skyra/Trio Siemens system and was implemented as described by Berrington et al[186]. Spectra were acquired both with and without VAPOR water suppression and outer volume saturation bands at a TR of 2000ms. A voxel of 20x20x20 mm^3 were positioned in the centre of the tumour. 64 averages were acquired using a 16 channel head coil. This cumulates in an acquisition time of 2mins 15seconds. For each patient the water flip angle was optimised by acquiring incremental scans whereby the flip angle was in increased by 2-10° for each acquisition with a bandwidth of 80Hz. and the angle that resulted in the optimal water

suppression was selected. An unsuppressed water scan was also acquired for absolute quantification purposes. Spectra were anonymised and files were exported in the .rda format.

Patient Number	Pathology	IDH status
1	Oligodendroglioma (III)	WT
2	GBM (IV)	Mut
3	GBM (IV)	WT
4	GBM (IV)	Mut
5	Astrocytoma (III)	WT
6	GBM (IV)	Mut

Table 5.1 The IDH status and pathological diagnosis of each of the patients included in this study. Roman numerals in bracket denote the WHO grade of each of the tumours at time of scanning.

5.2.3 Analysis

Basis sets were provided by our collaborator, and the construction of these is described by Berrington et al[186]. Metabolite assignment was carried out using LCModel[187] and the basis sets comprised of the metabolites 2-HG, alanine (Ala), ascorbate (Asc), aspartate (Asp), Gamma aminobutyric acid (GABA), total creatine (Cr+PCr:tCr) glutamine (Gln), glutamate (Glu), glycine. (Gly), lactate (Lac), N-acetyl aspartate (NAA), phosphocholine (PCho), phosphatidylethanolamine (PE), scyllo-inositol (scyllo-Ins), taurine (Tau), glucose (Glc), and glutathione (GSH). The values in the %SD column of each of the LCModel outputs represent the estimated standard deviations or Cramér-Rao lower bounds (CRLB) which are expressed as a percentage of the estimated concentrations. Multiplication of these lower bounds by 2 gives the 95% confidence interval. A CRLB >50 suggests that the concentration of the metabolite may range from zero to double the estimated concentration, and it can therefore be determined that the metabolite is not reliably detectable within this spectrum. A CRLB value of ≈20% suggests that only changes of approximately 40% that can be detected with reliability and a CRLB <20% may be considered as a rough criterion for estimates with acceptable reliability.

5.3 Results

We set out to assess the utility of the semi-LASER sequence with a TE of 110ms, developed and described by Berrington et al[186] for non-invasive detection of 2-HG in a clinical setting. Figure 5.6 shows the processed spectrum of patient 1 following analysis using LCModel[187]. Table 5.2 shows the concentration table displaying the calculated concentrations of metabolites of interest from the basis set, along with the CRLB (%SD). The plot shows the real part of the frequency domain data plotted as a thin black curve, whilst the thick red curve that overlies this is the LCModel fit to this data. At the top of the plot you can see the *residuals*, which equates to the actual data, minus the LCModel fit to the data, and provide a useful and sensitive diagnostic to the analysis. A residual plot that is randomly scattered about zero suggests a good fit within experimental error and no apparent systematic errors. The plot seen in figure 5.6 shows a poorly fitted spectrum, with a residual comprising of systematic peaks with high amplitude. This diagnostic metric suggests the data contained systematic errors. Further to this, it is not possible to distinguish resolved peaks of highly abundant metabolites that we would normally expect to see, such as NAA at 2.0 or the creatine peak at 3.0 ppm. When we consider the concentration table on the left, only 3 out of the 37 metabolites within the basis set had a CRLB <20% (highlighted in blue), and only 5 out of the 37 metabolites had an CRLB≈20% (up to 22%). This suggests that it is not possible to reliably detect these metabolites from this spectrum. This spectrum was acquired with only 8 averages, as shown in the printed acquisition parameters above the plot. Further to this, the tumour was located in the temporal pole of the temporal lobe, adjacent to the bone of the middle cranial fossa in this patient. This location resulted in poor shimming, and the proximity to the bone is likely to have introduced susceptibility artefact into the data.

Figure 5.7 shows the spectrum acquired from patient 2. Similar to the spectrum acquired from patient 1, this spectrum is also poorly fitted with a large residual and no identifiable peaks upon initial inspection. This patient was identified as IDH Mut following surgical resection, and although the concentration table (table 5.2) suggests that 2-HG was identifiable from these data, the CRLB is listed at 294%, indicating poor reliability for detection. This tumour was located in the superior part of the motor cortex, and was located much further away from skull or meningeal tissue when compared to the tumour seen in patient 1, was much less likely subject to susceptibility artefact within the data. Despite this, the spectrum is
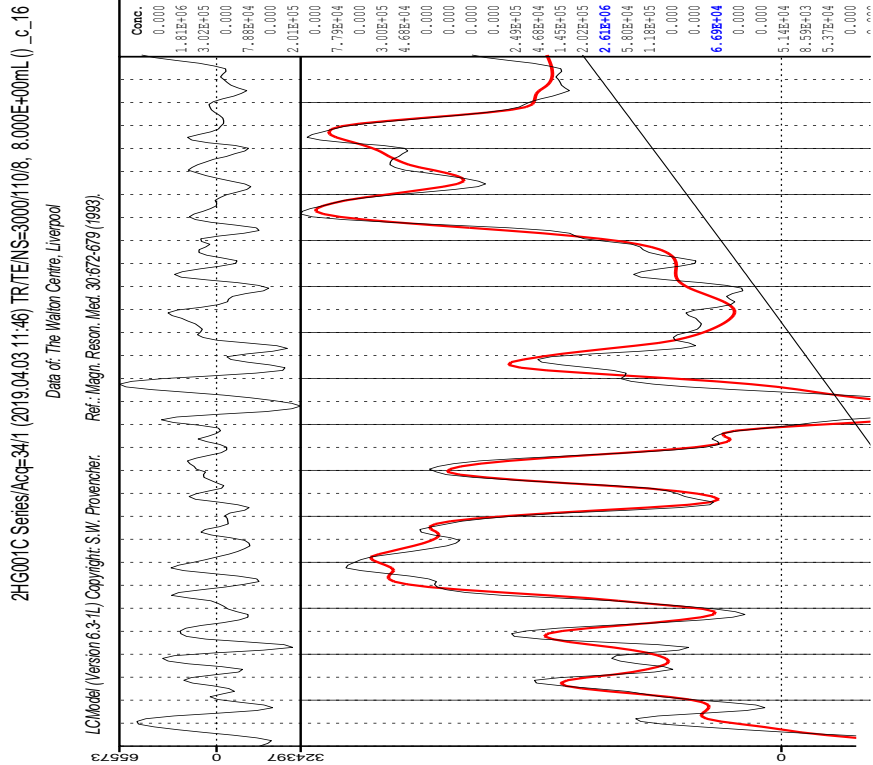
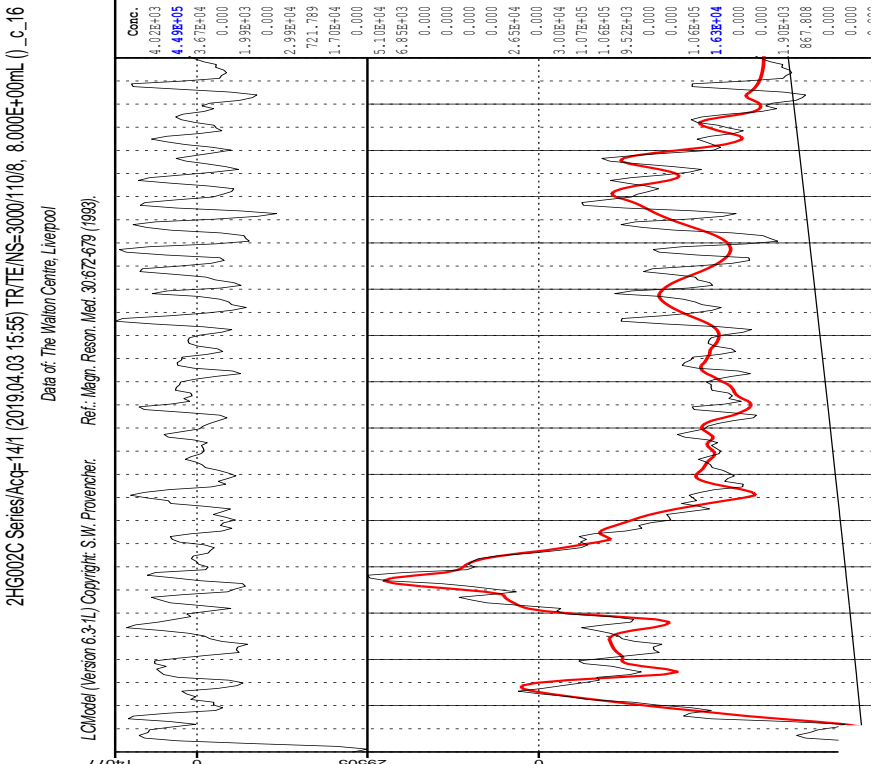**Fig 5.6 (a)** LCModel processed spectrum of patient 1. A large residual, very poor baseline and high SD values contribute to the poor quality of these data. It was not possible to resolve any potential 2-HG peaks. **(b)** The processed spectrum from patient 2. There is a large amount of residual in the plot, however 2-HG was correctly identified but the CRLB suggests it is not a reliable calculation from this fit.

Metabolite	Patient 1		Patient 2		Patient 3		Patient 4		Patient 5		Patient 6	
	CRLB (%SD)	Conc.	CRLB (%SD)	Conc.	CRLB (%SD)	Conc.	CRLB (%SD)	Conc.	CRLB (%SD)	Conc.	CRLB (%SD)	Conc.
2-HG	999%	0	294%	4.02E+03	999%	0	22%	1.16E+05	67%	2.991	**20%**	2.431
Ala	38%	3.02E+05	40%	3.67E+04	213%	6.53E+03	42%	3.04E+04	107%	1.531	999%	0
GABA	999%	0	206%	721.789	142%	3.06E+03	155%	2.89	129%	0.352	57%	0.141
Cr	21%	2.01E+05	**16%**	2.99E+04	**12%**	1.25E+05	**8%**	1.08E+05	12%	6.553	**6%**	2.603
Gln	57%	7.79E+04	65%	1.7E+04	99%	3.27E+04	59%	3.26E+04	101%	1.767	**18%**	2.069
Glu	999%	0	999%	0	106%	3.14E+04	**14%**	1.55E+05	36%	4.725	999%	0
Myo-Ins	324%	4.68E+05	999%	0	28%	1.35E+05	52%	6.96E+04	71%	2.726	**17%**	3.823
Lac	67%	1.45E+05	30%	3E+04	24%	2.19E+05	24%	4.4E+04	39%	2.872	**17%**	0.725
NAA	999%	0	999%	0	37%	3.15E+04	**7%**	1.08E+05	9%	7.035	999%	0
PCho	999%	0	121%	9.52E+03	999%	0	122%	3.29E+04	999%	0	80%	0.631
tCr	**11%**	3.2E+05	**16%**	2.99E+04	**12%**	1.25E+05	**7%**	1.34E+05	9%	8.673	**6%**	2.77

Table 5.2 Shows the CRLB (%SD) and calculated concentration values for metabolites of interest calculated by LCModel fitting. Acceptable CRLB (<20%) in boldface.

of poor quality, and this may be attributed to the low number of averages as only 8 were acquired. However the shimming routine was not optimised, and these data may be considered as low quality based on a number of diagnostic metrics including large systematic residual, poor baseline, high CRLB values for the vast majority of metabolites and the inability to visually identify clearly resolved peaks that we would expect to see in a spectrum of a human glioma such as NAA, lactate, choline and creatine. During the acquisition of both datasets from patient 1 and 2, we struggled to correctly implement the procedure for the calibration of the optimum flip angle for water suppression. In turn these data are likely to contain residual water signal, and the values given in the concentration table may not be considered as absolute concentrations.

Figure 5.8 shows the spectrum acquired from patient 3. Upon first inspection there are a number of clearly distinguishable peaks at 3.2 (choline), 3.0 (creatine), 2.0 (NAA) and a poorly phased peak at 1.3 (lactate). This tumour was diagnosed as IDH WT by the pathologist following resection. When we consider the residual plot at the top of the figure, there are a number of systematic peaks of high amplitude throughout the plot suggesting a number of experimental errors, which may be attributed to the low number of averages. Similar to the previous two

patients these data were acquired with only 8 averages, which has led to the presence of a
large amount of noise within the data, and a poor fit. When we consider the concentration
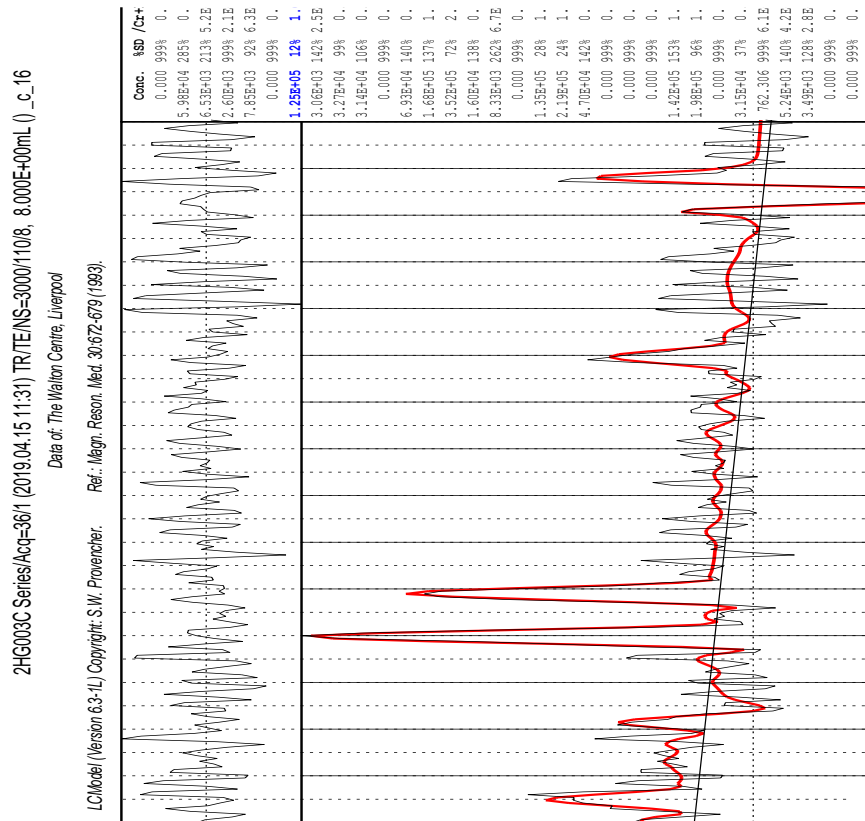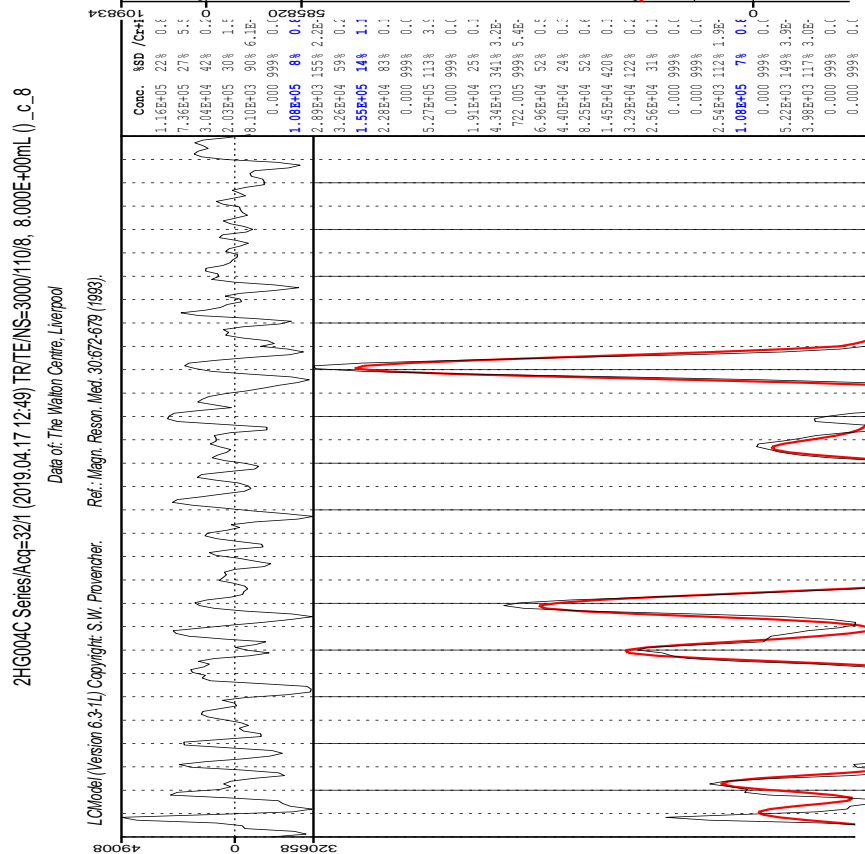

Fig 5.7 (a) The processed spectrum from patient 3, where no 2-HG was observed and its WT status was
confirmed by a pathologist. Although it was possible to resolve some peaks on this spectrum (Cre, NAA

etc.) the residual is large suggesting poor fit. **(b)** Processed spectrum from patient 4, improved baseline, it was possible to detect 2-HG (table 5.2)

table (table 5.2), there are only 2 metabolites from within the basis set shown to have a CRLB <20%. This large amount of variation within the data may again be attributed to the low averages and high level of noise. There was, however, no 2-HG detected, and although this spectrum is far from reliable in order to form a firm diagnostic decision, this result does correlate with the histopathological findings. During the acquisition of these data we were able to identify the correct brain shimming routine, however we did not have a clear value for the water suppression flip angle following the calibration protocol, which may have given rise to latent water signal within the data and poor calculation of absolute concentrations of each metabolite in LCModel.

The processed spectrum from patient 4 can be seen in figure 5.7. Initial examination of the plot shows a number of clearly identifiable peaks within the spectrum such as choline (3.2), creatine (3.0), glutamate/glutamine/GABA (2.2-2.4) and NAA (2.0). There is also an inverted lactate peak, however this has a much smaller amplitude when compared to patient 3, which is likely due to the pathological diagnosis as patient 3 was a grade IV GBM, whereas patient 4 was diagnosed as a grade III astrocytoma. We would normally expect to see increased lactate in the high grade tumours due to increased extracellular acidification. When we consider the concentration table (table 5.2) 2-HG has been identified as present within the spectrum, and the CRLB was calculated at 22%, which suggests that the metabolite peak was present in the spectrum. The identification of 2-HG and therefore the suggestion that this patient is IDH Mut is conducive with the pathological diagnosis as this patient was shown to be carrying the mutation following histopathological investigations. The residual plot at the top of the spectrum shows a large amount of noise which is reflective of a low signal to noise ratio (SNR) and poor fit of the data, which can be attributed to only 8 averages being used to acquire this dataset. During the acquisition of these data both the water flip angle calibration and the shimming routine was conducted smoothly, and this is reflected in the linewidths. It is possible to distinguish 2 peaks at 2.2-2.4, which are contributed to by glutamate, glutamine and GABA, which reflects the increase in quality of the shimming routine when compared to previous spectra. If additional averages had been acquired, it may have been possible to resolve these metabolites. All these factors assist in determining the quality and reliability of the data. Although the detection of 2-HG from this spectrum appears to be within the boundaries of reliability according to the %SD of CRLB, all the other attributes of the dataset suggest that the spectrum is of poor quality and was difficult to fit.
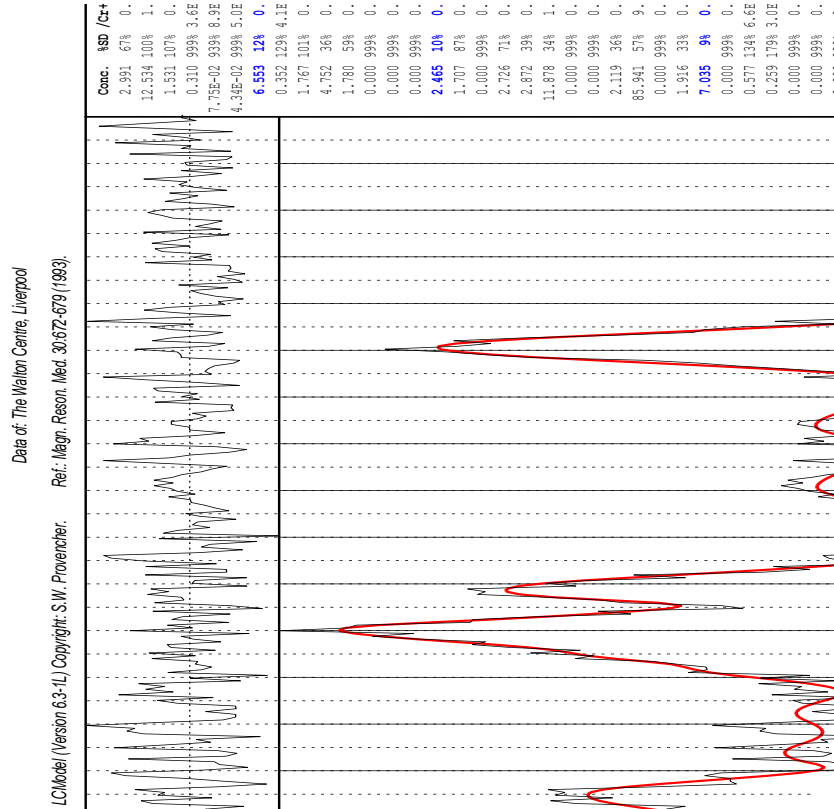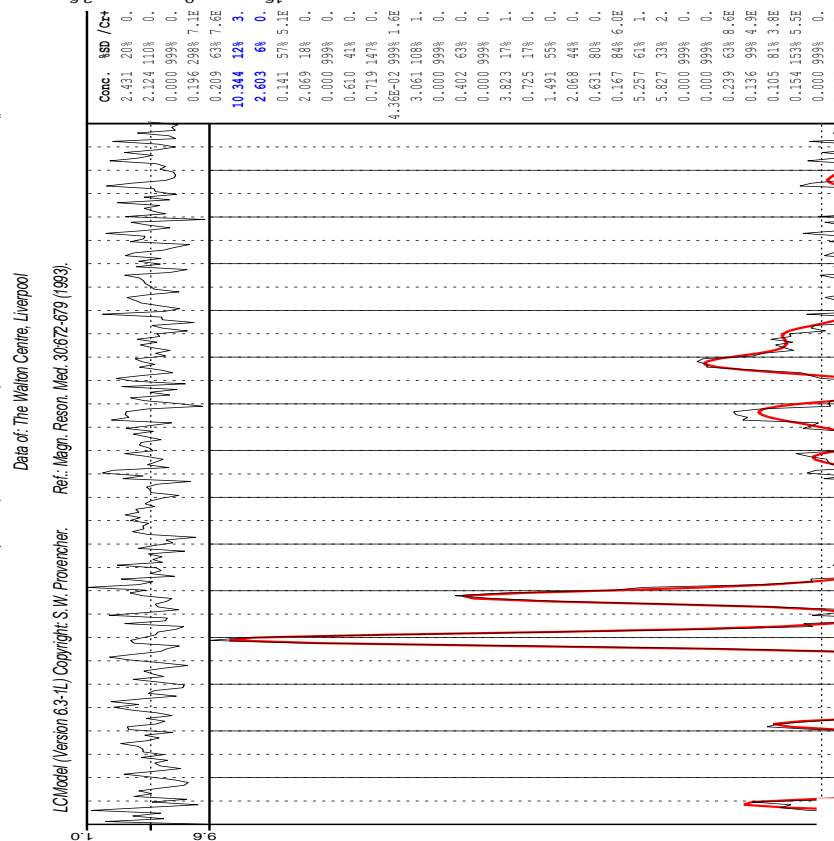
Fig 5.8 (a) Processed spectrum from patient 5. The increased number of averages for this acquisition has contributed to a much better fit of the data. Although 2-HG was shown to be present in these data this was found to be a false positive, and the broad linewidths may be attributed to bad shimming. **(b)** Processed spectrum from patient 6. Improved linewidths and residual indicate a much better fit of the data then previous patients. 2-HG was found to be present, Mut status confirmed by histopathology.

Figure 5.8 shows the processed spectrum acquired from patient 5. These data were acquired with 64 averages, and when we consider the residual plot at the top of the figure it is evident that the fit of these data was much improved when compared to the fit of the previous patients where only 8 averages were used. The residual plot contains randomly scattered peaks with low amplitude about zero. When we consider the fitted data plot, there are clearly identifiable peaks in the spectrum (choline, creatine and NAA) however the linewidth of these peaks is broad and has contributed to difficulties in resolving smaller multiplets such as glutamate/glutamine and 2-HG. The difficulties in resolution of the 2-HG peak have been reflected in %SD as shown in the table 2 for this metabolite, whereby the CRLB is 67%. This patient was determined to be IDH WT, based on histopathology, and therefore according to these data has shown to be somewhat of a false positive, however as the CRLB is >50% the metabolite may well range from zero to twice the stated value, and therefore the metabolite may be considered as undetectable within these data and the fit may have contributed to a false positive.

The LCModel output for patient 6 can be seen in figure 5.8. These data have very clearly identifiable peaks at 3.2 and 3.0 from choline and creatine respectively. Furthermore, the linewidth of each of these peaks is narrower than the previous patient indicating improved shimming. The residual plot at the top of the figure shows random peaks of much lower amplitude than previous patients, suggesting an improved fit of the data. When considering the values in table 5.2, the CRLB of 20% suggests that the concentration estimate may be considered as reliable[187]. The identification of 2-HG within this spectrum was confirmed by histopathology and the diagnosis for this patient was IDH Mut. The improved linewidths and data fit may be attributed to the improved quality of the acquisition of this spectrum, as with 64 averages and good shimming and water flip angle calibration we were able to resolve a number of peaks, including the 2-HG peak, more robustly than previous efforts in this study.

5.4 Discussion

The results that have been presented here demonstrate that the MRS sequence was able to accurately, and with confidence, identify IDH mutation by way of 2-HG detection on one occasion, however a larger sample size is required in order to ratify these findings. It has also highlighted the technical difficulties of implementing such sensitive MRS sequences in clinical

practice, and how clinical translation of novel MR advancements requires both academic and industrial institutions to work more closely in the future, in order to deliver new technologies into modern day medicine.

5.4.1 semi-LASER 110ms TE for 2-HG Detection

As the delivery of cancer care continuously evolves to incorporate a more individualised approach, there is a requirement to procure as much information about a tumours' profile at the earliest possible stages in order to deliver the most efficient and effective treatment for the patient. The non-invasive detection of IDH mutation by way of 2-HG measurement using MRS has shown promise, and since it was first proposed in 2012 by Pope et al.[19] a variety of studies have emerged that explore a number of strategies for the improved detection and localisation of the 2-HG signal. As was previously discussed, there are a range of technical limitations that exist around the detection of the oncometabolite including spectral overlap from a number of peaks produced by metabolites found in abundance within both healthy and diseased brain tissue. Here we investigated the utility of an optimised semi-LASER sequence with a TE of 110ms at 3T, proposed by Berrington et al.[186] for the detection of 2-HG *in vivo*. We chose to investigate the clinical potential of this sequence for several reasons. Firstly, the use of a broadband adiabatic refocusing pulse allows for retention of the H3/H3' peak at 1.9 ppm, a characteristic of the 2-HG spectrum that is lost when using optimised PRESS sequences at 3T. Further to the peak retention, the use of the refocusing pulse allows for a significant reduction in compartmental artefacts and facilitates a flatter baseline within the often crowded and difficult to resolve ppm range of *in-vivo* MRS data.

This study comprises of only 6 patients, 3 IDH WT and 3 IDH Mut tumours from different pathological origins. The initial intention of this investigation was to acquire a minimum of 40 good quality spectra with a test group split of approximately 20 IDH WT and 20 IDH Mut tumours. As this was the first time our lab was to form a collaboration with the Walton Centre NHS foundation trust, there were a number of hurdles to overcome before work would be approved to commence. The process of obtaining the relevant ethical permissions for both conducting the study as well as installation of a research pulse sequence on the magnet took 2 years. We were finally able to begin to install the binaries on the clinical system in October 2018. From this point we suffered further delays as the system needed to be updated to later version in order to facilitate the binaries shared with us by the developers at the FMRIB group in Oxford. Once the system had been updated to support the installation, the protocol for

patient recruitment also needed to be established and it wasn't until April 2019 when we were able to acquire the first dataset. Due to these issues we were unable to achieve the 40 patient cohort that we initially anticipated.

When considering the diagnostics that form part of the LCModel output of each of the spectra, the spectra acquired from patients 1,2,3 and 4 were of insufficient quality in order to obtain any true determination as to whether the 2-HG oncometabolite was present. This can be primarily determined by the residual plot seen at the top of figures 5.6-5.8. These residual plots suggest that the data were fitted poorly to produce the main plot as well as determination of the values within the concentration table. The poor fit of each of these spectra can likely be attributed to the presence of a large amount of noise within the data which is a consequence of the low number of averages. As this was the first time both installing and operating a custom MRS sequence on a Siemens system, we were unfamiliar with the setup process. As a result of unfamiliarity with the sequence and system, alongside complexity of acquisition set up with manual water suppression FA calibration and shimming, it was not recognised that the scans we were acquiring did not have a sufficient number of averages in order to acquire suitable data with a reasonable signal to noise ratio. Following the initial scans the developers were consulted, and they advised that the number of averages was too low and in fact needed to be at least 64, which was how we acquired the datasets for patient 5 and 6. Although the dataset for patient 5 was acquired using the appropriate number of averages, the shimming for this particular tumour was poor, due to the anatomical position at the base of the skull in the temporal lobe. Shimming has a high impact on signal to noise ratio, and this is evident if we compare both the magnitude of the residual plot as well as the linewidths of the spectrum of patient 5 with that of patient 6[52]. The shimming routine should homogenise the magnetic field of the MRS voxel so that the entire region that is being scanned is subject to the same field strength and therefore the same Larmor frequency. The Larmor equation dictates that if there is a variation of B_0 across the scanned region, there will in return be a distribution of frequencies for each of the resonance lines, causing a broadening of the NMR lines[52].

The impact of poor shimming is a reality faced in the acquisition of *in vivo* data, especially when imaging brain tumours that are located in close proximity to the skull and sinuses due to air-bone, and bone-tissue interface susceptibility. These factors are likely to have contributed to the production of the false positive peak from 2-HG seen in patient 5. This investigation set out to determine the clinical applicability of using the proposed semi-LASER

TE 110ms in order to detect the presence of 2-HG. Although the sequence has been optimised in order to circumvent spectral overlap of the 2-HG peak, the complexity in acquisition and the sensitivity of the routine to sources of variability, including poor shims due to anatomical location, raise the question as to whether this practice could be routinely and successfully carried out in a clinical setting with confidence that the result would be indicative of the true IDH status of the patient.

The spectrum acquired from patient 6 allowed for true positive identification of 2-HG with a calculated CRLB = 20% representing the most reliable detection of 2-HG[187]. Whilst this was shown to be a true positive, it was the only dataset that was of suitable quality in order to extract reliable information from. These results are therefore inadequate in determining whether or not this semi-LASER pulse sequence is able to consistently measure 2-HG in IDH Mut glioma patients within a clinical setting. The complexity surrounding the implementation of this particular pulse sequence including the water suppression flip angle calibration and shimming routine presented a number of opportunities for the introduction of experimental error during the scanning. Moreover, there are further limitations to the translatability of this technique into the clinic when we consider the practicality of radiographers carrying out the calibration and acquisition on each patient. There is a very tight restriction on time required for each scan when using clinical systems, and the relatively long iterations between shimming, FA calibration and acquisition of both water supressed and unsuppressed spectra cumulates into an MRS routine that is difficult to ubiquitously employ throughout healthcare systems. Although the sensitivity of this pulse sequence for 2-HG detection was outlined by Berrington et al.[186] there is yet to be any multicentre validation. Perhaps this has highlighted that in order to accelerate translation and overcome the disparity between academic innovation and clinical implementation, there needs to be a closer collaboration between industrial manufacturers and research institutions, in order to make such techniques more accessible. Increased accessibility and coherence with user interfaces of clinical systems could propel validation studies and facilitate the translation of cutting edge technology into the delivery of personalised healthcare.

One of the major limitations of this study, as previously discussed, was the ethical approval and technical installation of the sequence. In addition to these limiting factors, one other substantial confounder is that the 2-HG peak appears within the spectrum proximally to the NAA peak. 2-HG is most often seen in lower grade (II-III) tumours, or those higher grade

(IV) tumours that have progressed from a lower grade lesion. NAA is known to occur in high abundance within LGG[210]. The large NAA peak in LGG may mask the 2-HG peak in a number of tumours, especially those lower grade, more juvenile lesions that have not built up 2-HG concentration to high levels. This can impact the clinical translatability of these methods. This confounding factor could be overcome by implementing 2D *in vivo* spectroscopy. This would allow decoupling of the 2-HG signal, and clear differentiation of the peak, as has been demonstrated by Verma et al[195]. Although 2D *in vivo* spectroscopy may present as a clear alternative for overcoming the limitations of 1D MRS, the long acquisition times, such as the 17 minutes required for the 2D L-COSY MRS described by Verma et al, are unfeasible within a clinical setting. The reason for this is due to the requirement to reduce the amount of time a patient spends in the scanner. Although MRS can provide information regarding the metabolic phenotype of the tumour, the diagnostic information is limited, as it does not contain any spatial data. Typically, the longest time a clinic will allow any patient to be in a scanner is somewhere between 45-60mins. Using 30-40% of the imaging time on spectroscopy alone will sacrifice the acquisition of other diagnostic information. Clinics often aim to cap MRS acquisition at somewhere around 5minutes, or in certain cases 10 minutes maximum, in order to allow for other sequences to be implemented. One potential method for reducing these long acquisition times would be to decrease the number of averages. If one decreases the number of averages, then the signal will also decrease, and in the case of LGG where 2-HG concentration may not be so high, the important peak may be lost within the spectrum.

The acquisition time for the scans performed in this study was 3 min, and therefore one way in which the data collection may be improved is to increase the number of averages per scan. Here, the scans carried out on patients 5 and 6 were acquired with 64 averages, this may be doubled to 128 averages, increasing the acquisition time to 6 minutes and increasing the signal, thereby facilitating better separation of the 2-HG peak from NAA. This amount of acquisition time would also be clinically feasible, and may facilitate better translation of the sequence into diagnostic practice.

5.5 Concluding Remarks

Here we aimed to investigate the ability of the proposed semi-LASER sequence in the detection of 2-HG in IDH Mut glioma patients in a clinical setting. A combination of technical and administrative hurdles contributed to the late commencement of the study, and time constraints resulted in the acquisition of only a single reliable spectrum. Although we have been unable to determine the utility of this sequence in non-invasive detection of IDH mutation, this work may be considered as a foundation upon which further validation investigations may be conducted. If the proposed sequence is shown to be reliable at consistently identifying IDH Mut patients by way of MRS, this would be an invaluable tool in the diagnosis and monitoring of glioma pathologies.

6. Summary

Tumours of the central nervous system represent approximately 2% of all identified primary tumours in the UK, whereby each year roughly 12,000 people are newly diagnosed and around 5,500 of those are malignant.[211] Although these tumours only account for a small percentage of newly diagnosed malignancies, the inherent infiltrative and aggressive nature by which they propagate along and between vital tissue within the brain makes treatment of the disease difficult, and mortality rates high. Only around 30% of adults diagnosed with a primary glioma survive past 1 year post-diagnosis, and only 13% survive more than 5 years[211]. The median survival for patients with anaplastic astrocytoma is approximately 2-3 years, whereas the median survival for GBM patients is around 1 year post-diagnosis[211]. The importance of understanding the molecular and genetic composition of a tumour was highlighted by the incorporation of a number of molecular markers into the WHO diagnostic criteria in 2016. Amongst others, mutations in the IDH1/2 gene were highlighted as integral for the correct diagnosis and prognostication of gliomas.

It is thought that in those tumours carrying an IDH mutation, this genetic aberration is one of the first events to occur in the development of a primary glioma. The subsequent production and accumulation of 2-HG has been suggested to drive the cellular phenotype towards malignancy, potentially aiding in the facilitation of a number of molecular events. Despite this, interestingly those patients who harbour the mutation have increased survival and better prognosis. The current standard practice for determining IDH mutational status, where immediate resection is not necessary, is to obtain a biopsy *via* an invasive procedure, and conduct a combination of immunohistochemical and sequencing investigations. Clearly there is an unmet need in the non-invasive procurement of molecular diagnostic information, as well as understanding further the downstream effects of IDH mutation on the development and progression of the disease. In this thesis we have presented a number of investigations that have been aimed at using various magnetic resonance methods to non-invasively probe for IDH mutation, as well as understanding the cellular metabolic consequences.

Radiomics is an emerging and rapidly evolving technology, and the utility of it is being investigated in a wide range of applications and modalities, including computed tomography (CT), positron emitting tomography (PET) and MRI[101,212–214]. Although the utility of the technology has been demonstrated in a number of proof-of-concept studies, the translation of this technique into clinical settings has been hampered by issues surrounding the reliability and replicability. In an attempt to address this the IBSI published guidelines for the correct

conduct of radiomics studies[81]. In this thesis we explored the application of both genetic algorithms and random forests in combination with multiple linear models for variable selection and classifier construction. Although we were unable to build a robust model for classification of IDH tumour based on T_1-post contrast image texture features, our investigations highlighted the importance of image quality and consistency across data acquisition when carrying out radiomics studies. Further to this, it has also raised the question of clinical applicability when considering radiomics models. All the data used here were collected within the same centre by radiographers during routine clinical examinations. Despite this we found that the large number of variables that were introduced during the acquisition presented bias within the data and impeded on the ability to single out discriminatory features that would facilitate the classification of IDH Mut tumours. This is a factor to consider when contemplating the scaling up of any particular model that may have been shown to allow classification of images during a study. Every imaging department in every hospital operates differently, with a range of different hardware and software. If these factors are likely to affect the ability of a model to classify images then this must be taken into account in order to enable clinical translation. In addition to the variability within the data used in the construction of radiomics models, the processing of images, generation of texture features and final classification is a long and convoluted process that requires a large amount of user input. Within the busy schedules of radiologists, implementation of these techniques *via* the currently available platforms is arduous and unachievable. If a model is found to be robust across multiple centres with varying acquisitions, then the process of radiomics classification must be streamlined in order to be considered by healthcare professionals as a viable diagnostic tool. This being said, radiomics is showing a large amount of promise, and the use of artificial intelligence in order to extract as much information from images is an interesting and potentially invaluable tool. The technology is still in its infancy, and it may be the case that with continued development, aided by international collaboration, will see the beginning of clinical translation in years to come.

Here we have presented NMR metabolomics in order to assess the metabolic phenotype of IDH mutant tumours. Although the importance of the mutation has been acknowledged for a number of years, it is not yet clear how its presence and subsequent accumulation of 2-HG affects the process of gliomagenesis and disease progression. By employing a combination of 1D and 2D NMR, in conjunction with both univariate and

multivariate analyses, we have been able to distinguish a potential metabolic phenotype of IDH Mut tumours. We identified a number of metabolites that appear to be consistently changed within the Mut tumours, which may aid in further investigations to understand the cellular impact of the mutation. Cellular metabolism is an inherently complex matrix of interlocking pathways, each one having both up- and downstream effects on an array of other intracellular processes. Changes in metabolism can act as driving forces in the evolution and advancement of disease, and by observing and quantifying these changes we may begin to understand the impact that IDH mutation has. These findings may be of particular importance when considering future drug development, especially those aimed at IDH inhibition[215]. Here the combined utility of both 1D and 2D NMR, as well as the use of quantitative 2D NMR for metabolomics has been demonstrated and it is hoped that this will continue to grow, as it has been shown here that incorporation of this technique into metabolomics studies can add a great deal of value.

The most obvious method for non-invasively probing for the mutation is to use *in-vivo* MRS in order to detect the 2-HG resonance. As has been previously outlined, there are a large number of technical difficulties that hamper the robust measurement of the metabolite. Here we were able to collect one dataset that was able to correctly identify the presence of the metabolite and therefore the mutation, it may be considered that the current convoluted acquisition, and subsequent specialist analysis remains as a barrier for clinical implementation. With continued development, and potential industrial collaboration, this method of detection will begin to make its way into hospitals worldwide, however there needs to be a considerable collaborative effort between scanner manufacturers and academic research institutes in order to facilitate this.

Taken together the work presented here demonstrates the versatility of magnetic resonance for both the potential non-invasive diagnosis, as well as unwinding the mysteries of tumour biology of IDH Mut tumours. With continued development and collaboration, incorporating multi-centre studies and cross validation, magnetic resonance can play an integral role in guiding the future of personalised treatment of IDH Mut gliomas.

References

1. Louis, D. N. *et al.* The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. *Acta Neuropathol.* **131**, 803–820 (2016).
2. Ostrom, Q. T. *et al.* The epidemiology of glioma in adults: A state of the science review. *Neuro. Oncol.* **16**, 896–913 (2014).
3. Weller, M. *et al.* Weller et al. - 2015 - Glioma. (2015). doi:10.1038/nrdp.2015.17
4. Masui, K., Mischel, P. S. & Reifenberger, G. Molecular classification of gliomas. in *Handbook of Clinical Neurology* (2016). doi:10.1016/B978-0-12-802997-8.00006-2
5. Caul, S. & Broggio, J. Cancer registration statistics, England 2017. *Off. Natl. Stat.* 1–16 (2019).
6. Killela, P. J. *et al.* Mutations in IDH1, IDH2, and in the TERT promoter define clinically distinct subgroups of adult malignant gliomas. *Oncotarget* **5**, 1515–1525 (2014).
7. Yan, H. *et al.* Mutations in Gliomas. *N. Engl. J. Med.* **360**, 765–773 (2009).
8. Balss, J. *et al.* Analysis of the IDH1 codon 132 mutation in brain tumors. *Acta Neuropathol.* **116**, 597–602 (2008).
9. Watanabe, T., Vital, A., Nobusawa, S., Kleihues, P. & Ohgaki, H. Selective acquisition of IDH1 R132C mutations in astrocytomas associated with Li-Fraumeni syndrome. *Acta Neuropathol.* **117**, 653–656 (2009).
10. Ohgaki, H. *et al.* Genetic pathways to glioblastoma: A population-based study. *Cancer Res.* **64**, 6892–6899 (2004).
11. Grimm, S. A. & Chamberlain, M. C. CNS Oncology Anaplastic astrocytoma. *CNS Oncol.* **5**, 145–157 (2016).
12. Furnari, F. B. *et al.* Malignant astrocytic glioma: Genetics, biology, and paths to treatment. *Genes Dev.* **21**, 2683–2710 (2007).
13. Sturm, D. *et al.* Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* (2012). doi:10.1016/j.ccr.2012.08.024
14. Waitkus, M. S., Diplas, B. H. & Yan, H. Isocitrate dehydrogenase mutations in gliomas. *Neuro. Oncol.* **18**, 16–26 (2016).
15. Dang, L. *et al.* Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–44 (2009).
16. Choi, C. *et al.* 2-hydroxyglutarate detection by magnetic resonance spectroscopy in IDH-mutated patients with gliomas. *Nat. Med.* **18**, 624–9 (2012).
17. Andronesi, O. C. *et al.* Detection of 2-Hydroxyglutarate in IDH-Mutated Glioma Patients by In Vivo Spectral-Editing and 2D Correlation Magnetic Resonance Spectroscopy. *Sci. Transl. Med.* **4**, 116ra4-116ra4 (2012).
18. Emir, U. E. *et al.* Noninvasive quantification of 2-hydroxyglutarate in human gliomas with IDH1 and IDH2 mutations. *Cancer Res.* **76**, 43–49 (2016).
19. Pope, W. B. *et al.* Non-invasive detection of 2-hydroxyglutarate and other metabolites in IDH1 mutant glioma patients using magnetic resonance spectroscopy. *J. Neurooncol.* **107**, 197–205 (2012).
20. Turcan, S. *et al.* Efficient induction of differentiation and growth inhibition in IDH1 mutant glioma cells by the DNMT Inhibitor Decitabine. *Oncotarget* **4**, 1729–1736 (2013).
21. Schumacher, T. *et al.* A vaccine targeting mutant IDH1 induces antitumour immunity. *Nature* **512**, 324–327 (2014).

22.   Pellegatta, S. *et al.* Effective immuno-targeting of the IDH1 mutation R132H in a murine model of intracranial glioma. *Acta Neuropathol. Commun.* **3**, 4 (2015).

23.   McLendon, R. *et al.* Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).

24.   Parsons, D. W. *et al.* An Integrated Genomic Analysis of. *Science (80-.).* **1807**, 1807–1813 (2010).

25.   Leather, T., Jenkinson, M. D., Das, K. & Poptani, H. Magnetic resonance spectroscopy for detection of 2-hydroxyglutarate as a biomarker for IDH mutation in gliomas. *Metabolites* **7**, (2017).

26.   Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of ??-ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).

27.   Guerra, C. A. *et al.* Reports 30. *Science (80-.).* **324**, 261–265 (2009).

28.   Noushmehr, H. *et al.* Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma. *Cancer Cell* **17**, 510–522 (2010).

29.   Lu, C., Ward, P., Kapoor, G. & Rohle, D. IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* **483**, 474–478 (2012).

30.   Yang, H., Ye, D., Guan, K. L. & Xiong, Y. IDH1 and IDH2 mutations in tumorigenesis: Mechanistic insights and clinical perspectives. *Clin. Cancer Res.* **18**, 5562–5571 (2012).

31.   Flavahan, W. A. *et al.* Insulator dysfunction and oncogene activation in IDH mutant gliomas. *Nature* **529**, 110–114 (2016).

32.   Koivunen, P. *et al.* Transformation by the (R)-enantiomer of 2-hydroxyglutarate linked to EGLN activation. *Nature* **483**, 484–8 (2012).

33.   Molenaar, R. J., Radivoyevitch, T., Maciejewski, J. P., van Noorden, C. J. F. & Bleeker, F. E. The driver and passenger effects of isocitrate dehydrogenase 1 and 2 mutations in oncogenesis and survival prolongation. *Biochim. Biophys. Acta - Rev. Cancer* **1846**, 326–341 (2014).

34.   Chaumeil, M. M. *et al.* Non-invasive in vivo assessment of IDH1 mutational status in glioma. *Nat. Commun.* **4**, 2429 (2013).

35.   Popovici-Muller, J. *et al.* Discovery of the first potent inhibitors of mutant IDH1 that lower tumor 2-HG in vivo. *ACS Med. Chem. Lett.* **3**, 850–855 (2012).

36.   Bittinger, M. a *et al.* An Inhibitor of Mutant IDH1 Delays. *Science (80-.).* **340**, 626–630 (2013).

37.   Viswanath, P., Chaumeil, M. M. & Ronen, S. M. Molecular Imaging of Metabolic Reprograming in Mutant IDH Cells. *Front. Oncol.* **6**, 60 (2016).

38.   Serkova, N. J. & Brown, M. S. Quantitative analysis in magnetic resonance spectroscopy: from metabolic profiling to in vivo biomarkers. *Bioanalysis* **4**, 321–41 (2012).

39.   Gokcen, C., Isikay, S. & Yilmaz, K. L-2 Hydroxyglutaric aciduria presenting with anxiety symptoms. *Case Reports* **2013**, bcr2013009512–bcr2013009512 (2013).

40.   Xia, L. *et al.* Prognostic role of IDH mutations in gliomas: a meta-analysis of 55 observational studies. *Oncotarget* **6**, 17354–65 (2015).

41.   Kleihues, P. & Ohgaki, H. Primary and secondary glioblastomas: from concept to clinical diagnosis. *Neuro. Oncol.* **1**, 44–51 (1999).

42.   Ohgaki, H. & Kleihues, P. Population-based studies on incidence, survival rates, and genetic alterations in astrocytic and oligodendroglial gliomas. *J. Neuropathol. Exp. Neurol.* **64**, 479–489 (2005).

43.   Leu, S., von Felten, S., Frank, S., Boulay, J.-L. & Mariani, L. IDH mutation is associated

with higher risk of malignant transformation in low-grade glioma. *J. Neurooncol.* **127**, 363–72 (2016).

44. Juratli, T. A. *et al.* The prognostic value of IDH mutations and MGMT promoter status in secondary high-grade gliomas. *J. Neurooncol.* **110**, 325–333 (2012).

45. Kim, Y. H. *et al.* Molecular classification of low-grade diffuse gliomas. *Am. J. Pathol.* **177**, 2708–2714 (2010).

46. Sanson, M. *et al.* Isocitrate dehydrogenase 1 codon 132 mutation is an important prognostic biomarker in gliomas. *J. Clin. Oncol.* **27**, 4150–4154 (2009).

47. Juratli, T. A. *et al.* a Biomarker for Malignant Progression. **15**, 682–690 (2013).

48. Van Den Bent, M. J. *et al.* IDH1 and IDH2 mutations are prognostic but not predictive for outcome in anaplastic oligodendroglial tumors: A report of the European Organization for Research and Treatment of Cancer Brain Tumor Group. *Clin. Cancer Res.* **16**, 1597–1604 (2010).

49. Gliomas, L. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).

50. Combs, S. E. *et al.* Prognostic significance of IDH-1 and MGMT in patients with glioblastoma: One step forward, and one step back? *Radiat. Oncol.* **6**, 115 (2011).

51. Millward, C. P. *et al.* The impact of MGMT methylation and IDH-1 mutation on long-term outcome for glioblastoma treated with chemoradiotherapy. *Acta Neurochir. (Wien).* **158**, 1943–1953 (2016).

52. Keeler, J. Understanding NMR Spectroscopy. *John Wiley Sons* 409–416 (2010). doi:10.1007/SpringerReference_67582

53. Spin, P. & Experiment, S. Particle Spin and the Stern-Gerlach Experiment. *Lect. Notes, Pysics 301* (2009).

54. Blink, J. E. mri Physics. *Foot Ankle Spec.* **11**, 242–245 (2004).

55. Holger Försterling, F. Spin dynamics: Basics of Nuclear Magnetic Resonance, Second Edition . *Med. Phys.* **37**, 406–407 (2009).

56. Canavagh, J., Fairbrother, W. J., Palmer II, A. G., Rance, M. & Skelton, N. J. Theoretical Description of NMR Spectroscopy. *Protein NMR Spectrosc. 2nd Ed.* (2007).

57. Mansfield, P. Multi-planar image formation using NMR spin echoes. *J. Phys. C Solid State Phys.* **10**, (1977).

58. LAUTERBUR, P. C. Image Formation by Induced Local Interactions: Examples Employing Nuclear Magnetic Resonance. *Nature* **242**, 190–191 (1973).

59. Sayed, S., Nassef, M., Badr, A. & Farag, I. A Nested Genetic Algorithm for feature selection in high-dimensional cancer Microarray datasets. *Expert Syst. Appl.* **121**, 233–243 (2019).

60. Atkinson, A. J. *et al.* Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin. Pharmacol. Ther.* **69**, 89–95 (2001).

61. Lambin, P. *et al.* Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446 (2012).

62. Maley, C. C. *et al.* Genetic clonal diversity predicts progression to esophageal adenocarcinoma. *Nat. Genet.* **38**, 468–473 (2006).

63. Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: A looking glass for cancer? *Nat. Rev. Cancer* **12**, 323–334 (2012).

64. Chicklore, S. *et al.* Quantifying tumour heterogeneity in 18F-FDG PET/CT imaging by texture analysis. *Eur. J. Nucl. Med. Mol. Imaging* **40**, 133–140 (2013).

65. Fisher, R., Pusztai, L. & Swanton, C. Cancer heterogeneity: Implications for targeted

therapeutics. *Br. J. Cancer* **108**, 479–485 (2013).

66. Endesfelder, D. *et al.* new england journal. (2012).

67. Henriksson, E. *et al.* 2-Deoxy-2-[18F]fluoro-D-glucose uptake and correlation to intratumoral heterogeneity. *Anticancer Res.* **27**, 2155–2159 (2007).

68. Basu, S. *et al.* Evolving role of molecular imaging with PET in detecting and characterizing heterogeneity of cancer tissue at the primary and metastatic sites, a plausible explanation for failed attempts to cure malignant disorders. *Eur. J. Nucl. Med. Mol. Imaging* **38**, 987–991 (2011).

69. Knopp, M. V. & Yang, X. Quantifying tumor vascular heterogeneity with dynamic contrast-enhanced magnetic resonance imaging: A review. *J. Biomed. Biotechnol.* **2011**, (2011).

70. Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577 (2016).

71. Gillies, R. J., Anderson, A. R., Gatenby, R. A. & Morse, D. L. The biology underlying molecular imaging in oncology: from genome to anatome and back again. *Clin. Radiol.* **65**, 517–521 (2010).

72. Tixier, F. *et al.* Intratumor Heterogeneity Characterized by Textural Features on Baseline 18F-FDG PET Images Predicts Response to Concomitant Radiochemotherapy in Esophageal Cancer. *J. Nucl. Med.* **52**, 369–378 (2011).

73. Vaidya, M. *et al.* Combined PET/CT image characteristics for radiotherapy tumor response in lung cancer. *Radiother. Oncol.* **102**, 239–245 (2012).

74. Guggenbuhl, P. *et al.* Reproducibility of CT-based bone texture parameters of cancellous calf bone samples: Influence of slice thickness. *Eur. J. Radiol.* **67**, 514–520 (2008).

75. Galavis, P. E., Hollensen, C., Jallow, N., Paliwal, B. & Jeraj, R. Variability of textural features in FDG PET images due to different acquisition modes and reconstruction parameters. *Acta Oncol. (Madr).* **49**, 1012–1016 (2010).

76. Mayerhoefer, M. E., Szomolanyi, P., Jirak, D., Materka, A. & Trattnig, S. Effects of MRI acquisition parameter variations and protocol heterogeneity on the results of texture analysis and pattern discrimination: An application-oriented study. *Med. Phys.* **36**, 1236–1243 (2009).

77. Raja, R. *et al.* Assessment of tissue heterogeneity using diffusion tensor and diffusion kurtosis imaging for grading gliomas. *Neuroradiology* **58**, 1217–1231 (2016).

78. Miles, K. A., Ganeshan, B. & Hayball, M. P. CT texture analysis using the filtration-histogram method: What do the measurements mean? *Cancer Imaging* **13**, 400–406 (2013).

79. Soni, X. N., Priya, S. & Bathla, X. G. Texture analysis in cerebral gliomas: A review of the literature. *Am. J. Neuroradiol.* **40**, 928–934 (2019).

80. Hainc, N., Stippich, C., Stieltjes, B., Leu, S. & Bink, A. Experimental texture analysis in glioblastoma: A methodological study. *Invest. Radiol.* **52**, 367–373 (2017).

81. Zwanenburg, A., Leger, S., Vallières, M., Löck, S. & Initiative, for the I. B. S. Image biomarker standardisation initiative. (2016). doi:10.17195/candat.2016.08.1

82. Chatfield, C. *Statistical Analysis and Data Display*. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **168**, (2005).

83. Mayerhoefer, M. E. *et al.* Effects of magnetic resonance image interpolation on the results of texture-based pattern classification a phantom study. *Invest. Radiol.* **44**, 405–411 (2009).

84. Galloway, M. M. Texture analysis using gray level run lengths. *Comput. Graph. Image Process.* **4**, 172–179 (1975).

85. Rizzo, S. *et al.* Radiomics: the facts and the challenges of image analysis. *Eur. Radiol. Exp.* **2**, (2018).

86. Dinstein, I., Shanmugam, K. & Haralick, R. M. Textural Features for Image Classification. *IEEE Trans. Syst. Man. Cybern.* **SMC-3**, 610–621 (1973).

87. Avants, B. B., Tustison, N. & Song, G. Advanced Normalization Tools (ANTS). *Insight J.* 1–35 (2009). doi:http://hdl.handle.net/10380/3113

88. Pieper, S., Lorensen, B., Schroeder, W. & Kikinis, R. The NA-MIC Kit: ITK, VTK, pipelines, grids and 3D slicer as an open platform for the medical image computing community. *2006 3rd IEEE Int. Symp. Biomed. Imaging From Nano to Macro - Proc.* **2006**, 698–701 (2006).

89. Van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).

90. Hou, Z. A review on MR image intensity inhomogeneity correction. *Int. J. Biomed. Imaging* **2006**, 1–11 (2006).

91. Vovk, U., Pernuš, F. & Likar, B. A review of methods for correction of intensity inhomogeneity in MRI. *IEEE Trans. Med. Imaging* **26**, 405–421 (2007).

92. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).

93. Bakker, R., Tiesinga, P. & Kötter, R. The Scalable Brain Atlas: Instant Web-Based Access to Public Brain Atlases and Related Content. *Neuroinformatics* **13**, 353–366 (2015).

94. Lim, I. A. L. *et al.* Human brain atlas for automated region of interest selection in quantitative susceptibility mapping: Application to determine iron content in deep gray matter structures. *Neuroimage* **82**, 449–469 (2013).

95. Nooner, K. B. *et al.* The NKI-Rockland sample: A model for accelerating the pace of discovery science in psychiatry. *Front. Neurosci.* (2012). doi:10.3389/fnins.2012.00152

96. Job, D. E. *et al.* A brain imaging repository of normal structural MRI across the life course: Brain Images of Normal Subjects (BRAINS). *Neuroimage* **144**, 299–304 (2017).

97. Winkler, A. M., Kochunov, P. & Glahn, D. FLAIR templates. (2012).

98. Upadhyay, N. & Waldman, A. D. Conventional MRI evaluation of gliomas. *Br. J. Radiol.* **84**, 107–111 (2011).

99. Fedorov, A. *et al.* 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magn. Reson. Imaging* **30**, 1323–1341 (2012).

100. Van Griethuysen, J. J. M. *et al.* Computational radiomics system to decode the radiographic phenotype. *Cancer Res.* **77**, e104–e107 (2017).

101. Vallières, M., Freeman, C. R., Skamene, S. R. & El Naqa, I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys. Med. Biol.* **60**, 5471–5496 (2015).

102. Han, Y. S., Yoo, J. & Ye, J. C. Deep Residual Learning for Compressed Sensing CT Reconstruction via Persistent Homology Analysis. **4**, (2016).

103. Africa, S. HHS Public Access. **4**, 910–914 (2017).

104. Society, E. ESR statement on the stepwise development of imaging biomarkers. *Insights Imaging* **4**, 147–152 (2013).

105. Sullivan, D. C. *et al.* Metrology Standards for Quantitative Imaging Biomarkers.

Radiology **277**, 813–825 (2015).

106. Vaquerizas, J. M. *et al.* GEPAS, an experiment-oriented pipeline for the analysis of microarray gene expression data. *Nucleic Acids Res.* **33**, (2005).

107. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 6567–6572 (2002).

108. Li, L., Weinberg, C. R., Darden, T. A. & Pedersen, L. G. Gene selection for sample classification based on gene expression data: study of sensitivity to choice of parameters of the GA/KNN method. *Bioinformatics* **17**, 1131–1142 (2001).

109. Ooi, C. H. & Tan, P. Genetic algorithms applied to multi-class prediction for the analysis of gene expression data. *Bioinformatics* **19**, 37–44 (2003).

110. Sha, N. *et al.* Bayesian variable selection in multinomial probit models to identify molecular signatures of disease stage. *Biometrics* **60**, 812–819 (2004).

111. Trevino, V. & Falciani, F. GALGO: An R package for multivariate variable selection using genetic algorithms. *Bioinformatics* **22**, 1154–1156 (2006).

112. Taylor, C. E. Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence. Complex Adaptive Systems. John H. Holland . *Q. Rev. Biol.* **69**, 88–89 (1994).

113. Langs, G. *et al.* Machine learning: from radiomics to discovery and routine. *Radiologe* **58**, 1–6 (2018).

114. Pavlov, Y. L. Random forests. *Random For.* 1–122 (2019). doi:10.1007/978-3-662-56776-0_10

115. Rodríguez, J. D., Pérez, A. & Lozano, J. A. Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **32**, 569–575 (2010).

116. Lu, C. F. *et al.* Machine learning–based radiomics for molecular subtyping of gliomas. *Clin. Cancer Res.* **24**, 4429–4436 (2018).

117. Ren, Y. *et al.* Noninvasive Prediction of IDH1 Mutation and ATRX Expression Loss in Low-Grade Gliomas Using Multiparametric MR Radiomic Features. *J. Magn. Reson. Imaging* **49**, 808–817 (2019).

118. Choi, Y. S. *et al.* Fully automated hybrid approach to predict the IDH mutation status of gliomas via deep learning and radiomics. *Neuro. Oncol.* 1–10 (2020). doi:10.1093/neuonc/noaa177

119. Li, Z., Wang, Y., Yu, J., Guo, Y. & Cao, W. Deep Learning based Radiomics (DLR) and its usage in noninvasive IDH1 prediction for low grade glioma. *Sci. Rep.* **7**, 1–11 (2017).

120. Bangalore Yogananda, C. G. *et al.* A novel fully automated MRI-based deep-learning method for classification of IDH mutation status in brain gliomas. *Neuro. Oncol.* **22**, 402–411 (2020).

121. Zhang, X. *et al.* IDH mutation assessment of glioma using texture features of multimodal MR images. *Med. Imaging 2017 Comput. Diagnosis* **10134**, 101341S (2017).

122. Yu, J. *et al.* Noninvasive IDH1 mutation estimation based on a quantitative radiomics approach for grade II glioma. *Eur. Radiol.* **27**, 3509–3522 (2017).

123. Liu, X. *et al.* IDH mutation-specific radiomic signature in lower-grade gliomas. *Aging (Albany. NY).* **11**, 673–696 (2019).

124. Inglese, M. *et al.* Whole-brain N-acetylaspartate spectroscopy and diffusion tensor imaging in patients with newly diagnosed gliomas: A preliminary study. *Am. J.*

Neuroradiol. **27**, 2137–2140 (2006).

125. Eis, M., Els, T., Hoehn-Berlage, M. & Hossmann, K. A. Quantitative diffusion MR imaging of cerebral tumor and edema. *Acta Neurochir. Suppl. (Wien).* **60**, 344–346 (1994).

126. Pei, J., Tseng, V. S., Conference, P. & Goebel, R. *Advances in Knowledge Discovery in Databases*. *Journal of Applied Economic Sciences (JAES)* **III**, (2008).

127. Gliomas, G. *et al.* Peritumoral Diffusion Tensor Imaging of High-. *Ajnr. Am. J. Neuroradiol.* 937–941 (2003).

128. Gimenez, U. *et al.* Microscopic DTI accurately identifies early glioma cell migration: correlation with multimodal imaging in a new glioma stem cell model. *NMR Biomed.* **29**, 1553–1562 (2016).

129. Thanh Noi, P. & Kappas, M. Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel).* **18**, (2017).

130. Saeys, Y., Inza, I. & Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**, 2507–2517 (2007).

131. Bischl, B. *et al.* Mlr: Machine learning in R. *J. Mach. Learn. Res.* **17**, 1–5 (2016).

132. Yang, F., Dogan, N., Stoyanova, R. & Ford, J. C. Evaluation of radiomic texture feature error due to MRI acquisition and reconstruction: A simulation study utilizing ground truth. *Phys. Medica* **50**, 26–36 (2018).

133. Saha, A., Yu, X., Sahoo, D. & Mazurowski, M. A. Effects of MRI scanner parameters on breast cancer radiomics. *Expert Syst. Appl.* **87**, 384–391 (2017).

134. Ford, J., Dogan, N., Young, L. & Yang, F. Quantitative Radiomics: Impact of Pulse Sequence Parameter Selection on MRI-Based Textural Features of the Brain. *Contrast Media Mol. Imaging* **2018**, (2018).

135. Xia, J., Broadhurst, D. I., Wilson, M. & Wishart, D. S. Translational biomarker discovery in clinical metabolomics: An introductory tutorial. *Metabolomics* **9**, 280–299 (2013).

136. Differences, S. of of NMR. **170**, 598–600 (1985).

137. Bollard, M. E. *et al.* Metabolic profiling of the effects of D-galactosamine in liver spheroids using 1H NMR and MAS-NMR spectroscopy. *Chem. Res. Toxicol.* **15**, 1351–1359 (2002).

138. Beckonert, O. *et al.* Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nat. Protoc.* **2**, 2692–2703 (2007).

139. Nicholson, J. K., Connelly, J., Lindon, J. C. & Holmes, E. Metabonomics: A platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* **1**, 153–161 (2002).

140. Wilson, D. M., Burlingame, A. L., Cronholm, T. & Sjövall, J. Deuterium and carbon-13 tracer studies of ethanol metabolism in the rat by 2H, 1H-decoupled 13C nuclear magnetic resonance. *Biochem. Biophys. Res. Commun.* **56**, 828–835 (1974).

141. Weber, S. G. & Wishart, D. S. Quantitative metabolomics using NMR. *TrAC Trends Anal. Chem.* **27**, 228–237 (2008).

142. Nicholson, J. K. & Wilson, I. D. High resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectrosc.* **21**, 449–501 (1989).

143. Markley, J. L. *et al.* The future of NMR-based metabolomics. *Current Opinion in Biotechnology* (2017). doi:10.1016/j.copbio.2016.08.001

144. Duarte, I. F., Diaz, S. O. & Gil, A. M. NMR metabolomics of human blood and urine in

disease research. *Journal of Pharmaceutical and Biomedical Analysis* (2014). doi:10.1016/j.jpba.2013.09.025

145. Wishart, D. S. *et al.* HMDB: The human metabolome database. *Nucleic Acids Res.* **35**, 521–526 (2007).

146. Weljie, A. M., Newton, J., Mercier, P., Carlson, E. & Slupsky, C. M. Targeted pofiling: Quantitative analysis of1H NMR metabolomics data. *Anal. Chem.* **78**, 4430–4442 (2006).

147. Fan, T. W. M., Bandura, L. L., Higashi, R. M. & Lane, A. N. Metabolomics-edited transcriptomics analysis of Se anticancer action in human lung cancer cells. *Metabolomics* **1**, 325–339 (2005).

148. Borlak, J. *et al.* Verapamil: Identification of novel metabolites in cultures of primary human hepatocytes and human urine by LC-MSn and LC-NMR. *Xenobiotica* **33**, 655–676 (2003).

149. Giraudeau, P. Quantitative 2D liquid-state NMR. *Magn. Reson. Chem.* **52**, 259–272 (2014).

150. Bingol, K. & Brüschweiler, R. Multidimensional APPROACHES to NMR-based metabolomics. *Anal. Chem.* **86**, 47–57 (2014).

151. Dumas, M. E., Canlet, C., Andŕ, F., Vercauteren, J. & Paris, A. Metabonomic assessment of physiological disruptions using 1H-13C HMBC-NMR spectroscopy combined with pattern recognition procedures performed on filtered variables. *Anal. Chem.* **74**, 2261–2273 (2002).

152. Reitman, Z. J. & Yan, H. Isocitrate dehydrogenase 1 and 2 mutations in cancer: Alterations at a crossroads of cellular metabolism. *J. Natl. Cancer Inst.* **102**, 932–941 (2010).

153. Parsons, D. W. *et al.* An Integrated Genomic Analysis of. *Science (80-.).* **321**, 1807–1812 (2008).

154. Al-Mutawa, Y. K. *et al.* Effects of hypoxic preconditioning on neuroblastoma tumour oxygenation and metabolic signature in a chick embryo model. *Biosci. Rep.* **38**, 1–15 (2018).

155. Chignola, F. *et al.* The CCPN metabolomics project: A fast protocol for metabolite identification by 2D-NMR. *Bioinformatics* **27**, 885–886 (2011).

156. No Title. *http://www.bmrb.wisc.edu/search/multiple_peak_search.php*

157. Broadhurst, D. I. & Kell, D. B. Statistical strategies for avoiding false discoveries in metabolomics and related experiments. *Metabolomics* **2**, 171–196 (2006).

158. Westerhuis, J. A. *et al.* Assessment of PLSDA cross validation. *Metabolomics* **4**, 81–89 (2008).

159. Hyung, G., Park, J., Yang, S. & Baek, H. hydroxyglutarate and other brain metabolites in human brain tumor tissue extracts. 1–16 (2018).

160. Jalbert, L. E. *et al.* Metabolic profiling of IDH mutation and malignant progression in infiltrating glioma. *Sci. Rep.* **7**, 1–10 (2017).

161. Cuperlovic-Culf, M., Ferguson, D., Culf, A., Morin, P. & Touaibia, M. 1 H NMR metabolomics analysis of glioblastoma subtypes: Correlation between metabolomics and gene expression characteristics. *J. Biol. Chem.* **287**, 20164–20175 (2012).

162. Shao, W. *et al.* Malignancy-associated metabolic profiling of human glioma cell lines using 1 H NMR spectroscopy. *Mol. Cancer* **13**, 1–12 (2014).

163. Paterson, C., Nowak, M. A. & Waclaw, B. An exactly solvable, spatial model of mutation accumulation in cancer. *Sci. Rep.* **6**, 1–19 (2016).

175

164.   Waclaw, B. *et al.* A spatial model predicts that dispersal and cell turnover limit intratumour heterogeneity. *Nature* **525**, 261–264 (2015).

165.   Nowak, B. M. A. & Waclaw, B. Genes, environment, and "bad luck". **355**, 1266–1268 (2017).

166.   Saccenti, E., Hoefsloot, H. C. J., Smilde, A. K., Westerhuis, J. A. & Hendriks, M. M. W. B. Reflections on univariate and multivariate analysis of metabolomics data. *Metabolomics* **10**, 361–374 (2014).

167.   Sch, J. Shrinkcov2005. (2005).

168.   Koskela, H., Kilpeläinen, I. & Heikkinen, S. Some aspects of quantitative 2D NMR. *J. Magn. Reson.* (2005). doi:10.1016/j.jmr.2005.02.002

169.   Morris, G. A. Systematic sources of signal irreproducibility and t1 noise in high-field NMR spectrometers. *J. Magn. Reson.* **100**, 316–328 (1992).

170.   Sources of tI Noise in Two-Dimensional. **323**, (1984).

171.   Bizzarri, M., Dinicola, S. & Cucina, A. Myoinositol and Inositol Hexakisphosphate in the Treatment of Breast Cancer: Molecular Mechanisms. in *Pre-Menopause, Menopause and Beyond: Volume 5: Frontiers in Gynecological Endocrinology* (eds. Birkhaeuser, M. & Genazzani, A. R.) 233–241 (Springer International Publishing, 2018). doi:10.1007/978-3-319-63540-8_20

172.   Koguchi, T., Tanikawa, C., Mori, J., Kojima, Y. & Matsuda, K. Regulation of myo-inositol biosynthesis by p53-ISYNA1 pathway. *Int. J. Oncol.* **48**, 2415–2424 (2016).

173.   Liu, G., Song, Y., Cui, L., Wen, Z. & Lu, X. Inositol hexaphosphate suppresses growth and induces apoptosis in HT-29 colorectal cancer cells in culture: PI3K/Akt pathway as a potential target. *Int. J. Clin. Exp. Pathol.* **8**, 1402–1410 (2015).

174.   Huang, C., Ma, W. Y., Hecht, S. S. & Dong, Z. Inositol hexaphosphate inhibits cell transformation and activator protein 1 activation by targeting phosphatidylinositol-3' kinase. *Cancer Res.* **57**, 2873–2878 (1997).

175.   Dinicola, S. *et al.* Inositol induces mesenchymal-epithelial reversion in breast cancer cells through cytoskeleton rearrangement. *Exp. Cell Res.* **345**, 37–50 (2016).

176.   Warburg, O. The metabolism of carcinoma cells 1. *J. Cancer Res.* **9**, 148–163 (1925).

177.   Michelakis, E. D., Webster, L. & Mackey, J. R. Dichloroacetate (DCA) as a potential metabolic-targeting therapy for cancer. *Br. J. Cancer* **99**, 989–994 (2008).

178.   Sun, R. C. *et al.* Reversal of the glycolytic phenotype by dichloroacetate inhibits metastatic breast cancer cell growth in vitro and in vivo. *Breast Cancer Res. Treat.* **120**, 253–260 (2010).

179.   Beuster, G. *et al.* Inhibition of alanine aminotransferase in Silico and in vivo promotes mitochondrial metabolism to impair malignant growth. *J. Biol. Chem.* **286**, 22323–22330 (2011).

180.   Gupta, R. K. *et al.* Relationships between choline magnetic resonance spectroscopy, apparent diffusion coefficient and quantitative histopathology in human glioma. *J. Neurooncol.* **50**, 215–226 (2000).

181.   Michel, V., Yuan, Z., Ramsubir, S. & Bakovic, M. Choline transport for phospholipid synthesis. *Exp. Biol. Med.* **231**, 490–504 (2006).

182.   McKnight, T. R. *et al.* Choline metabolism, proliferation, and angiogenesis in nonenhancing grades 2 and 3 astrocytoma. *J. Magn. Reson. Imaging* **33**, 808–816 (2011).

183.   Maurer, G. D. *et al.* Differential utilization of ketone bodies by neurons and glioma cell lines: A rationale for ketogenic diet as experimental glioma therapy. *BMC Cancer*

11, (2011).

184. Xiao, M. *et al.* Inhibition of α-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* **26**, 1326–1338 (2012).

185. Xia, J., Sinelnikov, I. V., Han, B. & Wishart, D. S. MetaboAnalyst 3.0-making metabolomics more meaningful. *Nucleic Acids Res.* **43**, W251–W257 (2015).

186. Berrington, A. *et al.* Improved localisation for 2-hydroxyglutarate detection at 3T using long-TE semi-LASER. *Tomogr. a J. imaging Res.* **2**, 94–105 (2016).

187. Provencher, S. W. Automatic quantitation of localized in vivo 1H spectra with LCModel. *NMR Biomed.* **14**, 260–264 (2001).

188. Bal, D. & Gryff-Keller, A. 1H and 13C NMR study of 2-hydroxyglutaric acid and its lactone. *Magn. Reson. Chem.* **40**, 533–536 (2002).

189. Sener, R. N. R. L-2 Hydroxyglutaric Aciduria: Proton Magnetic Resonance Spectroscopy and Diffusion Magnetic Resonance Imaging Findings. *J. Comput. Assist. Tomogr.* **27**, 38–43 (2003).

190. An, Z. *et al.* Detection of 2-hydroxyglutarate in brain tumors by triple-refocusing MR spectroscopy at 3T in vivo. *Magn. Reson. Med.* **00**, 1–9 (2016).

191. Andronesi, O. C. *et al.* Detection of 2-Hydoxyglutarate in IDH-mutated Glioma Patients by Spectral-editing and 2D Correlation Magnetic Resonance Spectroscopy. *Sci Transl Med* **4**, (2012).

192. Goffette, S. M. *et al.* L-2-Hydroxyglutaric aciduria: Clinical, genetic, and brain MRI characteristics in two adult sisters. *Eur. J. Neurol.* **13**, 499–504 (2006).

193. Andronesi, O. C. *et al.* Review series Detection of oncogenic IDH1 mutations using magnetic resonance spectroscopy of 2-hydroxyglutarate. *J. Clin. Invest.* **123**, 3659–3663 (2013).

194. Andronesi, O. C. *et al.* Treatment response assessment in IDH-mutant glioma patients by noninvasive 3D functional spectroscopic mapping of 2-hydroxyglutarate. *Clin. Cancer Res.* **22**, 1632–1641 (2016).

195. Verma, G. *et al.* Non-invasive detection of 2-hydroxyglutarate in IDH-mutated gliomas using two-dimensional localized correlation spectroscopy (2D L-COSY) at 7 Tesla. *J. Transl. Med.* **14**, 274 (2016).

196. Kim, H., Kim, S., Lee, H. H. & Heo, H. In-Vivo Proton Magnetic Resonance Spectroscopy of 2-Hydroxyglutarate in Isocitrate Dehydrogenase-Mutated Gliomas: A Technical Review for Neuroradiologists. *Korean J. Radiol.* **17**, 620–32 (2016).

197. Reitman, Z. J. *et al.* Profiling the effects of isocitrate dehydrogenase 1 and 2 mutations on the cellular metabolome. *Pnas* **108**, 3270–3275 (2011).

198. Gioacchino Tedeschi *et al.* Increased choline signal coinciding with malignant degeneration of cerebral gliomas: a serial proton magnetic resonance spectroscopy imaging study. *J. Neurosurg.* **87**, 516–524 (1997).

199. Dowling, C. *et al.* Preoperative proton MR spectroscopic imaging of brain tumors: Correlation with histopathologic analysis of resection specimens. *Am. J. Neuroradiol.* **22**, 604–612 (2001).

200. De La Fuente, M. I. *et al.* Integration of 2-hydroxyglutarate-proton magnetic resonance spectroscopy into clinical practice for disease monitoring in isocitrate dehydrogenase-mutant glioma. *Neuro. Oncol.* **18**, 283–290 (2016).

201. Allen, P. S., Thompson, R. B. & Wilman, A. H. Metabolite-specific NMR spectroscopy in vivo. *NMR Biomed.* **10**, 435–444 (1997).

202. Kim, H., Thompson, R. B., Hanstock, C. C. & Allen, P. S. Variability of metabolite yield using STEAM or PRESS sequences in vivo at 3.0 T, illustrated with myo-inositol. *Magn. Reson. Med.* **53**, 760–769 (2005).

203. Thompson, R. B. & Allen, P. S. A new multiple quantum filter design procedure for use on strongly coupled spin systems foundin vivo: Its application to glutamate. *Magn. Reson. Med.* **39**, 762–771 (1998).

204. Mescher, M., Merkle, H., Kirsch, J., Garwood, M. & Gruetter, R. Simultaneous in vivo spectral editing and water suppression. *NMR Biomed.* **11**, 266–272 (1998).

205. Ganji, S. K. *et al.* In vivo detection of 2-hydroxyglutarate in brain tumors by optimized point-resolved spectroscopy (PRESS) at 7T. *Magn. Reson. Med.* **944**, 936–944 (2016).

206. Izquierdo-Garcia, J. L. *et al.* Metabolic reprogramming in mutant IDH1 glioma cells. *PLoS One* **10**, (2015).

207. Howe, F. A. *et al.* Metabolic profiles of human brain tumors using quantitative in vivo 1H magnetic resonance spectroscopy. *Magn. Reson. Med.* **49**, 223–232 (2003).

208. Kaiser, L. G., Young, K. & Matson, G. B. Numerical simulations of localized high field 1H MR spectroscopy. *J. Magn. Reson.* **195**, 67–75 (2008).

209. Gliomas, I. L. *et al.* Magnetic Resonance of 2-Hydroxyglutarate in. *Sci. Transl. Med.* **4**, 1–10 (2012).

210. Usinskiene, J. *et al.* Optimal differentiation of high- and low-grade glioma and metastasis: a meta-analysis of perfusion, diffusion, and spectroscopy metrics. *Neuroradiology* **58**, 339–350 (2016).

211. Longworth, L. National Institute For Health And Clinical Excellence Overview Carmustine implants and temozolomide for the treatment of newly diagnosed high-grade glioma. 1–13 (2007).

212. Aerts, H. J. W. L. *et al.* Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat. Commun.* **5**, (2014).

213. Huang, Y. Q. *et al.* Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. *J. Clin. Oncol.* **34**, 2157–2164 (2016).

214. Zhang, B. *et al.* Radiomics features of multiparametric MRI as novel prognostic factors in advanced nasopharyngeal carcinoma. *Clin. Cancer Res.* **23**, 4259–4269 (2017).

215. Lester McCully, C. *et al.* Plasma and cerebrospinal fluid pharmacokinetics of the DNA methyltransferase inhibitor, 5-azacytidine, alone and with inulin, in nonhuman primate models. *Neuro-Oncology Adv.* **2**, (2020).

Appendix 1

Imaging acquisition parameters for radiomics study

Patients for this study were identified by histology report. All glioma patients who had undergone IDH mutation investigation and had been imaged were included.

All T_1 images were acquired as gradient echo images with gadolinium contrast enhancement, across 3 scanners manufactured by Philips, Siemens and GE. The acquisition parameters for the images collected on each scanner were as follows:

GE –

TR:8.132ms, TE: 3.164ms, Number of averages: 1, Echo train length:1, Bandwidth: 244.14, Acquisition time: 4mins 22seconds , Resolution 256x256x280, Voxel spacing: 1.016x1.016x0.7, Number of slices: 180

A total of 107 patients were scanned with this sequence and included in the analysis.

Philips –

TR: 9.0ms, TE: 1.369ms, Number of Averages: 1, Echo train length: 1, Bandwidth: 343, Acquisition time: 4min 38seconds, Resolution 288x288x180, Voxel spacing: 0.868x0.868x1, Number of slices: 180

A total of 68 patients were scanned with this sequence and included in the analysis.

Siemens –

TR: 8.42ms, TE: 1.362ms, Number of Averages: 1, Echo train length: 1, Bandwidth: 287, Acquisition time: 4min 32seconds, Resolution 256x256x280, Voxel spacing: 0.946x0.946x1, Number of slices: 180

A total of 36 patients were scanned with this sequence and included in the analysis.

All T_2 images as Spin echo sequences as PROPELLER(GE)/BLADE(Siemens)/MultiVane(Philips) across three scanner. The acquisition parameters for the images collected were as follows:

GE –

TR: 6736.544ms, TE: 102.18ms, Number of averages: 2, Echo train length: 26, Bandwidth: 195, Acquisition time: 2min 04seconds, Resolution:512x512x29, Voxel spacing: 0.4688x0.4688x5, Number of slices, 29

A total of 77 patients were scanned with this sequence and included in the analysis.

Philips –

TR: 2999.99ms, TE: 80ms, Number of averages: 1, Echo train length: 15, Bandwidth: 204, Acquisition time: 3min 35seconds, Resolution: 512x512x27, Voxel spacing: 0.4492x0.4492x5, Number of slices: 27

A total of 43 patients were scanned with this sequence and included in the analysis.

Siemens –

TR: 3800ms, TE: 100ms, Number of averages: 2, Echo train length: 17, Bandwidth: 220, Acquisition time: 3min 47seconds, Resolution: 512x512x32, Voxel spacing: 0.4492x0.4492x5.2, Number of slices: 32

A total of 38 patients were scanned with this sequence and included in the analysis.

Tumour Type

GBMs from a number of different origins were included in this study. These were progressions from either astrocytic or oligodendrocytic tumours, as well as GMB. The breakdown of these tumour types are as follows:

Astrocytoma: 13

GBM: 129

Oligodendroglioma: 11

The segmented enhancing region of the tumour that was used in the study ranged in size from $1cm^3 - 65cm^3$.

Appendix 2

NMR Pattern file metabolites

2-HG
Adenosine
ADP
Alanine
Aspartate
ATP
Cystathionine
Cytidine

Formate

Fumarate

GDP

Glucose

Glutothione

Glycine

GTP

Guanosine

Histidine

Hypoxanthine

Imidazole

Inosine

Lactate

Lactulose

Malonate

Myo-insoitol

N-Acetylglutamate

NAD+

NADP

Niacinamide

Nicotinurate

Phenyalanine

s-adenosylhomocysteine

sn-Glycero-3-phosphocholine

Sucrose

Syringate

Trimethylamine-N-oxide

Tyrosine

Tyrosine

UDP-Galactose

UDP-Glucose

UDP-N-acetylglucosamine

UDP-N-acetylglucosamine-UDP-Glc

UMP

Uridine

Valine

Xanthine