RAMAN SPECTROSCOPIC AND MICROSCOPIC ANALYSIS OF TISSUE TYPE, MOLECULAR COMPOSITION, AND GLIOBLASTOMA IDENTIFICATION IN BRAIN TISSUE SECTIONS

by

Hülya TORUN

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Master of Science

in

Biomedical Sciences and Engineering

February 8, 2021

RAMAN SPECTROSCOPIC AND MICROSCOPIC ANALYSIS OF TISSUE TYPE, MOLECULAR COMPOSITION, AND GLIOBLASTOMA IDENTIFICATION IN BRAIN TISSUE SECTIONS

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a master's thesis by

Hülya TORUN

and have found that it is complete and satisfactory in all respects,

and that any and all revisions required by the final

examining committee has been made.

COMMITTEE MEMBERS

Prof. İhsan Solaroğlu (Advisor)

Asst. Prof. Mehmet Cengiz Onbaşlı (Co-Advisor)

Prof. Yasemin Gürsoy Özdemir

Assoc. Prof. Sedat Nizamoğlu

Assoc. Prof. İlke Gürol

Date: February 8, 2021

DEDICATION

I dedicated this work to;

My paternal grandfather, Hilmi Torun, had always desired his granddaughter to be a Professor, and unfortunately, his life was not long enough to see his dream come true. I can feel that you are so proud of me and I promise I will honor you one day by making your dream come true...

My enatic grandfather, Turgut Baykal, your girl will always remember your faith in her and strive to do her best in whatever she undertakes...

My grandmothers, Mesude Torun and Sadiye Baykal, I cannot thank you enough for your prayers and moral supports in everything I was thriving for. At least, one of you are witnessing me reaching my goals step by step, whilst the other is watching me from the other side, with love.

Mom, Dad, and Brother, Ayse, Mustafa, and Kenan Torun, for your endless love, support, and encouragement for everything I intended to do. I could not accomplish anything without you, and your infinite love and faith in me.

To all my teachers and mentors, who have taught me even a word, I have always been very lucky about my teachers and always met the best ones in their fields. I am so grateful to all of you for paving the way for me to be where I am today.

Kardelen Yangin, and Ece Cepni, for being far more than a best friend and for always being with me through my darkest and the brightest days.

To all women, who share similar goals and dreams with me, who thrive for becoming the best version of themselves every single day, and who never give up, no matter what the conditions are.

ABSTRACT

RAMAN SPECTROSCOPIC AND MICROSCOPIC ANALYSIS OF TISSUE TYPE, MOLECULAR COMPOSITION, AND GLIOBLASTOMA IDENTIFICATION IN BRAIN TISSUE SECTIONS

Hülya TORUN

Master of Science in Biomedical Sciences and Engineering

February 8, 2021

Glioblastoma (GB) is the most common primary malignant brain tumor. Despite improvements in treatments, survival probability has remained shorter than 2 years for most patients over the last 20 years. Accurate diagnosis of GB requires pathological evaluation of the tumor tissues using light microscopy, along with routine or specialized staining. Recent research also identified significant genetic/epigenetic alterations that influence diagnosis, prognosis, and treatment in addition to routine pathological evaluation. Identification requires the tissue to be sampled many times and analyzed using different methods that require additional time, resources, and expertise.

To determine whether the tissue used for routine analysis can also be used to perform more detailed and comprehensive analysis without staining, we propose to use Raman Spectroscopy (RS), which is a label-free and non-destructive technique. RS provides molecule-specific spectra from the chemical composition of the sample for rapid analysis. In this thesis, we investigated GB, white matter (WM), gray matter (GM), and necrosis (NC) regions of GB patients using RS to determine whether a similar precision can be achieved as the routine histomorphologic diagnostic process.

First, we proposed a refined protocol for effectively clearing paraffin from Formalin-Fixed Paraffin-Embedded brain tissue sections, without destroying the sample morphology and chemical composition, for eliminating the substantial Raman spectra of paraffin. We demonstrated that the less expensive and less toxic clearing agent Clearene^{TM} removes paraffin as effectively as p-Xylene, the mostly used clearing agent in histopathology laboratories. Thus, we suggest substituting Clearene^{TM} with p-Xylene for deparaffinization of brain tissue sections for Raman spectral analysis. Second, we optimized the choice of Raman spectrum acquisition parameters (excitation wavelength, acquisition time, accumulation count,), tissue thickness, and Raman substrate type (CaF_2, glass). Third, we acquired the Raman spectra of GB, WM, GM, and NC regions and analyzed the spectral profile regarding the Raman peaks given in the literature. Raman spectra of GB and WM regions ($n_{GB} = 20$, $n_{WM} = 18$), which were annotated by an expert neuropathologist, have been classified with 87.2±1% GB and 90.7±1% WM training/test accuracies using machine learning models (SVM, kNN, RF). The effect of pre-processing of Raman spectra on classification accuracies has been investigated.

Sample preparation conditions, Raman acquisition protocols, and machine learning classification models showed a successful proof-of-concept demonstration for the proposed Raman-based GB identification workflow. While there is room for further refining the machine learning models for improved training and validation accuracies, these protocols could be improved for eventual clinical utility. Once the clinical applicability and refined classification accuracies are demonstrated, these protocols might assist neuropathologists in error-free identification of GB in the clinics.

ÖZETÇE

BEYİN DOKU KESİTLERİNDE RAMAN SPEKTROSKOPİ VE MİKROSKOPİSİ KULLANILARAK DOKU TİPİ, MOLEKÜLER KOMPOZİSYON VE GLİOBLASTOMA TESPİTİ

Hülya TORUN

Biyomedikal Bilimler ve Mühendislik, Yüksek Lisans

8 Şubat, 2021

Glioblastoma (GB) en sık görülen birincil malign beyin tümörüdür. Tedavilerdeki gelişmelere rağmen, son 20 yılda çoğu hasta için hayatta kalma olasılığı 2 yıldan daha kısa kalmıştır. GB'nin doğru teşhisi, rutin veya özel boyama ile birlikte ışık mikroskobu kullanılarak tümör dokularının patolojik değerlendirmesini gerektirir. Yakın zamanda yapılan araştırmalar, rutin patolojik değerlendirmeye ek olarak tanı ve tedaviyi etkileyen önemli genetik/epigenetik değişiklikleri de tespit etmiştir. Tümör tespiti, dokunun birçok kez örneklenmesini ve ek zaman, kaynak ve uzmanlık gerektiren farklı yöntemler kullanılarak analiz edilmesini gerektirir.

Rutin analiz için kullanılan dokunun boyama veya genetik analiz yapmadan daha detaylı ve kapsamlı analiz yapmak için kullanılıp kullanılamayacağını belirlemek için etiketsiz ve tahribatsız bir teknik olan Raman Spektroskopisini (RS) kullanmayı öneriyoruz. RS, hızlı analiz için bir numunenin kimyasal bileşiminden moleküle özgü spektrumlar ölçülmesini sağlar. Bu tezde, rutin histomorfolojik tanı sürecindekine benzer bir tanı hassasiyetinin elde edilip edilemeyeceğini belirlemek için RS kullanarak glioblastoma, beyaz madde (BM), gri madde (GM) ve nekroz (NK) bölgelerini araştırdık.

İlk olarak, parafinin güçlü Raman spektrumunu ortadan kaldırmak için, doku morfolojisini ve kimyasal bileşimi bozmadan, formalin ile fikse edilmiş parafine gömülü beyin dokusu kesitlerinden parafini etkili bir şekilde temizlemek için geliştirdiğimiz bir protokol önerdik. İkinci olarak, Raman spektrum ölçüm parametreleri (uyarma dalga boyu, edinim süresi, ölçüm sayısı), doku kalınlığı ve Raman alttaş türü (CaF_2, cam) seçimini optimize ettik. Üçüncüsü, GB, BM, GM ve NK bölgelerinin Raman spektrumlarını ölçtük ve elde ettiğimiz spektral profillerini literatürde verilen Raman pikleri ile analiz ettik. Uzman bir nöropatolog tarafından belirlenmiş GB ve BM bölgelerinin (n_{GB} = 20, n_{WM} = 18) Raman spektrumları, makine öğrenimi modelleri kullanılarak %87.2±1 GB ve %90.7±1 BM eğitim / test doğruluğu ile sınıflanmıştır. Raman spektrumlarının ön işlenmesinin sınıflandırma doğruluklarına etkisi araştırılmıştır.

Örnek hazırlama koşulları, Raman ölçüm protokolleri ve makine öğrenimi sınıflama modelleri, önerilen Raman temelli GB tanımlama iş akışının başarıyla uygulandığını göstermektedir. Gelişmiş eğitim ve test doğrulukları için makine öğrenimi modellerinin daha da çalışılması gerekse de, bu protokoller ileride klinikte fayda sağlamak üzere geliştirilebilir. Bu tekniklerin kliniğe uygulanabilirliği ve daha yüksek örneklemle ilerletilen sınıflandırma doğruluğu gösterildikten sonra, bu protokoller kliniklerde patologlara GB'nin daha yüksek doğrulukla tanımlanmasında yardımcı olabilir.

ACKNOWLEDGMENTS

First and foremost, I would like to express my deepest and sincere gratitude and special regards to my master's thesis advisors, Professor Ihsan Solaroglu and Asst. Professor Mehmet Cengiz Onbasli, for their unwavering support and invaluable guidance. Without their encouragement and faith in me, this work would not have been achieved. Working with them and studying under their mentorship is always a great privilege and honor for me.

Besides my advisors, I would like to extend my sincere thanks to Professor Tarik Tihan for his technical, theoretical, and motivational contribution to this thesis study, as well as his inestimable supports to me. He has taught me the research methodology and played a significant role in my achievements in my master's thesis project. Sample supplies, technical, and research support provided by Prof. Tarık Tihan and UCSF Neuropathology are gratefully acknowledged.

I wish to express my appreciation to Dr. Ibrahim Kulac for the technical, theoretical, and motivational contribution to my master's thesis project. Without his persistent help at every stage of the project, this project would not have been accomplished. I also wish to pay my thanks to our collaborator Professor Kemal Baysal for their insightful comments, suggestions, and contributions to my thesis project.

I am deeply grateful to my colleague Buse Bilgin for her assistance and motivational support at every step of this project. Besides, I also would like to offer my special thanks and appreciation to Numan Batur for his contribution to the machine learning classification part of this project. His consistent efforts and help are gratefully acknowledged. I also would like to thank Assoc. Professor Ilke Gurol for her continuous support in facility use and Dr. Arzu Baygul for sample size calculation and statistical analysis.

I also would like to express my appreciation to Gokce Deniz Ardor and Roni Kazazi, for their infinite support and contributions in tissue selection and annotation from neuropathology archives.

We gratefully acknowledge the use of the services and facilities of the Koç University Research Center for Translational Medicine (KUTTAM), funded by the Presidency of Turkey, Presidency of Strategy and Budget.

Funding support by TÜBİTAK Project No. 119S362 is gratefully acknowledged.

I would like to express my deepest thanks to my family for their limitless love and support. Without their unconditional and endless support, I would not have been able to complete any of my achievements in my master's studies. They taught me to be professional and keep my faith and motivation even when the roughest parts of the road.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

CaF_2	Calcium Fluoride
CHR	Committee of Human Research
CNS	Central Nervous System
CT	Computed tomography
EGFR	Epidermal Growth Factor Receptor
DFA	Discriminant Functional Analysis
FDA	Food and Drug Administration
FFPE	Formalin-Fixed Paraffin-Embedded
fMRI	Functional Magnetic Resonance Imaging
FTIR	Fourier-Transform Infrared Spectroscopy.
GB	Glioblastoma
GUH	Gazi University Hospital
HOMO	Highest Occupied Molecular Orbital
H&E	Hematoxylin & Eosin
ICA	Independent Component Analysis
IDH	Isocitrate dehydrogenase
IHC	Immunohistochemistry
iMRI	Intraoperative Magnetic Resonance Imaging
IR	Infrared Spectroscopy
IRB	Institutional Review Board
kNN	k-Nearest Neighbors
KUH	Koç University Hospital
LDA	Linear Discriminant Analysis
LUMO	Lowest Unoccupied Molecular Orbital
MRI	Magnetic Resonance Imaging
NSC	Neural Stem Cells

PC	Principal Component
PCA	Principal Component Analysis
PET	Positron Emission Tomography
RF	Random Forest
ROC	Receiver Operating Characteristics
RS	Raman Spectroscopy
RF	Random Forest
SERS	Surface-Enhanced Raman Spectroscopy
SNR	Signal-to-Noise Ratio
SV	Support Vector
SVM	Support Vector Machine
TMA	Tissue Microarray
UCSF	University of California, San Francisco
US	Ultrasound
wt	Wild-Type

Chapter 1: **INTRODUCTION**

1.1 Glioblastoma

Glioblastoma (GB) is the most common and malignant Central Nervous System (CNS) neoplasm in adults and is considered a major tumor type among gliomas [1-3]. According to the World Health Organization (WHO) CNS tumor classification update in 2016, GB is classified as a WHO Grade IV tumor due to its histological and molecular features [2-4]. GB is the most lethal and invasive type of gliomas with highly aggressive cellular and molecular characteristics [5, 6]. GBs originate from neural stem cells (NSCs), NSC-derived astrocytes, and oligodendrocyte precursor cells and the tumor shows significant cellular and molecular heterogeneity [7]. GB also demonstrate aggressive and destructive features at the molecular level, such as diffuse infiltration pattern, mitotic activity, cellular pleomorphism, astrocytic differentiation, nuclear atypia, microvascular proliferation, and/or necrosis, with or without isocitrate dehydrogenase (*IDH*) gene mutations in addition to numerous morphological patterns [2, 3, 8].

IDH-mutant GB corresponds to ~10% of all GBs and is termed as 'secondary GB' as it develops by the progression of WHO Grade II diffuse astrocytoma or WHO Grade III anaplastic astrocytoma [2]. Morphologically, other than the lesser amount of necrosis, secondary GBs are histologically indistinguishable from *IDH*-wild type (wt) GB. *IDH* mutant GB is mainly observed in younger patients (average age of ~45) and has a remarkably better prognosis than *IDH*-wild type GB (Table 1.1) [2, 3]. *IDH*-wt GB corresponds to ~90% of all GBs and is termed as 'primary GB', as it naturally arises de novo, without any definite precursor lower grade tumors (i.e., Grade II or III). The incidence of *IDH*-wt GBs is common in relatively older patients, with an average age of ~62 and with a 1.35:1 male-to-female ratio. *IDH*-wt tumors are infiltrative and these tumors spread to neighboring or distant structures within the brain [3].

Table 1.1. Patient profiles of histologically diagnosed *IDH*-wild type (primary) and *IDH*-mutant (secondary) glioblastoma (GB) (Adapted from Louis et al., 2016) [3].

	IDH-wild type (primary) GB	**_IDH_-mutant GB (secondary)**
PTEN mutation	24%	5%
ATRX expression loss	0%	100%
TERT mutation	72%	26%
TP53 mutation	23%	74%
Loss of 19q	4%	32%
EGFR amplification	42%	4%
Age at diagnosis	59 years	43 years
Male-to-female ratio	1.4	1
Clinical history length	3.9 months	15.2 month

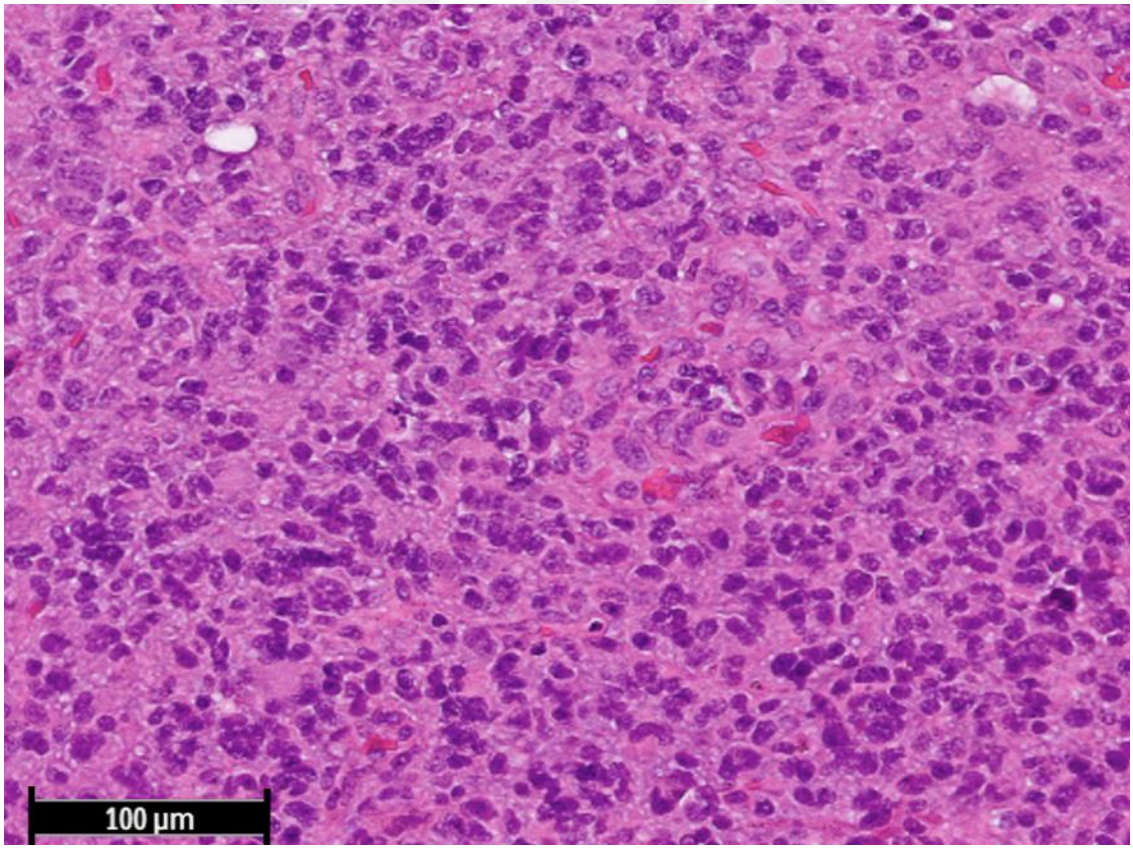


Figure 1.1. Hematoxylin and Eosin (H&E)-stained Glioblastoma (GB) tissue section, 5µm-thick. Image obtained from the University of California, San Francisco, Department of Pathology, Division of Neuropathology archival tissue blocks.

Despite the rapidly improving treatment modalities with current healthcare technologies, the prognosis of *IDH*-wt GB remains poor [3]. Histological and molecular variability of GBs requires expertise in neuropathology for accurate diagnosis and interpretation. Histological and morphological features are readily recognized using the standard hematoxylin and eosin (H&E) stain and light microscopy. H&E staining and light microscopy are known as the gold standard in the histopathological evaluation of GB (Figure 1.1). Despite the utility of this gold standard, recent studies suggest that detailed molecular tests are necessary for the identification of molecular alterations of the tumor before the final diagnosis can be established [9]. The recent updates in the classification of CNS tumors [10] demonstrated the molecular heterogeneity-related limitations in histological classification only and involved the genetic features and molecular markers into the neuropathological diagnosis criteria [6, 9, 11] (Table 1.2). While the diagnosis was carried out based on the histological patterns of H&E-stained tissue sections according to the preceding WHO classification [12], 2016 WHO updates in CNS tumors classification shifted the histomorphological evaluation to a different scheme based on the combination of morphological and molecular features. The essential classification parameters remained the same by classifying the tumors based on the constituents of the CNS cells, yet it is improved by consideration of the molecular features of these cells [9, 13]. These analyses often require methods such as immunohistochemistry, fluorescence in-situ hybridization (FISH), Sanger sequencing, or next-generation sequencing [5, 9]. The long examination time of molecular testing of tumor specimens and the subjective precision of histopathological evaluation, however, pose the most important challenge in the diagnosis of high-grade gliomas [14-16].

Table 1.2. Genetic Alterations of *IDH* wild-type Glioblastoma (GB) [3].

Genetic Alterations	**Percentage**
TERT promoter mutations	~80%
Homozygous deletion of *CDKN2A/CDKN2B*	~60%
Loss of chromosome 10p	~50%
Loss of chromosomes 10q	~70%
EGFR alterations	~55%
PTEN mutations or deletion	~40%
TP53 mutations	~25-30%
PI3K mutations	~25%

A high-resolution method that can analyze tumors histologically and analytically would be highly useful for further understanding of tumor biology. This method may improve diagnostic accuracy, save time, eliminate the use of excessive and hazardous reagents, and may help reduce expenses by assisting neuropathologists in the diagnostic examination of GB candidate tissues in the clinic.

1.1.1 Detection, Imaging, and Treatment Modalities of Glioblastoma (GB)

Despite the improvements in healthcare technologies, GB remained incurable with poor prognosis and complicated treatment due to the infiltration of the tumor [17, 18] to the surrounding brain parenchyma[18]. An overall median life expectancy of GB patients remained 15 months following the tumor resection and chemo/radiotherapy for the last decades [1, 5, 17, 19, 20].

Magnetic resonance imaging (MRI) is the most used technique in the detection of GB. Computed tomography (CT) is another modality of choice used in GB detection, which is mostly preferred for those patients who cannot tolerate MRI. Positron Emission Tomography (PET Scan) is preferred when anatomic imaging techniques (i.e. MRI or CT) are inadequate in differentiating remaining or recurrent tumors from treatment-related changes such as radiation necrosis or pseudo-progression. The PET scan also has limitations in the accuracy of interpretation of tumor boundaries [21]. Current treatment modalities centered on maximum safe surgical resection followed by adjuvant

chemo/radiotherapy [17, 22, 23]. Temozolomide and bevacizumab are the only U.S. Food and Drug Administration (FDA)–approved drugs for GB treatment [22].

In almost all patients with GB, the tumor has a recurrence within two years after the initial operation. The role of surgery has the utmost importance to determine the extent of resection which solely depends on the intraoperative perception of the neurosurgeon [24]. Maximum safe resection of the tumor is the most crucial factor for therapy management of patients with GB, in terms of decreasing the tumor burden before chemotherapy and radiotherapy, improving life expectancy [17], reducing the risk of recurrence, and achieving a better outcome for the patients [18]. The lower extent resections of non-eloquent lesions might increase the risk of worsening in the course of the disease [25]. One of the most conjectural prognostic factors of survival after surgery is a residual tumor for GB patients. One such study demonstrated that GB patients with residue tumors had ~6.6 times more mortality risk than the patients without residing tumors [24]. As the healthy brain tissue and the boundaries of the tumor cannot be distinguished by the human eye, the tumor boundaries are identified by the present pre and post-operative techniques such as a standard surgical microscope, intraoperative MRI (iMRI), ultrasound (US), fluorescence-guided microsurgery, and optical coherence tomography (OCT) [18].

Functional neuro-navigation is a technique preferred in most eloquent lesions based on ultrasound, pre-operative functional magnetic resonance imaging (fMRI) images, or diffusion tensor imaging. However, pre-operative MRI imaging is limited due to the existence of infiltrated tumor cells [25] that are away from the contrast-enhancing regions, as well as during microsurgery-assisted resection of gliomas. Moreover, intraoperative MRI imaging does not provide continuous guidance [18]. Despite the accessibility and availability of intraoperative ultrasound (IOUS) in comparison with the other intraoperative imaging techniques, the use of 3D-IOUS during tumor resection is considerably affected by the quality of ultrasound images, and the worth of this technique is diminished in the existence of edema and hemorrhage in the tumor and tumor environment [18, 25].

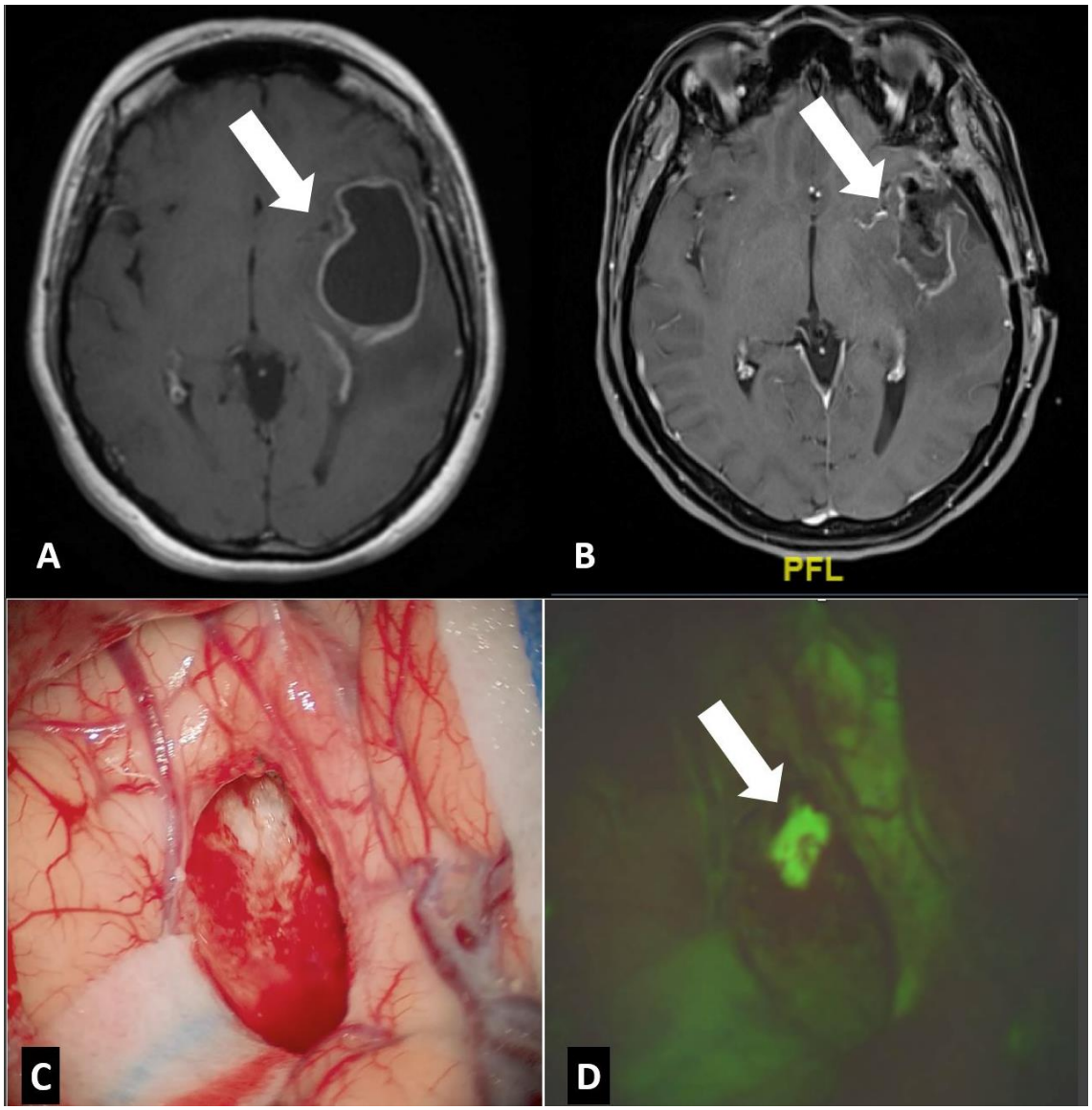


Figure 1.2 (A) Preoperative axial T1-weighted image after a contrast administration revealed a left temporal cystic tumor (arrow). (B) postoperative axial T1-weighted image after a contrast administration shows total removal of the tumor. (C) Intraoperative resection cavity view under white light. (D) At the same time, the same region presented high fluorescence yellow color (arrow) under the YELLOW 560 filter, shows the area of the remnant tumor that needed to be resected (Pictures were obtained from the archives of Prof. Ihsan Solaroglu).

Fluorescence-guided surgery has emerged as an advanced adjunctive technique, which had demonstrated an increase in the extent of tumor resection [26]. 5-aminolevulinic acid (5-ALA) is an agent that induces the production of fluorescent porphyrins which is selectively synthesized in malignant tissues after 20 mg administration per kg and provides detection of malignant gliomas to surgeon under the

xenon-light microscope during the surgical intervention [27, 28]. Due to the limitations of currently available techniques, sodium fluorescein-guided surgery recently gained more attention from neurosurgeons [26]. When combined with T1 weighted contrast-enhancement neuronavigation, sodium fluorescein-guided surgery was proved to be safe and effective in the resection of GBs (Figure 1.2). Despite the usefulness of 5-ALA as a supportive technique in identifying tumor boundaries that are not contrast-enhanced areas on T1 MRI [29], some research reported false-positive fluorescence enhancement in the non-neoplastic brain parenchyma [30, 31]. Moreover, it has several side effects [32] and requires patient consent prior to administration. Removal of the healthy tissue near the tumor environment during the resection of tumors located in functional areas of the brain or adjacent to the brain poses risk for neurological sequelae.

To prevent this risk, there is a need for an intraoperative technology that can distinguish normal and tumor tissue at high resolution during surgical intervention.

1.2 Raman Spectroscopy (RS)

Raman spectroscopy (RS) is a promising tool and could be used to develop technologies either for developing a digital pathology method to analyze pathology archival Formalin-Fixed Paraffin-Embedded (FFPE) tissue sections or fresh tissue as it is not affected by the water content and provides information on the chemical content of the given sample.

When a coherent light beam is exposed on a surface, photons scatter elastically and inelastically. In elastically scattered photons, also known as Rayleigh scattering, photons' kinetic energy is conserved, and the wavelength of the scattered photon remains the same as that of the incident photon. On the other hand, a rare fraction (10^{-6} or less) of the incident photons scatter from the surface inelastically. One of the forms of inelastic photon scattering is Raman scattering, in which the wavelengths of the incident and the re-emitted photons are different (Raman shift). There are two types of Raman shifts: Stokes and anti-Stokes shifts. In Stokes shift, the emitted photon is of lower energy (longer wavelength) than that of the incident light. In the anti-Stokes shift, the emitted photon has higher energy than that of the incident photon due to the molecules losing energy to the emitted photons (Figures 1.3-1.5). In this thesis, we focus only on Stokes

shifted (red-shifted) photons, which have positive Raman shift wavenumbers (cm^{-1}) since the brain tissue sections absorb photons [33, 34].

RS is a reagent and label-free, non-destructive, and highly specific analytical technique based on detecting the inelastically scattered photons. RS provides the molecular bond information of a sample under study. Raman spectral peaks (or shifts) originate from the scattering of photons by vibrational modes and their corresponding vibrational resonance frequencies of chemical bonds. Raman is a molecule and bond-specific method and highly useful in composition analysis of the samples [33, 34].

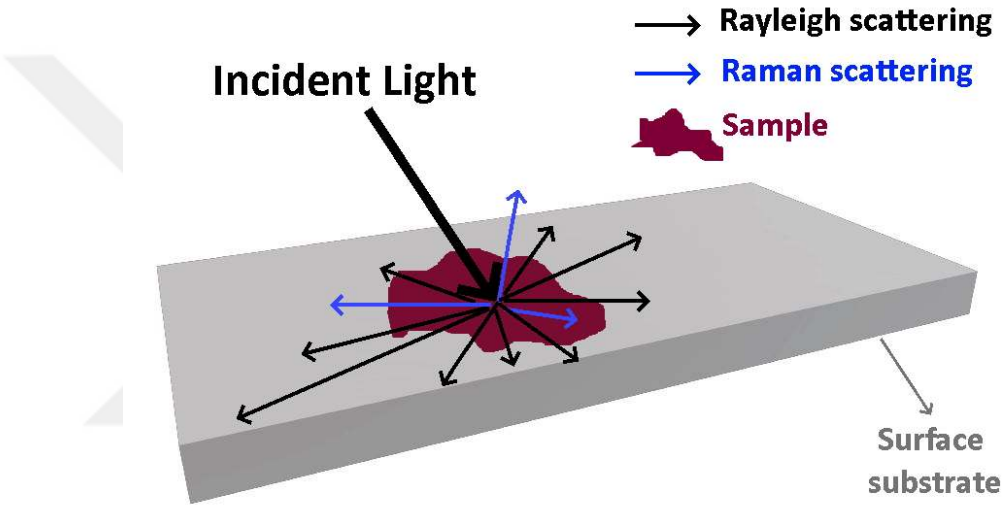


Figure 1.3 Raman and Rayleigh Scattering. Rayleigh scattered photons have the same wavelength as the incident light, Raman scattered photons (1/1.000.000 incidence) have a different wavelength from the incident light. Adopted from Egeland, N., 2015 [35].

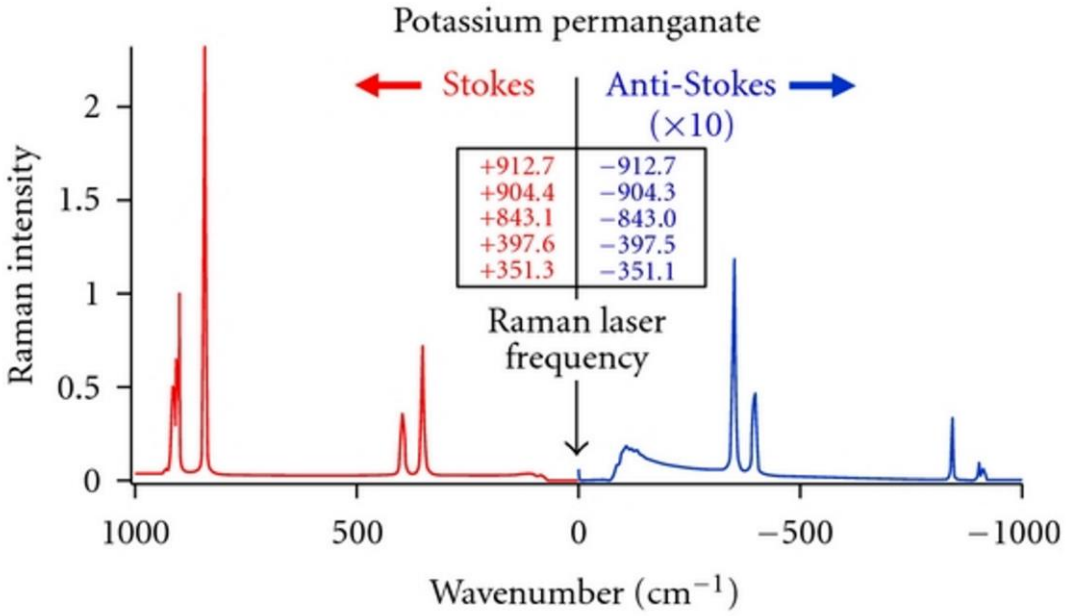


Figure 1.4. Representation of Stokes-Antistokes Shifts. Reprinted from Copyright © 2012 Timothy J. Johnson et al. under the Creative Commons Attribution License [36].

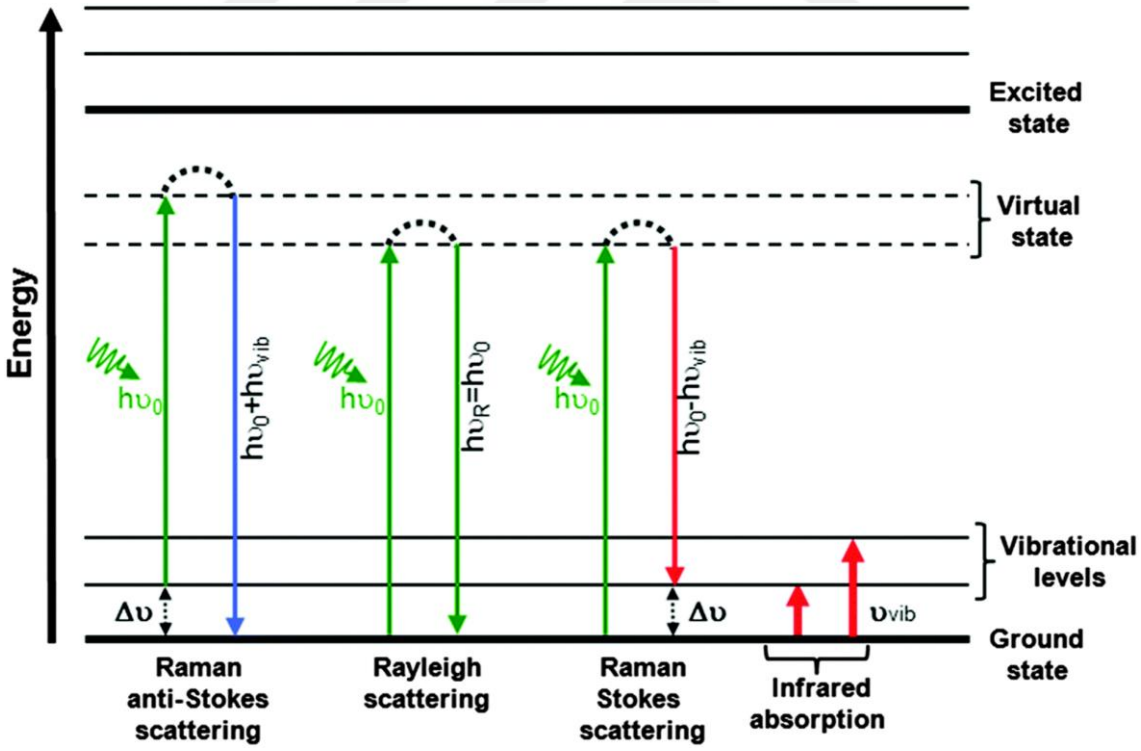


Figure 1.5. Jablonski diagram represents the vibrational energy states of infrared absorption, Rayleigh, Raman Stokes, and anti-Stokes scattering. $h\nu_o$ = incident laser energy, $h\nu_{vib}$ = vibrational energy, $\Delta\nu$ = Raman shift, ν_{vib} = vibrational frequency. Reprinted from Baker, Matthew J., et al., (2016) [37], with the permission of Copyright Clearance Center.

The key competence that enables this technology to distinguish tumors is being able to read the differences in the chemical content of normal and tumor cells by optical methods. Without requiring additional chemicals, a rapid and precise analytical technique is needed in assisting neuropathologists for error-free diagnosis of GB.

1.3 LITERATURE REVIEW

1.3.1 Raman Spectroscopy in Biomedical Applications

RS and microscopy are emerging techniques for characterization of composition analysis of biological materials as it provides highly specific chemical bonding information of a given molecule. RS is growing among biomedical applications as it rapidly offers fingerprint information of the biochemical composition without the use of additional reagents. RS has previously been used in biomarker and cancer research and tissue composition analysis for different diseases [34, 38-40] (Table 1.3). RS has previously been used as a tool in the research for bladder and prostate [41, 42], laryngeal [43, 44], breast [45, 46], lung [47, 48], cervical [49, 50], skin [51], stomach [52, 53], pancreatic [54], colon [55], brain cancers [13, 15, 56-61] and leukemia [62, 63].

Table 1.3. Application of Raman Spectroscopy for Various Cancer Types

Disease	**Biological material of interest & Data analysis method & References**	
Bladder and Prostate Cancer	Snap-frozen bladder and prostate tissue Principal components analysis (PCA)/ Linear Discriminant Analysis (LDA) [41]	Bladder and prostate cancer cell lines PCA [42]
Laryngeal Cancer	Fresh laryngeal tissue PCA/LDA + Savitzky Golay Filter [43]	Fresh laryngeal tissue PCA/LDA [44]
Breast Cancer	Fresh breast tissue	Fresh breast tissue

	PCA/LDA [45]	PCA [46]
Lung Cancer	Fresh bronchial tissue PCA [47]	Fresh tissue [48]
Cervical Cancer	Fresh cervix tissue maximum representation and discrimination feature, sparse multinomial logistic regression [49]	Pap-smear cervix Savitzky-Golay Filter [50]
Skin Cancer	Fresh skin tissue maximum representation and discrimination feature, sparse multinomial logistic regression [51]	
Stomach Cancer	Fresh gastric tissue PCA, multinomial logistic regression [52]	Fresh gastric tissue (w/probes) ant colony optimization (ACO), PCA/LDA [53]
Pancreatic Cancer	Fresh pancreatic tissue from a mouse model PCA and discriminant function analysis (DFA) [54]	
Colon Cancer	Fresh colon tissue Cluster analysis [55]	
Leukemia	ATCC cancer cell lines, normal lymphocytes, paraformaldehyde, and methanol-fixed cells PCA [62]	Normal human peripheral blood mononuclear cells (PBMCs) and leukemia cells PCA/LDA [63]
Brain Cancer	Frozen brain tissue PCA/DFA [59, 64] Fresh brain tissue PCA/LDA [65]	Fresh, cryosection, and Formalin-Fixed Paraffin-Embedded (FFPE) brain tissue PCA/LDA [56]

RS and microscopy have been used in the analysis of various tissues and tumors in recent years [39, 66]. Raman studies also include fresh brain tissue analysis or intraoperative frozen process of brain tumors for diagnosis. These studies also aimed at identifying specific tissue characteristics such as white and gray matter based on their lipid, nucleic acid, and protein contents [59, 64]. Raman spectroscopy and microscopy are also used for determining the brain tumor boundaries intraoperatively, and for aiding in brain tumor classification [67-69]. Others have employed RS for in-vivo and ex-vivo diagnosis of brain tumors in mouse and human xenografts [68-70]. Some studies attempted to develop algorithms to detect EGFR gene mutation/amplification status using Raman spectral analysis [71, 72].

The recent advances, the increasing number of recent Raman publications, and our initial experimental observations all suggest that Raman spectroscopic analysis can be a powerful tool for the diagnosis and characterization of human tissues. Many studies also showed that normal and tumor tissues can be distinguished, and molecular characteristics of tissues can also be determined with the use of this technique [59, 64]. When coupled with the current golden standard as well as additional techniques using different characteristics of light in different parts of the electromagnetic spectrum, analyzing tissues optically might be possible without the need for additional sophisticated molecular and genetic tests.

To obtain better quality Raman spectra for better analysis of tissue compositions, we investigated the key factors such as Raman substrate (i.e. glass, borosilicate, CaF_2), tissue type (i.e. fresh, frozen, or FFPE), sample preparation (i.e. deparaffinization protocols), and Raman measurement (excitation wavelength, laser acquisition time, laser accumulation count) conditions and improved these parameters concerning the earlier studies.

Sample Preparation and Substrate Choice for Raman Spectroscopy

For better experimental analysis of tissue samples and eliminating the overwhelming paraffin background in Raman spectra, sample preparation and paraffin clearing protocols must be developed. The importance of tissue clearing methods has been highlighted, as the challenge of deparaffinization of FFPE tissue sections increases due to their long-term storage [73]. Sample preparation conditions for RS for biological materials in different forms (fresh, snap-frozen, fixed, and liquid samples) were extensively reviewed by Butler

et. Al., 2016 pointing the following important points in the use of FFPE tissues. While one may analyze fresh tissues directly without processing, obtaining thin sections from FFPE tissues with desired thicknesses is easier. Despite this advantage of FFPE tissues, the deparaffinization step creates challenges as the paraffin or formalin residuals can substantially overlap with the specific Raman shifts of the sample. Paraffin or formalin residuals must therefore be removed effectively from tissues without altering tissue composition. Nonetheless, fixation can alter the tissue composition by creating cross-linking in the proteins and this may alter the protein-originated peaks in the Raman spectral range between 1500 and 1700 cm^{-1} [39]. Several Xylene [39, 74, 75], hexane, and histoclear [75]-based deparaffinization protocols for RS have been tested in the literature. Although these chemical clearing agents such as xylene, hexane, and histoclear reduced the paraffin, they could not eliminate the presence of paraffin residuals. Also, the tissues could be preserved in the clearing agent for up to 18 hours long for the optimized protocols [74, 75]. As prominent paraffin spectra interfere with the Raman spectral analysis of tissues, Ibrahim, O., et al. developed a digital deparaffinization protocol by subtracting the paraffin spectra from the tissue spectra using a least-squares analysis and non-negative constraints.

Ibrahim, O., et al., 2017 have emphasized the effect of glass slide substrate use in spectral analysis of tissues in medicine. Besides, the use of glass substrates in Raman spectroscopic analysis of biological samples with different excitation wavelengths (532 and 785 nm) were evaluated in addition to the natural substrates such as CaF_2 and magnesium fluoride. The background of the glass substrate is reported to be higher when compared to these natural substances. These natural substrates have higher costs than glass. It is also reported that the fluorescence contribution of glass substrates is significantly lower with 532 nm laser excitation wavelength when compared to the 785 nm [74]. Therefore, substrate and wavelength choice should be considered together to obtain better results, especially in biomedical research. Lewis, Aaran T., et al. evaluated the use of CaF_2, stainless-steel, quartz, and standard glass substrates and pointed that CaF_2 is the most widely used Raman substrate. The authors provided background absorption (except the characteristic peak at 321 cm^{-1}) and the transmission of visible and near-infrared light without high losses. Quartz and standard glasses were reported as mostly preferred in Surface Enhanced Raman Spectroscopy (SERS) and coherent anti-stokes RS studies as they have higher backgrounds than CaF_2. Despite the advantages of CaF_2 over

quartz and glass substrates, stainless steel substrates were reported as more practical for clinical use as they are cost-effective and provide high SNR and spatial resolution if the Raman device is adapted to the biomedical sample measurements [76]. A previous study [77] concluded that CaF_2 Raman substrates give the most consistent backgrounds for all excitation wavelengths despite their high cost. CaF_2 is increasingly used by researchers as a low-background Raman substrate [78-80].

Excitation Wavelength and Measurement Parameters for Raman Spectroscopy

A previous study [77] investigated some of the trade-offs in the laser excitation wavelength choice and the Raman substrate combinations for RS. Kerr, L. T. et. Al, 2015 evaluated the use of 473, 532, 660, 785, and 830 nm excitation wavelengths with different Raman substrates such as Raman-grade CaF_2, IR polished CaF_2, magnesium fluoride, standard glass, zinc selenide, sodium chloride, potassium bromide, fused silica while keeping the other experimental parameters constant as possible [77].

Butler et. al, 2016 have broadly examined the factors that affect the measurements such as spectral resolution, wavelength choice, fluorescent background, the numerical aperture of the objective, to avoid causing photoablation (sample burning) or saturation of the detector. In this study, the fluorescent background is reported to be susceptible to the background when 532 nm laser is used and is relatively free when high-energy (UV) and lower-energy IR wavelengths are used (i.e. 244 and 1064 nm). The most commonly used lasers are 786 and 830 nm (NIR region lasers) in biomedical research as they generally do not cause photodamage in the samples [39]. Although it is dependent on the sample, the laser acquisition time is defined to be sufficient when shorter than 10 seconds for quick measurements. It is also highlighted that the spectral quality relies on the instrumentation, sample type and suitability, laser power, laser acquisition time, and laser accumulation time.

Raman Spectral Data Processing and Classification

Recent studies have shown that cancer diagnostics are possible using RS and machine learning. Classification models are used together with RS for tissue identification and tumor classification [38, 81]. Gajjar et al. demonstrated that RS is a potential diagnostic approach in the accurate and intra-operative diagnosis of brain tumors, distinguishing the

GB and normal brain using PCA and LDA [15]. Koljenovic et. al. also developed an LDA-based prediction model for discriminating vital tumor tissue from necrotic tissue in frozen human glioblastoma samples and yielded 100% accuracy [82]. Kast et al. studied to identify the regions of the normal brain from glioblastoma and necrosis in frozen brain tissue sections using RS by creating Raman images [64]. Recently, Zhou et al discriminated against the normal brain, low grade (I-II), and high grade (III-IV) gliomas using visible resonance Raman VRR Spectroscopy with 75.1% total accuracy using PCA and Support Vector Machine (SVM) for classification [13]. In another study, Parlatan et al. applied RS to diagnose endometriosis using PCA with the SVM and k-Nearest Neighbors (k-NN) and yielded 100% specificity and sensitivity with unseen testing data [81]. Orringer et al. used stimulated Raman scattering microscopy in the operation room and acquired stimulated Raman histology images. Using multilayer perceptron-based image processing and by providing intraoperative pathologic consultation for brain tumor detection, the authors reported accuracies around 92% [60].

Developing an advanced RS diagnostic tool might allow for advanced analyses of tissues without using additional techniques or reagents. Our study aims to develop RS and microscopy as a viable adjunct to the current standard of light microscopy providing more advanced, comprehensive, and detailed analyses of GB.

1.4 HYPOTHESIS AND SPECIFIC AIMS:

1.4.1 HYPOTHESIS:

Raman spectroscopic techniques and machine learning analysis can provide additional actionable molecular level information in morphological analysis of tissue sections to further our understanding of tumor biology.

This hypothesis will be tested through the comparative analysis of normal brain tissue and a molecularly uniform set of *IDH*-wt GB samples using RS. With this approach, we aim to obtain further information beyond that gained through the study of routine pathological evaluation. This thesis project specifically focuses on understanding the Raman spectroscopic features of normal constituents of brain tissue and the differences of these spectroscopic features from a group of *IDH*-wt GBs.

1.4.2 SPECIFIC AIMS:

SPECIFIC AIM 1: To define Raman spectroscopic characteristics of FFPE normal brain tissue and major tissue types of the central nervous system, to construct a baseline for tissue diagnosis (See: Chapters 2 and 3).

Aim-1A. (OPTIMIZATION) The most abundant and available archival and routine material for brain tissues is found in the form of FFPE tissues. Since FFPE samples can provide repeated analysis of the same tissues by multiple methods (such as light microscopy and RS), we selected to use FFPE tissues as our material in this study. In AIM 1A, we aimed to develop a methodology that can generate spectroscopic data from this material. Establishing the methodology for FFPE allows us to use archival tissue and generate data that can be compared to the light microscopic information. For this aim, we collected paraffin blocks of tissues from the University of California, San Francisco (UCSF), Koç University Hospital (KUH) Department of Pathology, and Gazi University Hospital (GUH) as samples for our study. This stage also includes the optimization of parameters such as tissue thickness to maximize the signal-to-noise ratio (SNR) for precise Raman spectroscopic analysis of CNS tissue. Some of the parameters for

optimization involve the choice of excitation wavelength, accumulation time, acquisition conditions, laser power, and magnification.

First, we developed an optimal deparaffinization protocol for the efficient and effective removal of paraffin from FFPE sections to create a reliable Raman signal acquisition while retaining the morphological and chemical composition of the tissue samples. We determined the effectiveness of deparaffinization protocols and validate our protocol.

Second, we optimized the spectral analysis of CNS tissue by modifying analytical parameters such as tissue thickness, Raman substrate choice, spectral excitation wavelength parameters, magnification, scanning time, etc. (See: Chapter 2).

Aim-1B. (CHARACTERIZATION) We aimed to test these optimal parameters and generate specific signals from normal versus abnormal tissue types of CNS (i.e. GB vs white matter) for allowing the best molecular composition analysis by RS, once the Raman data acquisition parameters are optimized. (See: Chapter 3).

SPECIFIC AIM 2: To define Raman spectroscopic characteristics of *IDH*-wt GBs and compare them with morphological and molecular composition of tumor tissues in routine sections. The experiments in this phase aimed to develop characteristic profiles of tissue types such as gray matter, white matter, necrosis, and the tumor. Determine the distinguishing spectral features of GBs regarding the normal constituents of CNS (See Chapters 4 and 5).

Aim-2A**.** Using a genetically and pathologically uniform group of tumors (*IDH*-wt GBs), we expect to characterize cytological and histological features that can be compared with the normal constituents of CNS tissue. Specifically, we aimed to develop specific spectral profiles for these components of GBs such as necrosis, tumor, white matter, and gray matter. (See: Chapter 4)

Aim2B. The aim of this stage is to develop a digital pathology method as a proof-of-concept from the spectral database of tissue profiles that can be used as a reference base for analyzing samples to determine their composition and nature (neoplastic versus non-neoplastic). After collecting the spectral profiles for different tissue types (tumor,

necrosis, white matter, and gray matter), we aimed to classify these tissue groups using several machine learning algorithms. (See: Chapter 5)

1.5 Anticipated Results, Novelty, and Impact

- We anticipate being able to determine morphologic features of tissues in normal and tumor tissues using routine staining methods and construct the Raman spectroscopic profiles of the same tissues using FFPE samples. This will allow us to construct a database containing various qualities of types and normal and abnormal morphological features in tissues.
- The Raman spectral profiles are expected to identify the nature and qualities of the tissues without disturbing the tissue integrity and without using any chemicals or pigments for staining.
- The integration of additional spectral data on the routine sections is expected to provide more information about the cells and tissues beyond that provided by the routine staining methods such as H&E and might complement the information that could be obtained with the use of immunohistochemical or genetic techniques.
- These results could give us the ability to construct a database that can eventually be used in the clinic to determine the nature of a particular tissue without the need for any further tissue processing or staining.

Chapter 2:
MATERIALS & METHODS

In this thesis study, we analyzed FFPE brain tissue sections using RS and microscopy to identify tissue compositions and distinguish normal and abnormal tissue regions in patient samples with GB. In this chapter, we first explain the tissue preparation methods for RS measurements. Next, we present our pointwise and map Raman spectral data acquisition techniques. Finally, we present our data analysis techniques for the classification of tumor samples annotated by pathology specialist, Dr. İbrahim Kulaç.

2.1 Sample Size Calculation

Sample size calculation for this study was carried out by Biostatistician Dr. Arzu Baygül from Koç University, as described by Bujang et al [83].

The number of samples to be included in the study was determined as 24 for Sensitivity and 24 for specificity. Accordingly, considering the upper limit, approximately 30 people should be included in the study, considering a 20% loss suffix.

2.2 Material Transfer Agreement (MTA), Institutional Review Board (IRB), and Committee of Human Research (CHR) Approvals:

This study was conducted under the Koc University Institutional Review Board (IRB) approval number 2018.082.IRB1.011 and the University of California, San Francisco (UCSF) Committee of Human Research (CHR) approval number 17-23351.

This study was conducted under the ‘Material Transfer Agreement (MTA)’ between Koc University Research Center for Translational Medicine (KUTTAM) and the Department of Neuropathology, University of California, San Francisco. This study was conducted under the MTA between KUTTAM and the Department of Pathology, Gazi University Hospital (GUH), Ankara.

2.3 Pathological Evaluation and Diagnosis:

H&E staining and analysis of normal and tumor tissues and whole slide imaging by neuropathologists H&E staining were performed according to established and validated protocols [68]. An Aperio ScanScope XT system with 150 slide capacity (Leica Biosystems, Buffalo Grove, IL) was used and the digital images were obtained using publicly available software (Image Viewer, Aperio Technologies, Leica Biosystems Buffalo Grove IL, USA). The digital images were shared between the institutions for consensus assessment.

2.4 Material Selection

11 GB FFPE brain tissue blocks from 11 GB patients were transferred for this thesis study from the Division of Neuropathology, Department of Pathology at the UCSF, according to the MTA between Koç University and UCSF Neuropathology. 44 GB FFPE brain tissue blocks from 23 GB patients supplied from Koç University Hospital, Department of Pathology. 8 GB FFPE brain tissue blocks from 8 GB patients transferred from Gazi University Hospital (GUH), Department of Pathology, according to the MTA between Koc University and GUH Department of Pathology. We employed a total of 38 GB, 18 WM, 19 necrosis (NC), and 16 gray matter (GM) samples from 42 patients for this study (n=42).

The tissue samples were selected among previously diagnosed excision biopsies from the archives of KUH Pathology, GUH Pathology, and the UCSF Neuropathology.

One *IDH*-wild type WHO Grade IV GB tissue block was included in Raman optimization and 5 GB tissue blocks were included in deparaffinization protocol optimization studies. *IDH* mutant or lower grade glial tumors were not included in this study. 4 tissue blocks obtained from 2 recurrent GB patients were excluded from this study. Excess or residual tissues from diagnostic or surgical procedures were used in this study. In this study, we focused on the classification of GB, WM, GM, and NC tissue spectra using machine learning models.

2.5 *Sample Preparation*

2.5.1 Tissue Preparation for Histopathological Diagnostics:

Tissue samples were sectioned in 5µm-thick and mounted on positively charged glass slides (Thermo Scientific™ Polysine Adhesion Slides) by microtome (Leica Biosystems, RM2245). Sections were undergone deparaffinization and H&E staining steps for histopathological analysis using Tissue-Tek Prisma (Sakura Finetek, CA, USA) device according to the standard operation procedures of UCSF and KUH pathology laboratories. Histopathological diagnosis was carried out by two Neuropathologists from KUH and the UCSF (Dr. İbrahim Kulaç and Dr. Tarik Tihan, respectively) according to 2016 WHO CNS Tumor Classification [2, 3, 9, 10].

Two tissue slices were sectioned from the same tissue blocks: One was used for histopathological diagnostics, the other was used for Raman spectral measurements. The tumor and white matter regions on the sections were marked by a neuropathologist on an H&E-stained slide to be used as a guide for spectral measurements. Upon the completion of Raman measurements, the sections that were used in Raman measurements have undergone histopathological analysis for double-checking (Figure 2.1).

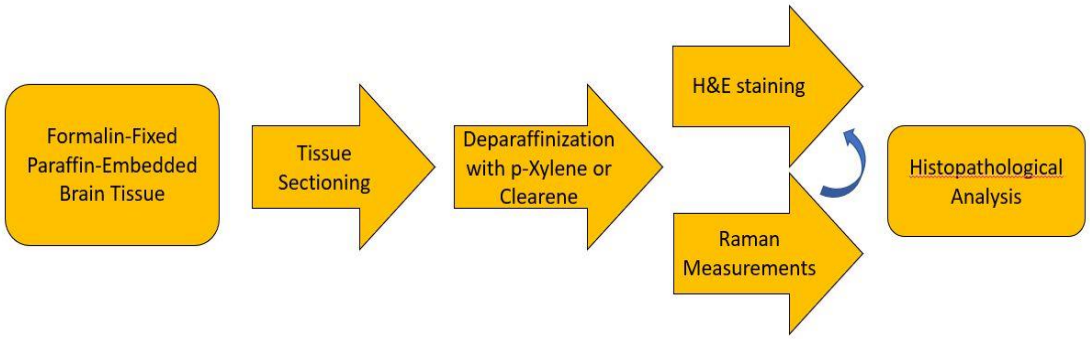


Figure 2.1. Workflow Diagram of FFPE Tissue Sections

2.5.2 Tissue Preparation for Deparaffinization Protocol Optimization:

Two FFPE tissue blocks were used in the development and optimization of deparaffinization protocols. Two blocks were sectioned in 5µm-thick onto glass slides in two sets, one for Clearene™[84] and another one for p-Xylene [85] deparaffinization.

The deparaffinization protocols were developed by applying several modifications to the protocols [39, 86].

Tissue sections were deparaffinized according to the following protocols:

Tissue sections were immersed in the cuvette filled with p-Xylene (p-Xylene for synthesis, Sigma-Aldrich, Merck Milipore) for 5 minutes at room temperature. This step was repeated twice to ensure the paraffin was dissolved effectively and replaced with xylene. Then, the tissue sections were immersed into the next cuvette filled with the Ethanol Absolute (99.9% by vol.) (ISOLAB 920.026) for 5 minutes to replace xylene and paraffin residuals with absolute ethanol. This step was repeated three times to ensure the effective clearing of paraffin and xylene residuals from the tissue sections. The sections were then immersed into the cuvette filled with 90% (by vol.) ethanol for 5 minutes to replace the alcohol with a less dilute one. After repeating this step 3 times, the tissue sections have been immersed into another cuvette filled with 70% (by vol.) ethanol for 5 minutes to remove the alcohol from the tissue sections by reducing the concentration. After repeating the last alcohol wash, the tissue was immersed into the cuvette filled with distilled water and held for 5 minutes once to remove the alcohol from the tissue section and supplemented water that the tissue sections lost during the paraffinization process.

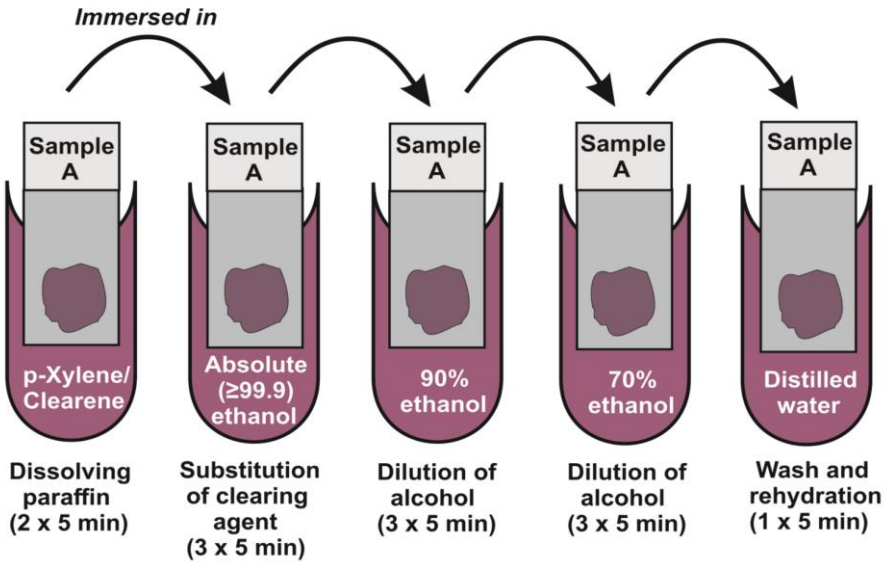


Figure 2.2. Deparaffinization Protocols w/p-Xylene or Clearene™ Solvents

The second deparaffinization protocol was applied by replacing the clearing agent p-Xylene [85] with Clearene^{TM} Solvent (Leica Biosystems, 3803600) [84] in the previously described protocol (Figure 2.2). Sections were then stained using the Tissue-Tek Prisma device for morphologic analysis under the light microscope.

2.5.3 Tissue Preparation for Tissue Thickness, Substrate, and Raman Measurement Optimization:

One GB tissue block was included in thickness, substrate, and Raman measurement optimization studies. One GB tissue block was determined as an optimization tissue block and was sectioned in 5µm and 10 µm-thick by microtome as two sets to be mounted onto positively charged glass and Calcium Fluoride (CaF_2) Raman substrates (CAFP76-26-1U, Crystran, U.K.).

2.5.4 Tissue Preparation for Raman Spectral Mapping

Tissue samples for spectral measurements were sectioned in 10µm-thick (x2) and mounted onto CaF_2 Raman substrates at Koç University Hospital, Department of Pathology.

Tissue sections were deparaffinized using Clearene^{TM} Solvent [84] according to the given protocol in Figure 2.2 and underwent the Raman spectral measurements after the sections were dried.

2.5.5 Sample Preparation for Tissue Microarray (TMA) Construction

38 FFPE tissue blocks from previously diagnosed 25 IDH-wt GB patients that were supplied from UCSF, Department of Pathology, Division of Neuropathology were used in Tissue Microarray (TMA) construction but only 11 FFPE tissue blocks in the TMA from 11 GB patients were included in this thesis study, the rest will be used in the future studies. All tissue blocks were sectioned onto positively charged glass slides as 5µm-thick and H&E stained at UCSF Neuropathology. Each H&E-stained section was analyzed by a board-certified Neuropathologist under the light microscope as a Gold standard and different tissue types (i.e., white matter, gray matter, necrosis, tumor) were marked onto

the slide with color codes (Figure 2.3). The slides were overlapped with the FFPE blocks and the same tissue regions were annotated for TMA preparation.

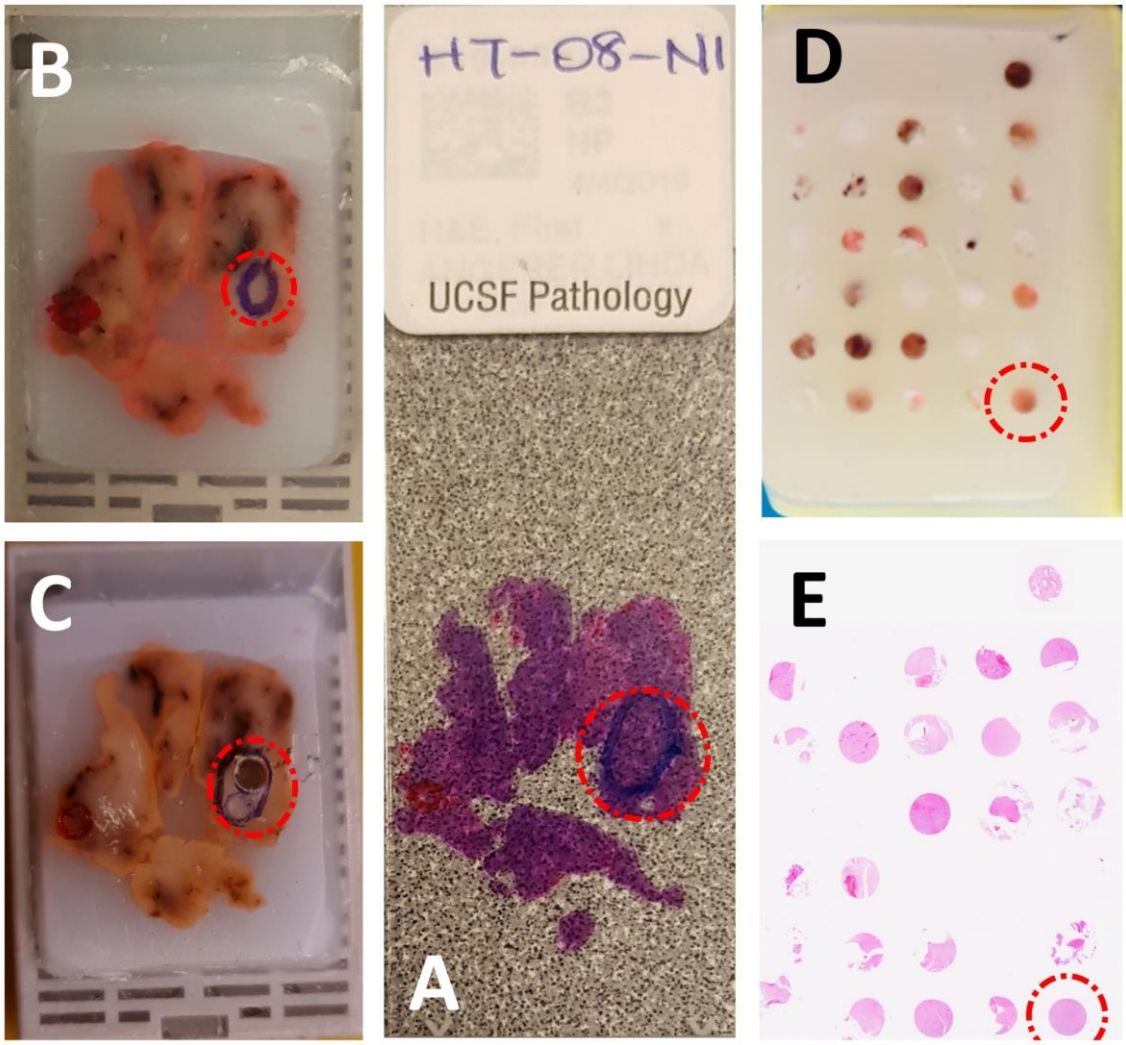


Figure 2.3. Tissue MicroArray (TMA) Preparation (A) H&E-Stained section was analyzed under the light microscope and the tumor region was marked with a blue pen on the slide. (B) The marked region was overlapped with the tissue block and the same region was marked on the tissue block. (C) The tumor region was removed using a punch needle via the TMA Master device and transferred into the TMA block (D). TMA block was sectioned, and H&E stained for histopathological analysis. (Red circles represent tumor regions)

120 regions including white matter, gray matter, necrosis, and the tumor that were marked on 38 FFPE tissue blocks and 4 blocks were placed into 3DHistech TMA Master device at a time. An array was prepared for constructing each tissue microarray (TMA) block. Four intact paraffin blocks were produced as TMA blocks and 31 holes (5x6 for samples and 1 control) were drilled onto these blocks using a 1.5 mm diameter needle

using the TMA Master. Drilled TMA block placed into the device with 4 other brain tissue blocks each time and height analysis were done. Tissue blocks were scanned and previously marked regions on the blocks were selected on the software. Regions of interest were transferred into the related previously drilled holes in the TMA block. (Figure 2.3) Each block was prepared involving 30 brain tissue cores and one lung tissue core as a control and annotating starting point. (Figure 2.3) Tissue microarray block (Figure 2.4.A) sectioned as sectioned in 5µm for histopathological analysis (Figure 2.4.C) and 10 µm-thick for Raman measurements (Figure 2.4.B)

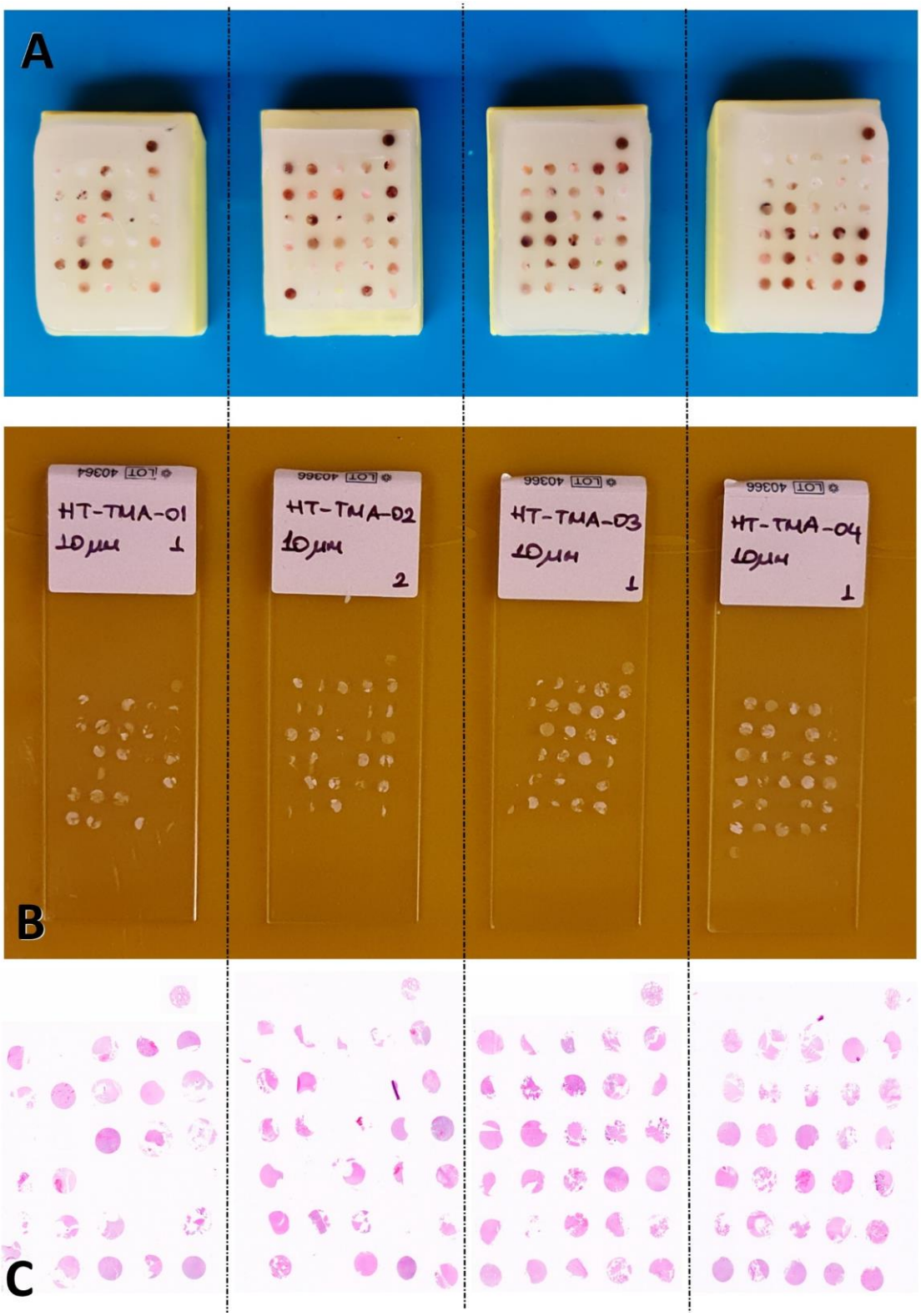


Figure 2.4. Tissue MicroArray (TMA) Block (A), Raman Section (B), and H&E-Stained Section (C)

2.6 Raman Spectral Measurements

All Raman spectra were acquired with a Renishaw inVia^{TM} Raman Microscope at TUBITAK MAM (Figure 2.5) using a 50x objective (Leica 50x/0.75). The device was calibrated for each measurement setup according to user instructions given in the manual provided by Renishaw.

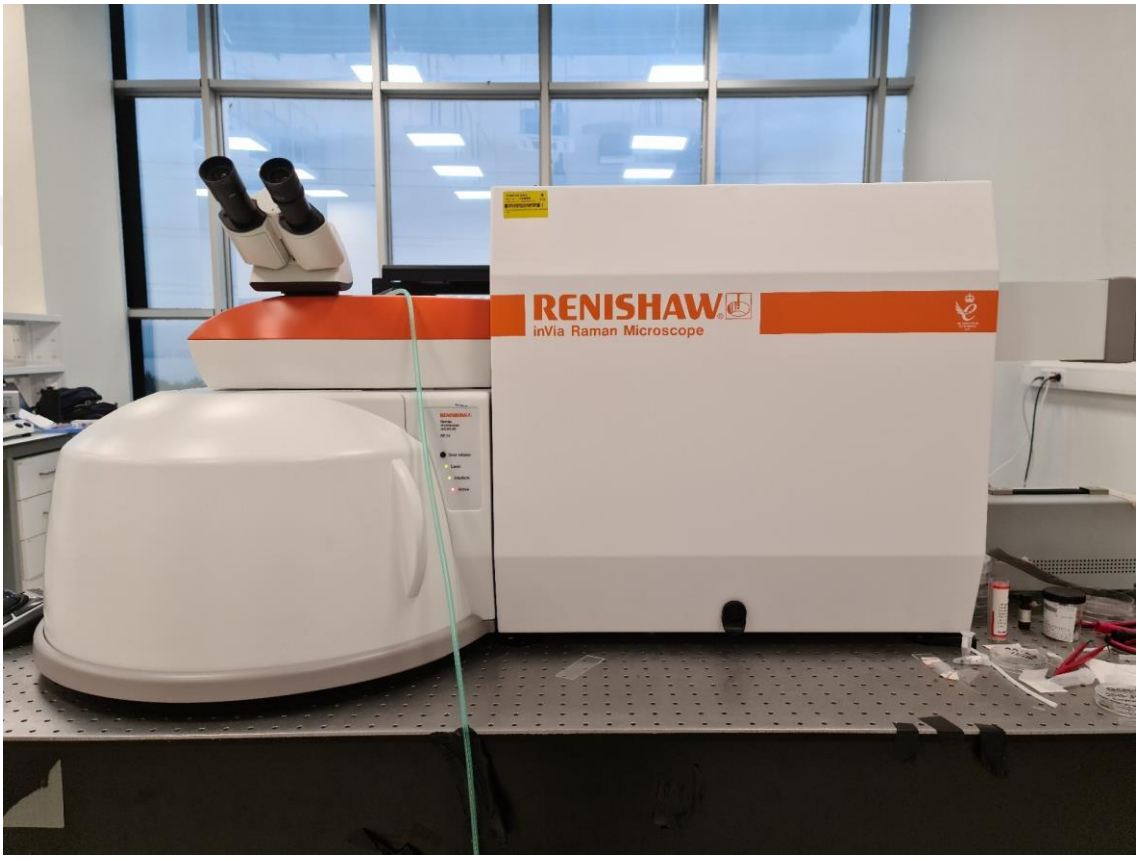


Figure 2.5. Renishaw inVia^{TM} Reflex Raman Microscope

2.6.1 Pointwise Raman Spectral Acquisition

Raman spectra of the sections were recorded using 532, 633, and 785 nm wavelength lasers under various acquisition times (1, 10, 20, 30, 40, 50 sec), accumulation counts (1, 3, 5, 10, 20), and laser power fractions (1, 10, 50, and 100%) for the optimization of Raman spectra acquisition parameters. Single point Raman spectra of previously annotated GB and white matter regions were recorded from each region with three repetitions to remove the cosmic rays using the Wire 4.3 Software of the instrument.

Raman spectra from at least one location point were recorded using previously given acquisition parameters for each excitation wavelength.

2.6.2 Spatial Map Acquisition of Raman Spectra

Raman spectra were recorded using the Renishaw Invia Raman Microscope after calibration was done for each measurement setup. 50x objective (Leica 50x/0.75) and 785 nm laser were used for mapping acquisitions. Previously optimized parameters (10 seconds acquisition, 100% power, and 1 accumulation count) were applied for map acquisition after the white light image was acquired for each region of interest or each core on TMA sections. Origin was set for each slide, calibration values, and the coordinates of each measurement were recorded for reproducible spectral acquisitions (Figure 2.6 and Figure 2.7). Cosmic Ray Removal option was selected for each measurement providing three repetitions for each spectrum acquisition.

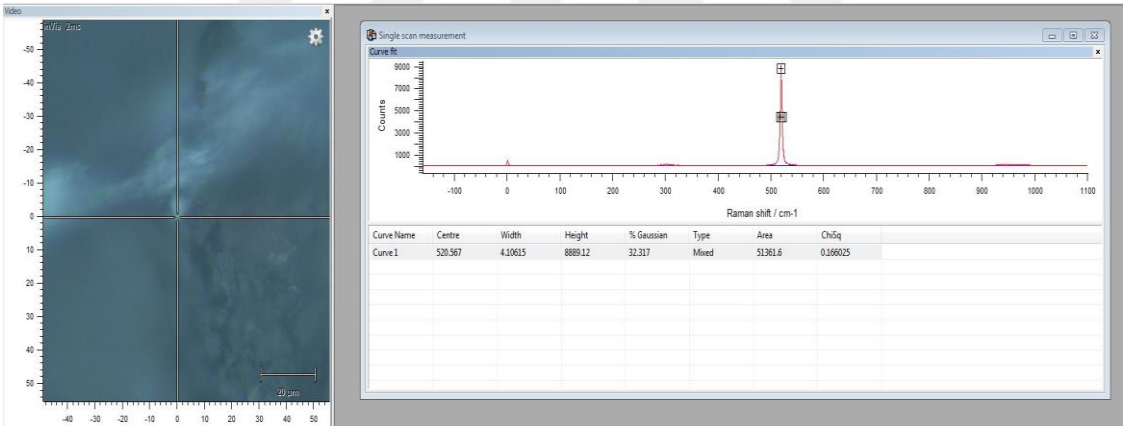


Figure 2.6. Origin setting and calibration of Renishaw InVia Raman Microscope using Wire 3.4 Software

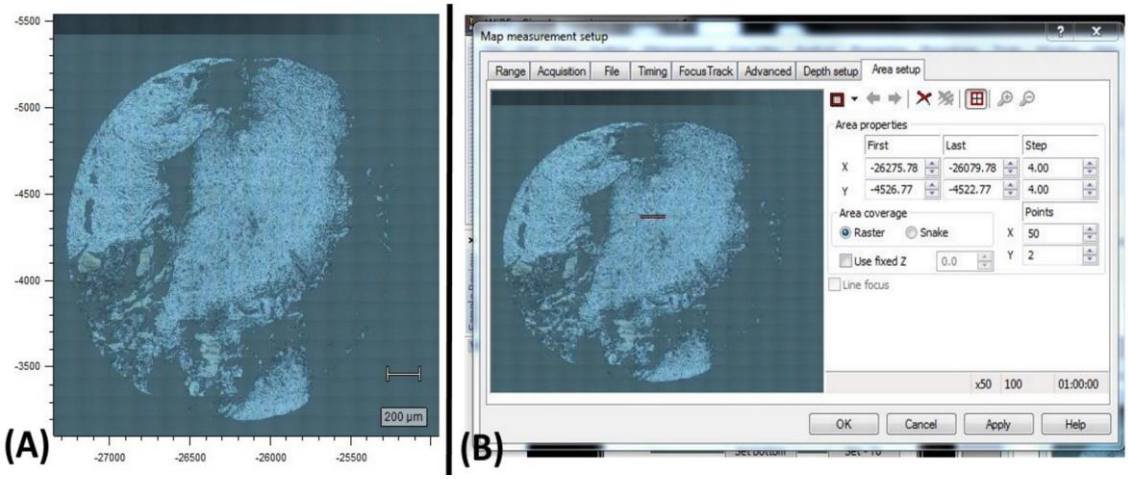


Figure 2.7. Raman Measurement Setup (A) White Light Image of the region of interest. (B) Area Setup and representative image of the measured area.

2.7 NIR Spectroscopy Measurements

UV-3600 Shimadzu UV-VIS-NIR Spectrophotometer was used for absorbance and transmittance measurements in tissue thickness optimization for Raman measurements. Absorbance and transmittance values of the tissue sections (10µm-thick) were recorded for three available Raman laser wavelengths (532nm, 633 nm, 785 nm) in our facility. Penetration depth values were calculated according to Beer-Lambert Law for each wavelength.

2.8 Data processing and analysis:

The cosmic-ray removal operation was applied to all datasets using Renishaw Wire (Windows-based Raman Environment) 4.3 software (©2014 Renishaw plc, U.K.) and plotted using MATLAB R2020a and Python. Then, spectra were classified either using the following pre-processing methods or without pre-processing with three different machine learning algorithms: SVM, kNN, and random forest classifiers (RF). Machine learning classifications of tissue spectra were performed with Numan Batur.

2.8.1 Pre-Processing

Normalization

Spectra were normalized with min-max normalization [87] according to the following formula:

x = Raman spectra

$x_{minimum}$ = minimum(x_1, x_2, …, x_n)

$x_{maximum}$ = maximum(x_1, x_2, …, x_n)

$x_{normalized} = (x - x_{minimum}) / (x_{maximum}-x_{minimum})$ [87]

The minimum and maximum values of the intensities of the spectra were determined and the Raman intensities of related peaks were replaced with the new intensity values by subtracting minimum values. Then, the new intensity values were divided by the difference between maximum and minimum intensities [87].

Savitzky-Golay Smoothing Filter

The Raman spectra were smoothed using the Savitzky-Golay smoothing filter, which is the most frequently used Raman spectra smoothing method based on the local polynomial fitting procedure. The filter fits a polynomial with the given order by the linear least-squares method to the points inside the window of the specified width to eliminate the noise from the spectra [87]. Although there is no suggested window size and polynomial order for denoising Raman spectra, we determined an optimum pair by iterative combinations that allow us to remove the low-frequency components in our dataset.

Principal Component Analysis (PCA)

Principle component analysis (PCA) is a multivariate pre-processing method based on the axis rotation technique in the dataset and was employed to understand the affecting factors on spectral variation within the samples. PCA aligns a new group of axes that are termed as principal components (PCs), with the largest directions of variation in the dataset and determined from the covariance matrix of the main data table. The determined

PCs are used as the new feature set of the data, on which the model is trained[88]. In our classification process, the PC set that is enough to explain 90% of the data is chosen within all pre-processing pipelines. Since all pre-processing operation changes the general structure of the data, the number of PCs that explains the 90% of the data could vary depending on the pre-processing pipeline.

Standard Scaling

The scaling operation was employed for each feature on the dataset by subtracting the mean from the feature and scaling by the variance of the spectrum (standard deviation). The purpose of this operation is to have zero mean and unit standard deviation for each feature. The scaling operation was applied by using the 'StandardScaler class' in the 'sci-kit learn' library of Python.

2.8.2 Machine Learning Algorithms

Support Vector Machine (SVM)

SVM is a machine learning algorithm based on the separation of labeled data points to positive and negative classes in real numbers ($\mathbb{R}^n$). The principle of the SVM model is to find the separating hyperplane with the possible widest margin of separation of two classes that are termed as the margin of error. This widest margin of the separating hyperplane carries data points that are the vectors from positive or negative class and these data points are termed as support vectors (SVs). The position of support vectors determines the location of the optimal separating hyperplane [89]. The 'C' parameter that resembles the trade-off between error toleration and maximization of the separation margin is one of the most important parameters in the SVM model [90].

k-Nearest Neighbors (kNN)

The kNN model is one of the basic machine learning algorithms and it relies on the principle is that the nearest samples to each other have similar characteristics. This leads to predicting the nearest neighbors' characteristics based on the ones whose

characteristics are already known, where the ‘k’ value is a positive integer [91]. kNN is the closest neighbor technique and is based on the idea that 'k' neighbors of any new sample can be classified by majority vote. kNN is one of the simplest and most straightforward data mining techniques and as training instances must be in memory at runtime, it is referred to as Memory Based Classification. The difference between the continuous attributes is calculated upon the Euclidean distance formula and the considerable issue while coping with the formula of Euclidean distance is the frequency of large values sinks the frequency of smaller values. Thus, continuous attributes are normalized to a defined interval to deal with this issue, so that both frequencies have the same effect on the distance value between instances [92].

Random Forest (RF)

RF is a machine learning method that is utilized for classification and regression of the dataset and is known for its high superiority and accuracy. RF is based on the combination of several decision trees and constructing a forest from decision trees [93] using training data bagging and creating an ensemble model. Each decision tree in the ensemble is created using a sample that replaces training data. Decision trees in the forest play a role as a base classifier for determining the class tag of an untagged instance. This is done by a majority vote where each classifier gets one vote for the predicted class tag, then the top-rated class tag is used for classifying the given sample [94].

Chapter 3:
OPTIMIZATION OF SAMPLE PREPARATION AND RAMAN SPECTRAL ACQUISITION PROTOCOLS

GB, the most frequent and lethal type of brain cancer, is diagnosed upon histological and genetic features of the tumors by neuropathologists. In the clinics, FFPE tissue sections are evaluated after H&E and/or several immunohistochemical staining and classified based on morphological features for final diagnosis. After the 2016 update of the WHO CNS tumor classification [10], molecular and genetic alterations of the tumors have been involved in the traditional histological evaluation-based diagnosis only. According to the 2016 update in WHO CNS tumor classification, highly heterogeneous molecular and genetic features of GB have also been started being evaluated in addition to the morphological and histological evaluation of GB [9]. This update increased the use of genetic tests used in the diagnosis of GB, however, prolonged evaluation time of tumor sections and high costs of the molecular tests have been recognized as the challenges in the diagnosis of GB. Moreover, diagnostic variability among pathologists due to the heterogeneity of the tumor and the variations in expertise level of pathologists may result in differences in the final diagnosis.

To minimize interobserver variability and to reduce the time and cost spent in the tumor evaluation; an analytical, rapid, and reproducible digital pathology technique would be highly useful to aid neuropathologists in error-free diagnosis. According to this motivation, we use RS to analyze the molecular composition of the tumor sections. While RS provides bond-specific signals, the signal contrast must be maximized by optimizing the brain tumor section sample preparation conditions and Raman spectrum acquisition parameters. In sample preparation, since the samples are obtained mainly from FFPE archival tissues, the paraffin content must be dissolved without damaging the morphology or the chemical content of the tissue section. The Raman spectrum acquisition parameters must also be optimal to avoid burning the samples with too much laser power density or heating effects while the spectral contrast must also be as high as possible to classify samples accurately.

In this chapter, we present our optimized tissue preparation parameters (i.e. tissue thickness, paraffin clearing protocol) for repeatable RS measurements in addition to safe, rapid, and high-contrast Raman measurement parameters (i.e. laser excitation wavelength, acquisition time, accumulation time, and accumulation count, substrate type). This optimization study (8 FFPE blocks, n=6) allows us to obtain the optimal signal-to-noise ratio (SNR) without damaging the tissue morphology. Thus, these methods could be used reproducibly in the analysis of FFPE GB tissue sections. Previous studies indicated that Raman spectra are highly affected by the sample [39], substrate type [39, 77-80], and measurement conditions [39, 77]. Kerr, L. T. et al presented a study about optimal laser excitation wavelength and Raman substrate choice for biological specimen analysis using RS. This study highlighted that CaF_2 is the optimal Raman substrate despite its high-cost and compared the Raman spectra of healthy human cheek cells using various excitation wavelengths [77]. As a low-background Raman substrate, the interest in CaF_2 substrate use is increased among researchers in recent years [78-80].

On the other hand, FFPE tissue section preparation is one of the most important factors to obtain reliable spectra since the strong paraffin-related spectrum suppresses the tissue spectrum and prevents tissue classification through the Raman spectrum. Figure 3.1 shows the Raman spectra of a representative brain tissue sample before (FFPE GB) and after (deparaffinized GB) dissolving paraffin. The characteristic bond peaks of the brain tissue sample appear after eliminating the paraffin peaks around 888, 1062, 1132, 1169, 1295, 1417, 1440, and 1460 cm^{-1} by dissolving paraffin (Figure 3.1) (Table 3.1) [75, 95-97]. Since the paraffin peaks provide broad and strong background, paraffin residuals need to be cleared from the tissue sections while preserving the tissue morphology and chemical composition of the samples. Eliminating paraffin background and preserving tissue composition intact provide repeatable and accurate analysis of brain tissue sections using RS. If the tissue preparation and paraffin clearing process are harsh and chemical composition is not preserved in deparaffinization, Raman spectra cannot be used for the identification or classification of tumors. For translational medicine applications and clinical testing of RS use in cancer diagnostics, repeatable brain tissue preparation and Raman measurement protocols must be established.

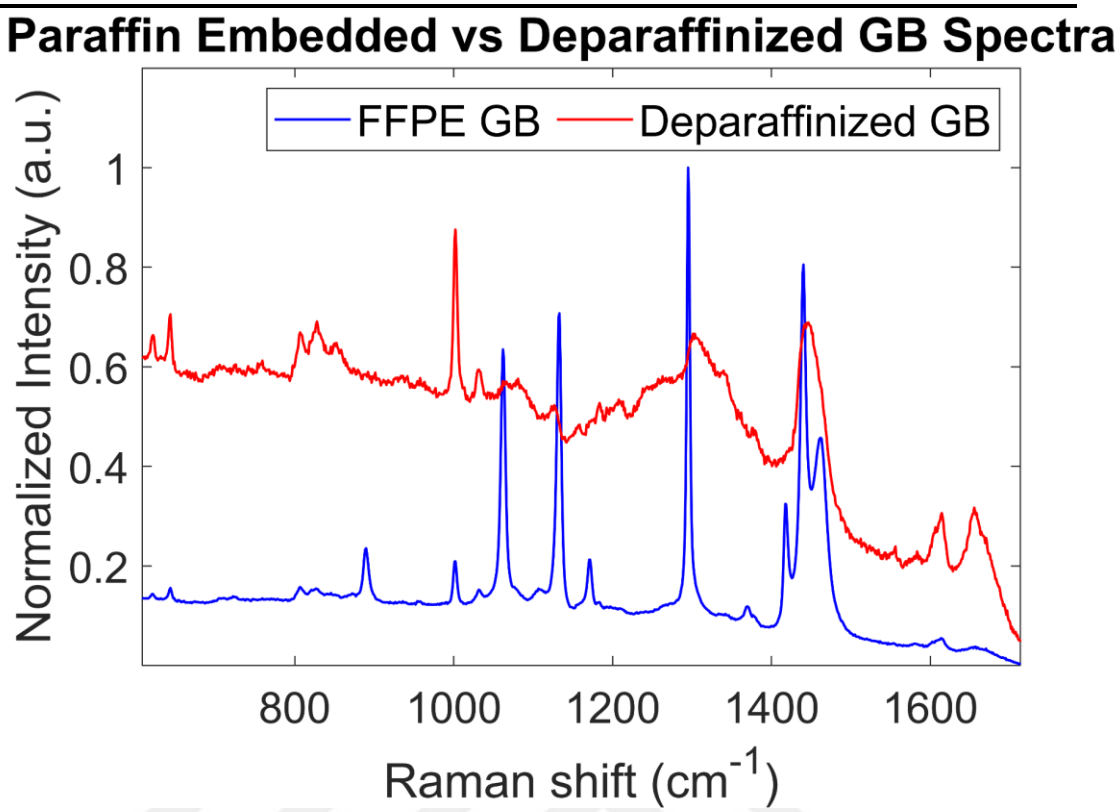


Figure 3.1. Paraffin-embedded vs deparaffinized glioblastoma (GB) tissue spectra. GB spectra from non-deparaffinized and deparaffinized tissue sections were acquired. Both spectra were recorded under full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. Deparaffinized tissue sections were deparaffinized according to the Clearene™-based deparaffinization protocol (See: Chapter 2).

We present a series of protocols for reproducible and repeatable FFPE brain tissue section preparation and Raman measurement protocols. We determined the minimum tissue section thickness (10 µm) by calculating the laser penetration depth in the tissue and investigated the effects of deparaffinization conditions and solvent type (p-Xylene, Clearene™). Moreover, we tested different substrate types (glass, CaF_2) to identify the optimal substrate with minimum Raman background. Then, we defined the optimal laser power (1%, 10%, 50%, 100%), laser acquisition time (1, 10, 20, 30, 40, 50 seconds), laser accumulation count (1, 3, 5, 10, 20), and laser excitation wavelength (532, 633 and 785 nm) to obtain the best SNR without damaging the tissue for RS. While the protocols presented in this optimization study are optimized for Raman spectroscopic analysis of FFPE brain tissue sections (8 FFPE GB block from 6 patients), the sample preparation

conditions, and Raman measurement parameters could be refined for other tissue types as well.

Refining the tissue section thickness, Raman substrate type, and deparaffinization protocols are necessary to obtain repeatable Raman spectra by reducing the variations during either the tissue preparation and Raman measurement process. Applying these optimized parameters also helps to eliminate the requirements for preprocessing steps during the data analysis (i.e. background removal).

3.1 Optimal minimum tissue thickness for Raman spectroscopy

Penetration depth is a measure of how deep light can penetrate the sample and is defined as the depth in which the radiation intensity in the sample falls to e^{-1} (36.78%) of its original value. The penetration depth value is calculated after measuring the optical properties of tissue [98]. Beer-Lambert Law is coined after the work of August Beer and Johan Lambert about the exponential relation between the transmission of the light and penetration depth. According to Beer-Lambert Law, the intensity of the transmitted beam exiting the sample logarithmically decreases with increasing concentration (August Beer) or path length (Johann Lambert). The optical absorption relies on the incident beam intensity, the concentration of absorbing species, the path length of the beam inside the sample, and the extinction coefficient of the tissue [99, 100].

First, the transmission values for 532, 633, and 785 nm excitation wavelengths were measured using NIR spectroscopy. Then, laser penetration depths for three different laser excitation wavelengths (532, 633, and 785 nm) were calculated using the following equations in MATLAB according to Beer-Lambert Law.

[99-101]

$$\mathrm{T(Transmittance)} = \frac{\mathrm{I\ (Transmitted\ light)}}{\mathrm{I_0\ (Incident\ Light)}}$$

$$\mathrm{T} = e^{-\frac{\mu(\mathrm{attenuation\ coefficient})}{\mathrm{d\ (tissue\ thickness)}}}$$

$$\mathrm{T} = e^{-\left(\frac{\mathrm{d\ (tissue\ thickness)}}{\delta\ (\mathrm{penetration\ depth})}\right)}$$

$$\delta = 1/\alpha$$

Absorbance (A) = 2 – log(%T)

α (absorption coefficient) = 2.303 × A /d

- For the 532 nm laser excitation wavelength, penetration depth was calculated as 8.691 (Figure 3.2).

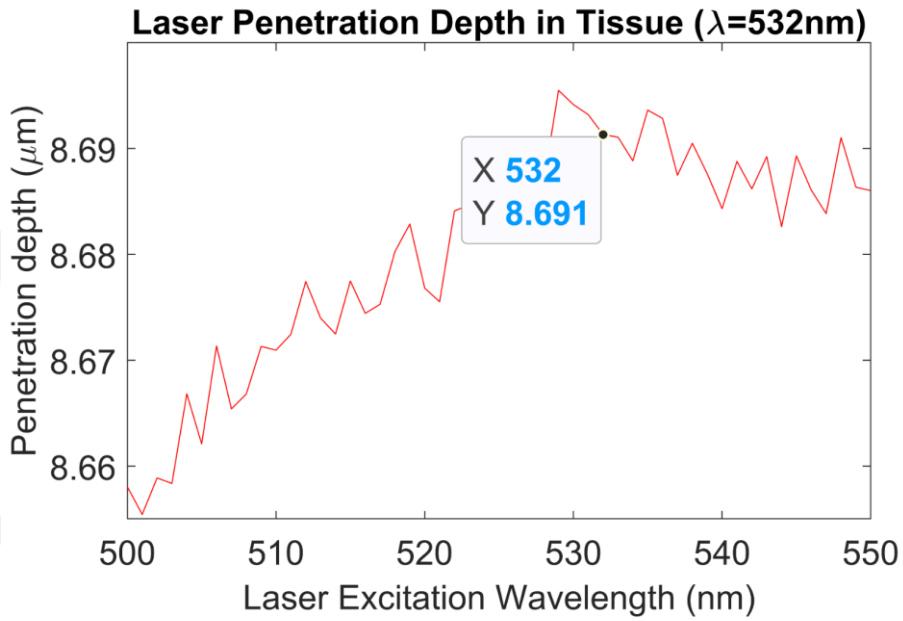


Figure 3.2. Penetration depth for 532 nm excitation wavelength

- For the 633 nm laser excitation wavelength, penetration depth was calculated as 8.8 (Figure 3.3).

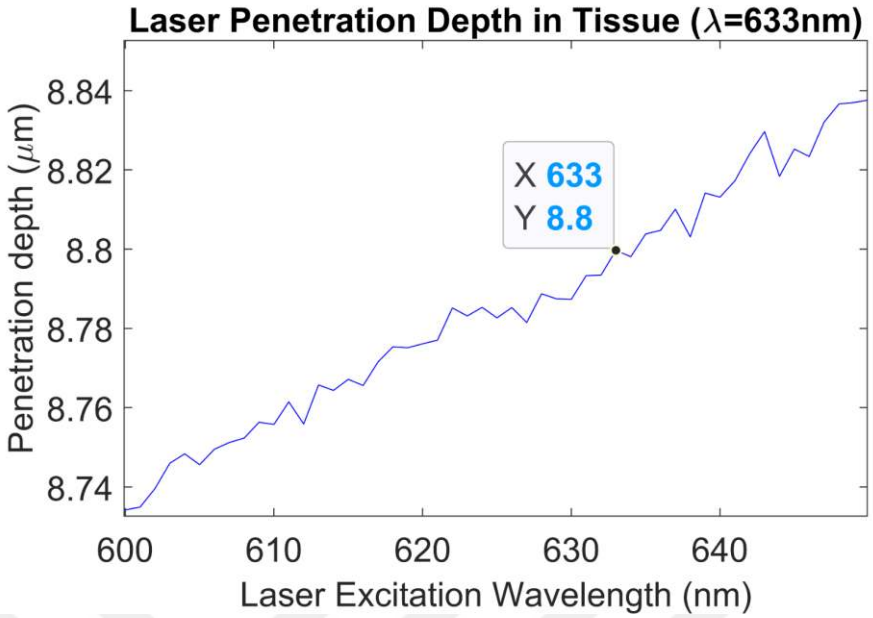


Figure 3.3. Penetration depth for 633 nm excitation wavelength

- For the 785 nm laser excitation wavelength, penetration depth was calculated as 9.128 (Figure 3.4).

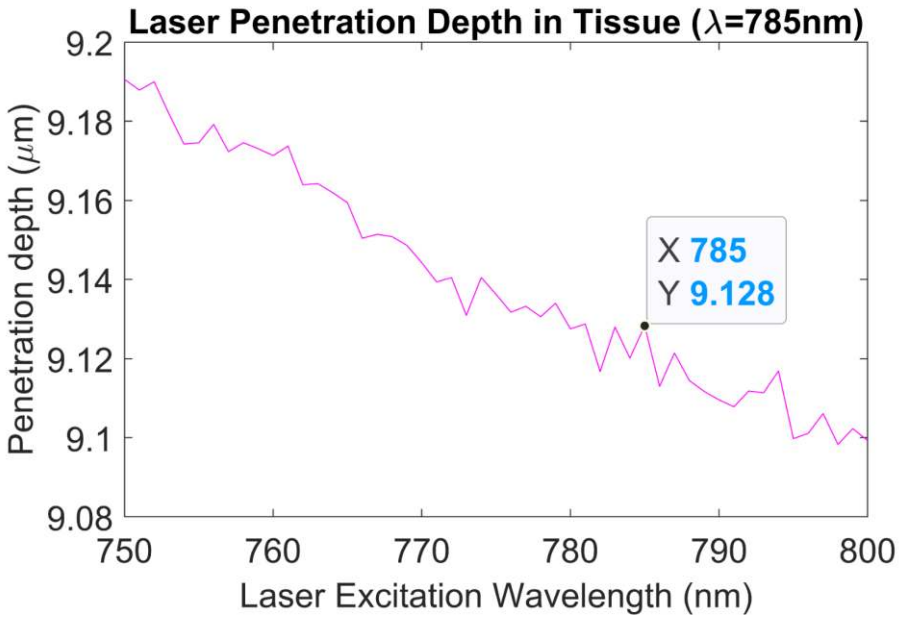


Figure 3.4. Penetration depth for 785 nm excitation wavelength

To avoid substrate background using RS, we suggest a minimum of 10 µm-thick brain tissue use for Raman spectra with better quality. Lower Raman spectrum

background intensity also provides minimum pre-processing of spectra during data analysis.

3.2 The Effect of excitation wavelength and the substrate type on the Raman background and signal-to-noise ratio (SNR)

In RS measurements, the background signal induced by the sample and substrate molecular structure and fluorescence may prevent reliable detection of Raman peaks. Fluorescence is an emission of light from the samples or the substrate after being illuminated by light. Electrons are excited from the highest occupied molecular orbital (HOMO) levels to the lowest unoccupied molecular orbitals (LUMO) with photons of energies greater than the LUMO-HOMO energy difference. Electrons thermalize from their excited states and emit fluorescence photons from multiple energy levels near HOMO. Owing to these multiple emissions, fluorescence background could be observed as strong and broadband in the Raman spectra. Since broadband fluorescence background is a strong factor that might suppress the Raman spectral peaks of the given sample, Raman measurements must use the optimal laser excitation wavelengths and substrate types for minimizing or eliminating fluorescence background and obtaining stable Raman peaks [102].

Laser excitation wavelength choice is one of the most important factors in avoiding fluorescent background and photodamaging [39] of the tissue, and for improving the SNR. Most frequently used excitation sources such as green or red (532 and 633 nm) lasers with relatively higher energy photons cause strong fluorescence background overwhelming the relatively weak Raman signal [103] and increase the risk of photodamaging in the tissue sections. Using an infrared laser with lower-energy photons, such as 785 nm or longer wavelength, results in relatively less fluorescence as fluorescent background occurs when the laser photon energy is close to the fluorescent transition energy of the sample and its substrate [39, 103, 104].

In this optimization study, we used GB and white matter (WM) regions from the same patient's FFPE tissue sections. After determining the optimal tissue thickness, the FFPE tissue block was sectioned in 10 µm-thick and the tissue

sections were tested under 532, 633, and 785 nm excitation wavelengths on different Raman substrates. Backgrounds of CaF_2 and glass substrates were compared under the exposure of three different excitation wavelengths (Figures 3.5-3.10).

For different excitation wavelengths, linewidths of the peaks and peak positions were observed differently. Figure 3.5 represents the spectra of both GB and WM on the CaF_2 slide and Figure 3.6 represents the spectra of both GB and WM on the glass slide under 532 nm excitation wavelength. These spectra show higher SNR and relatively low fluorescent background with respect to the other spectra obtained 633 nm and 785 nm excitation wavelengths. (Figures 3.7-3.10) Figure 3.7 represents the spectra of both GB and WM on the CaF_2 slide and Figure 17 represents the spectra of both GB and WM on the glass slide using 633 nm excitation wavelength. Among these three different lasers, the spectra recorded under the 633 nm excitation wavelength showed poor quality and significantly low SNR caused by the highest fluorescent background (Figure 3.7 and Figure 3.8) with no visible tissue damage (Figure 3.11), consistent with the studies previously presented in the literature [39, 104]. Despite the other excitation wavelengths, 785 nm showed the lowest fluorescent background among the others (Figure 3.9 and Figure 3.10), minimal detector saturation, and tissue damage risk (Figure 3.11). According to these results, we suggest using the 785 nm laser source (Figure 3.9 and 3.10) and CaF_2 Raman slides for ideal Raman spectra with constant and low background and the highest SNR.

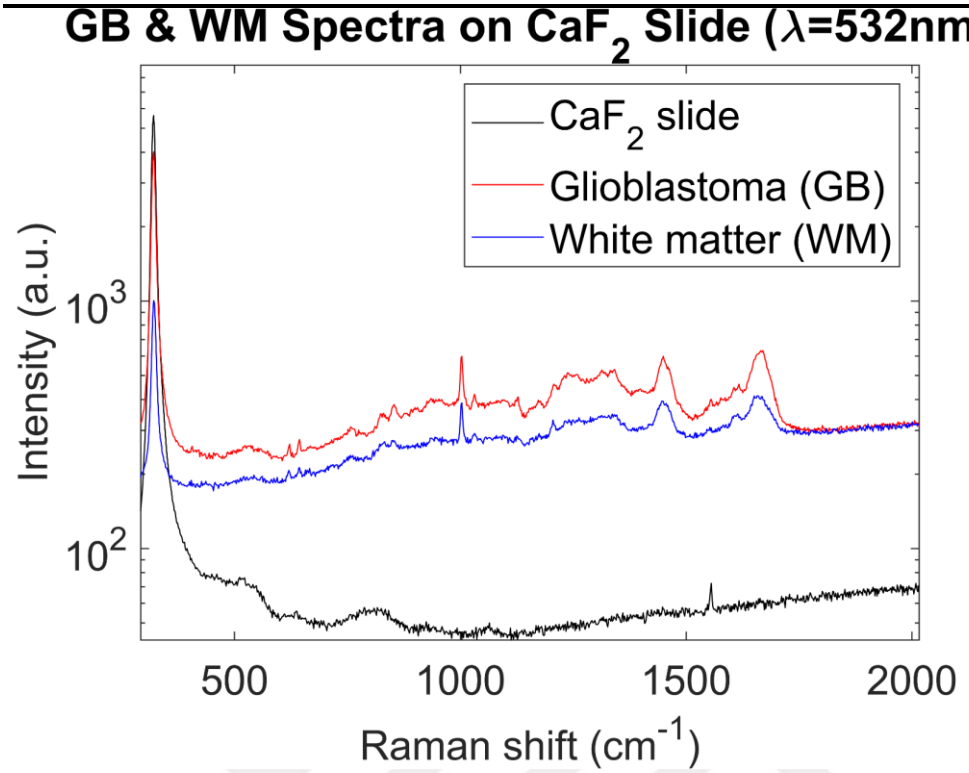


Figure 3.5. Glioblastoma (GB) and white matter (WM) tissue spectra on Calcium Fluoride (CaF_2) slide under 532 nm excitation wavelength

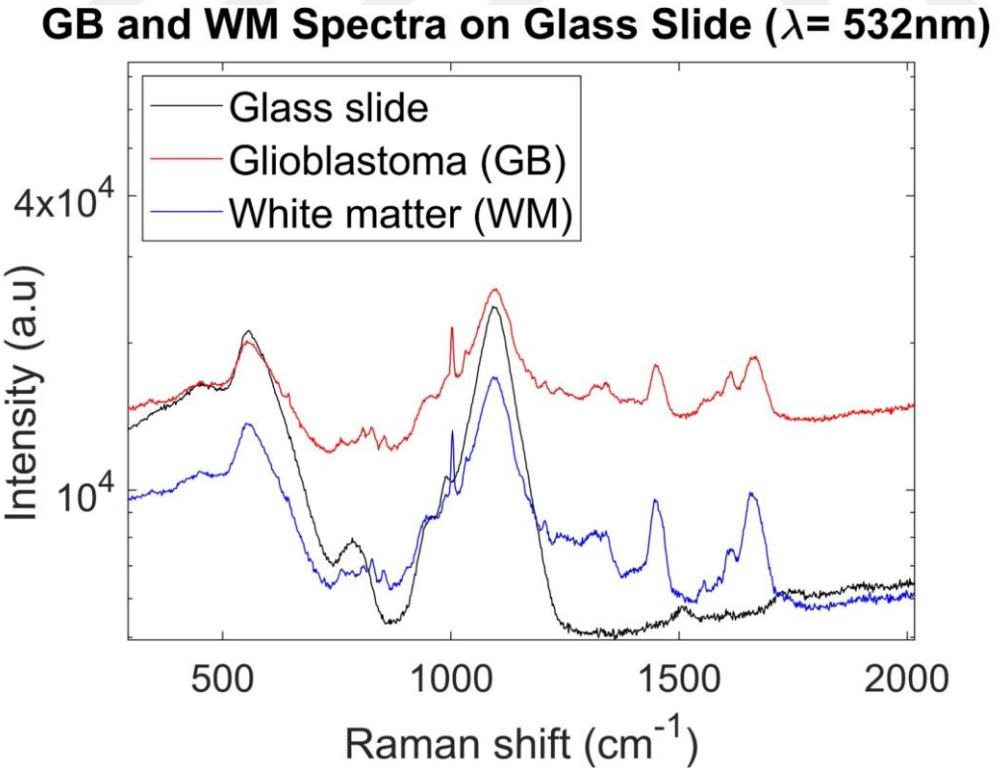


Figure 3.6. Glioblastoma (GB) and white matter (WM) tissue spectra on the glass slide under 532 nm excitation wavelength

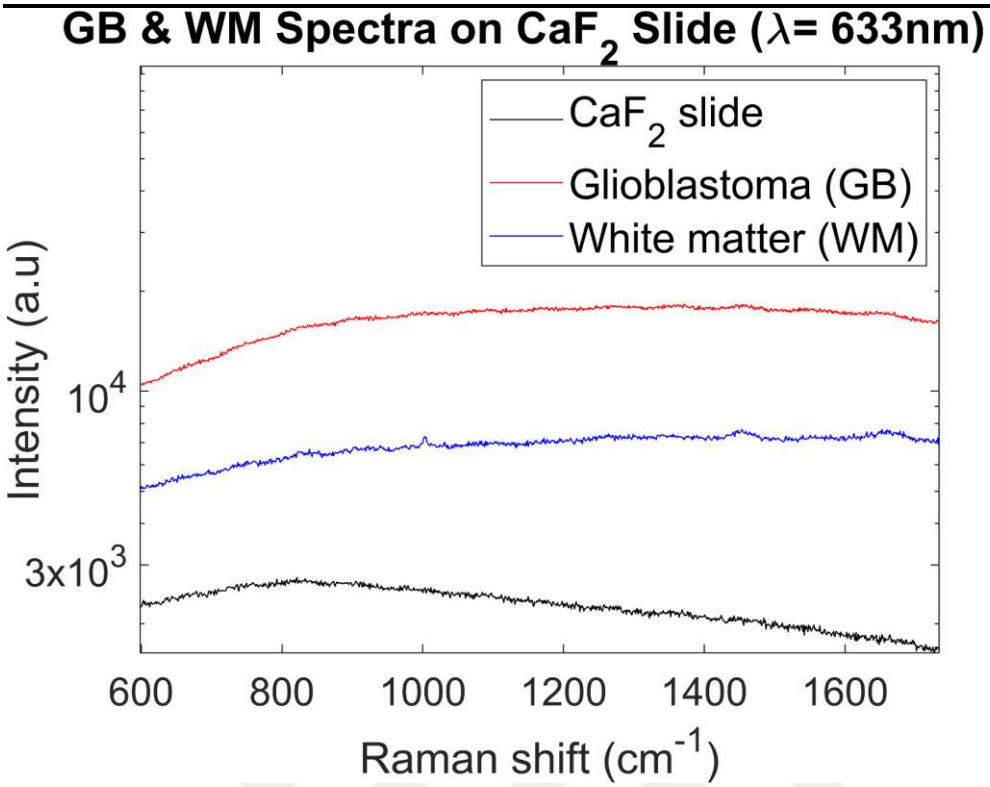


Figure 3.7. Glioblastoma (GB) and white matter (WM) tissue spectra on Calcium Fluoride (CaF_2) slide under 633 nm excitation wavelength

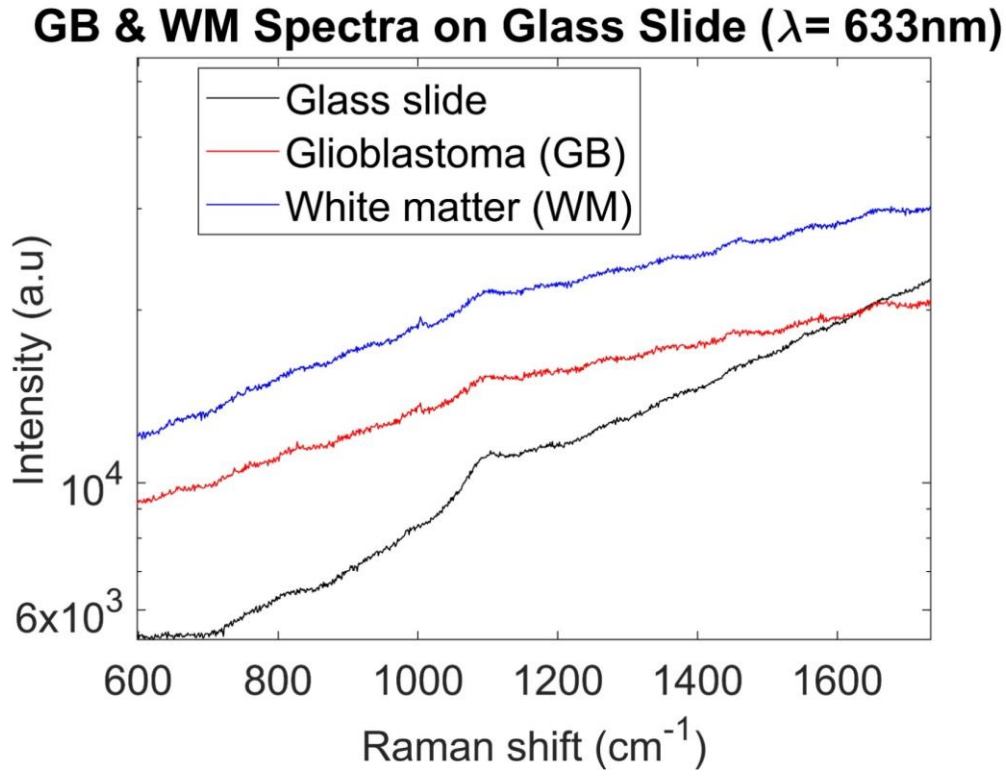


Figure 3.8. Glioblastoma (GB) and white matter (WM) tissue spectra on the glass slide under 633 nm excitation wavelength

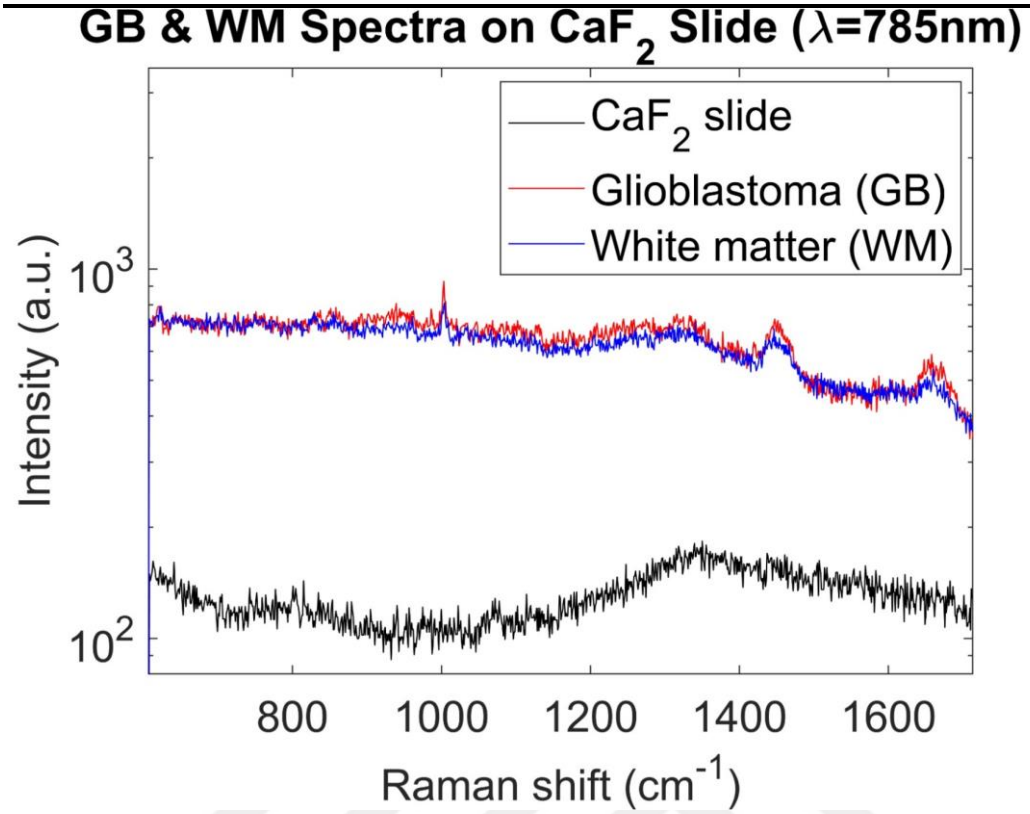


Figure 3.9. Glioblastoma (GB) and white matter (WM) tissue spectra on Calcium Fluoride (CaF_2) slide under 785 nm excitation wavelength

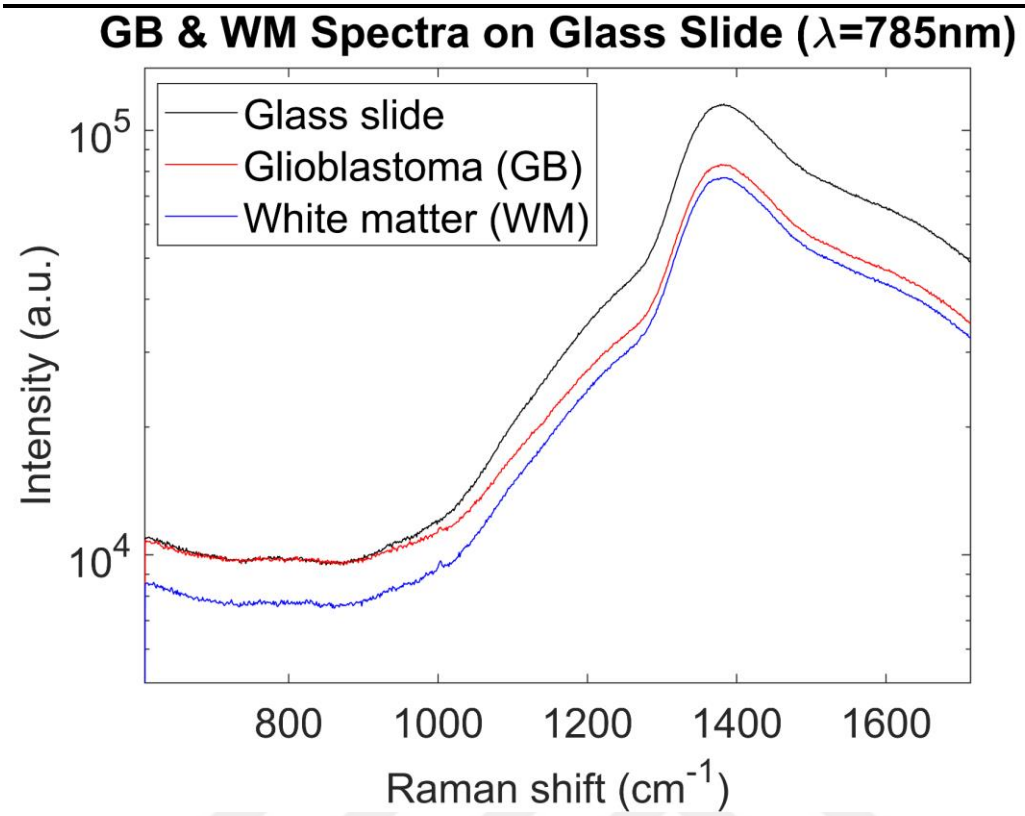


Figure 3.10. Glioblastoma (GB) and white matter (WM) tissue spectra on the glass slide under 785 nm excitation wavelength

Although no change was observed on the tissue morphology in the optical image of the Raman microscope after the laser exposure when 633 nm and 785 nm lasers were applied, a color change was observed in the Raman microscope image of the tissue surface when 532 nm laser was applied. (Figure 3.11) Even though we obtained the higher SNR using 532 nm excitation wavelength among other lasers, since preserving tissue morphology is the most important factor in this study, we suggest using 785 nm excitation wavelength for the reproducible analysis of tissue sections using RS.

Moreover, we observed that the fluorescent background was reduced and stabilized after multiple laser exposure to the tissue (Figure 3.12), however, multiple laser exposure increases the risk of tissue damage while using a 532 nm excitation wavelength. Increasing the laser power, acquisition time, or accumulation count, increased the risk of damaging the tissue (Figure 3.11). Besides, the spectra samples' Raman emission saturated the detector as presented in the next section.

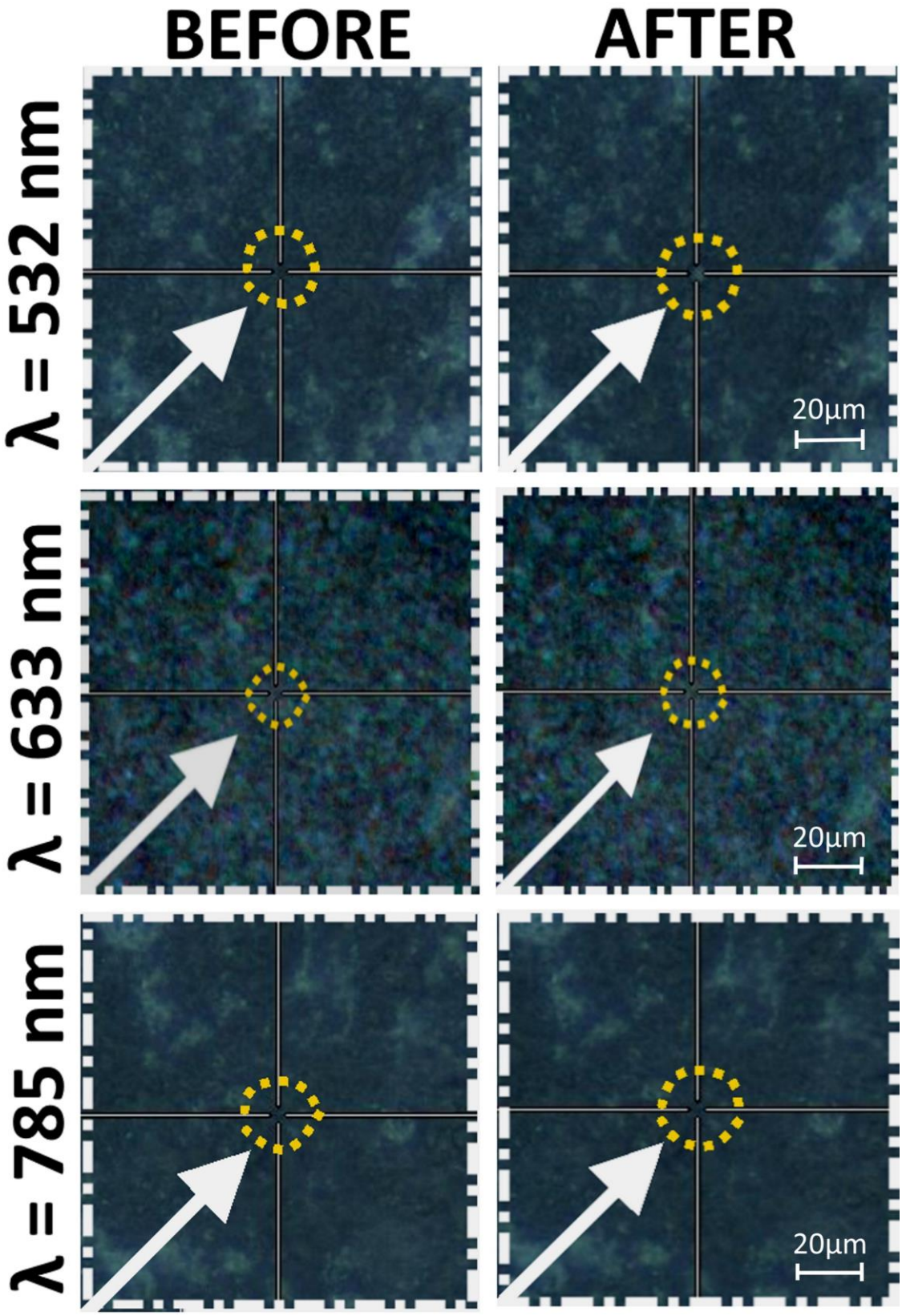


Figure 3.11. Photodamage effect of lasers with 532, 633, and 785 nm excitation wavelengths on brain tissue

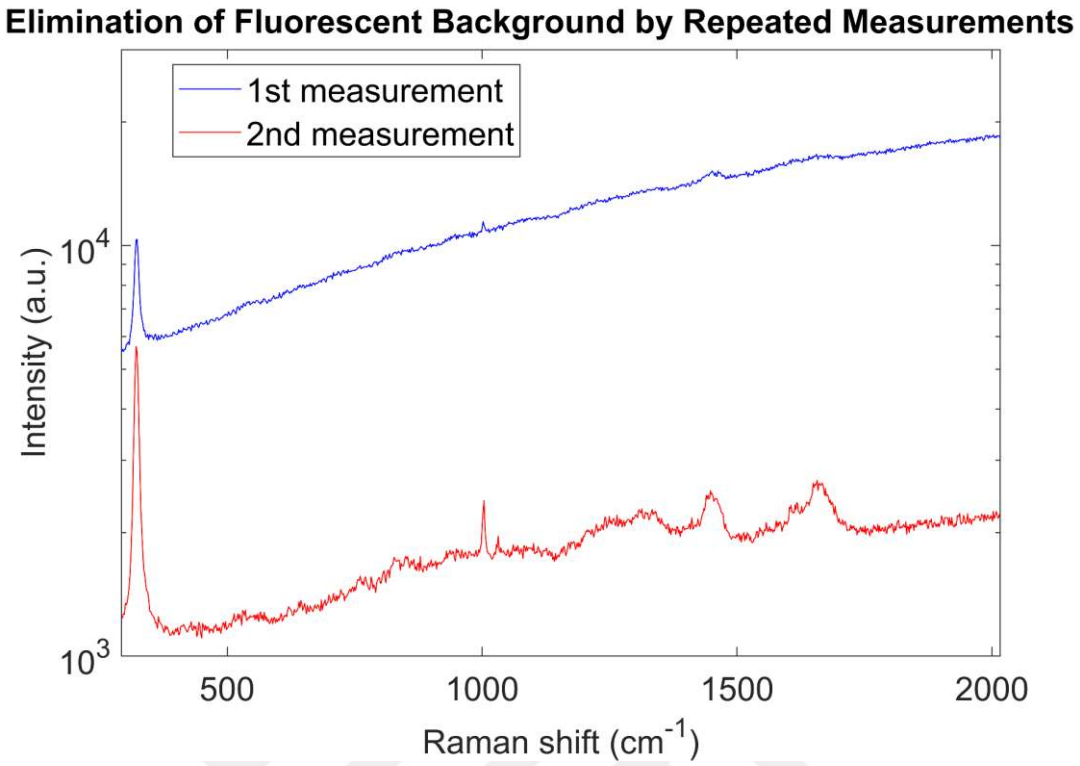


Figure 3.12. Reduction of fluorescent background by multiple laser exposures (λ=532 nm)

3.3 The Effect of Raman Measurement Parameters on Raman Spectral Contrast

To obtain the highest-SNR Raman spectra in a minimum duration for clinical applications of RS in brain cancer, optimizing Raman measurement parameters is necessary. In this section, we investigated the effect of laser power (as a percentage of maximum available power density in $mW \cdot \mu m^{-2}$), accumulation time, and accumulation count for three laser excitation wavelengths (532, 633, and 785 nm) and defined the optimal values for each for brain tissue analysis of RS. For each laser excitation wavelengths, Raman spectra were acquired from 10 µm-thick GB and WM tissue sections.

3.3.1 Optimization of Raman Measurement Parameters for 532 nm Excitation Wavelength

Laser power optimization for 532 nm excitation wavelength

Raman spectra of 10 µm-thick GB and WM tissue sections on both CaF_2 and glass substrates were acquired. Laser power was varied from 1, 10, 50, and 100% (of a maximum of ~6.11 $mW \cdot \mu m^{-2}$ power density) and was recorded for each power application. Each of these measurements was recorded under 10-second laser acquisition time and 1 laser accumulation count application. The SNR of the spectra increased with the increasing laser power and resulted in the highest value when 100% laser power was applied as its maximum on either CaF_2 or the glass substrate (Figure 21-25).

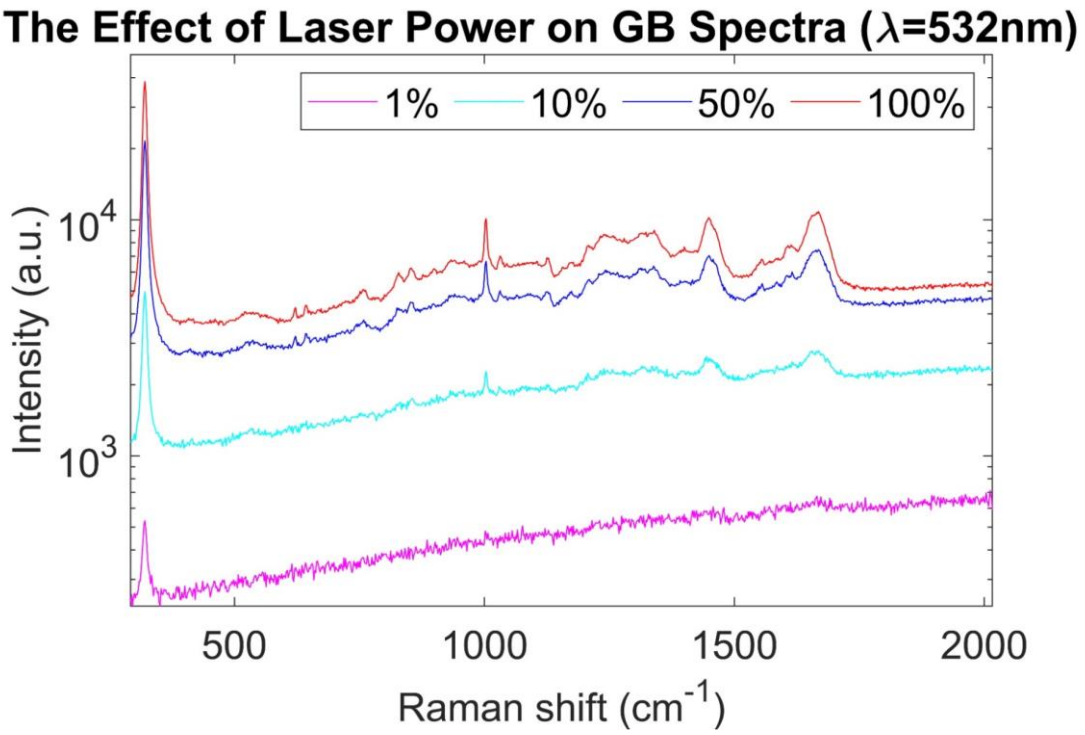


Figure 3.13. The effect of laser power on glioblastoma (GB) spectra on CaF_2 slide (λ=532nm)

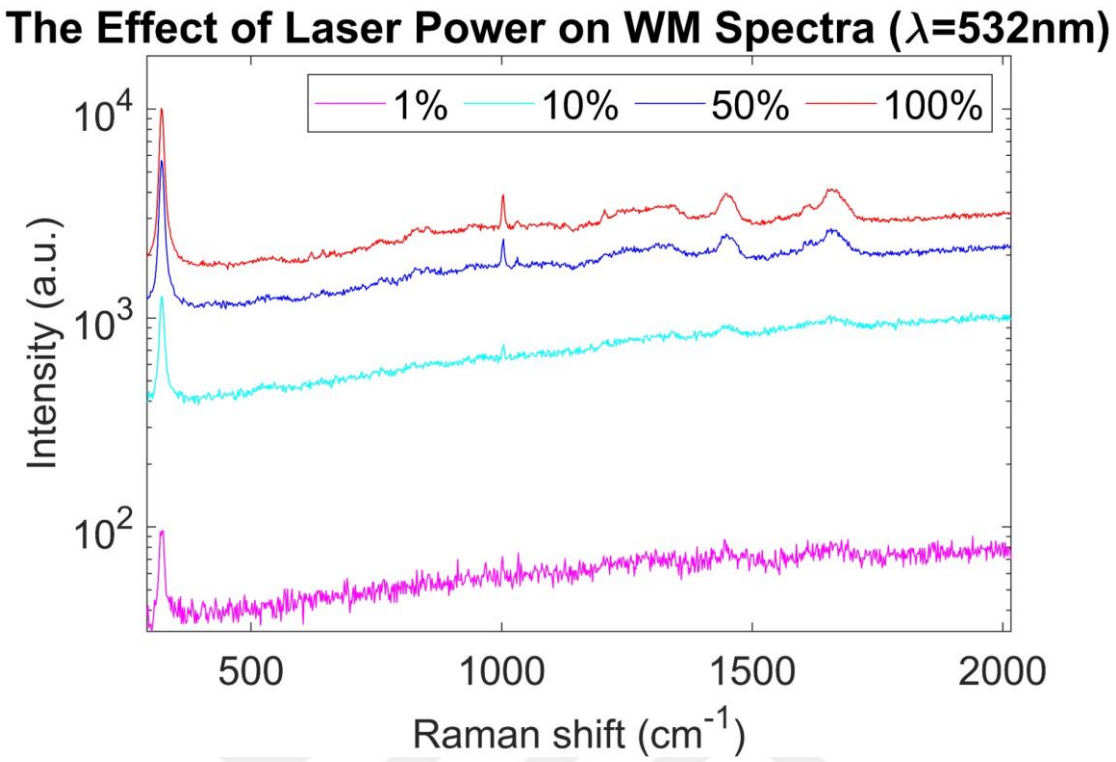


Figure 3.14. The effect of laser power on white matter (WM) spectra on CaF_2 slide (λ=532nm)

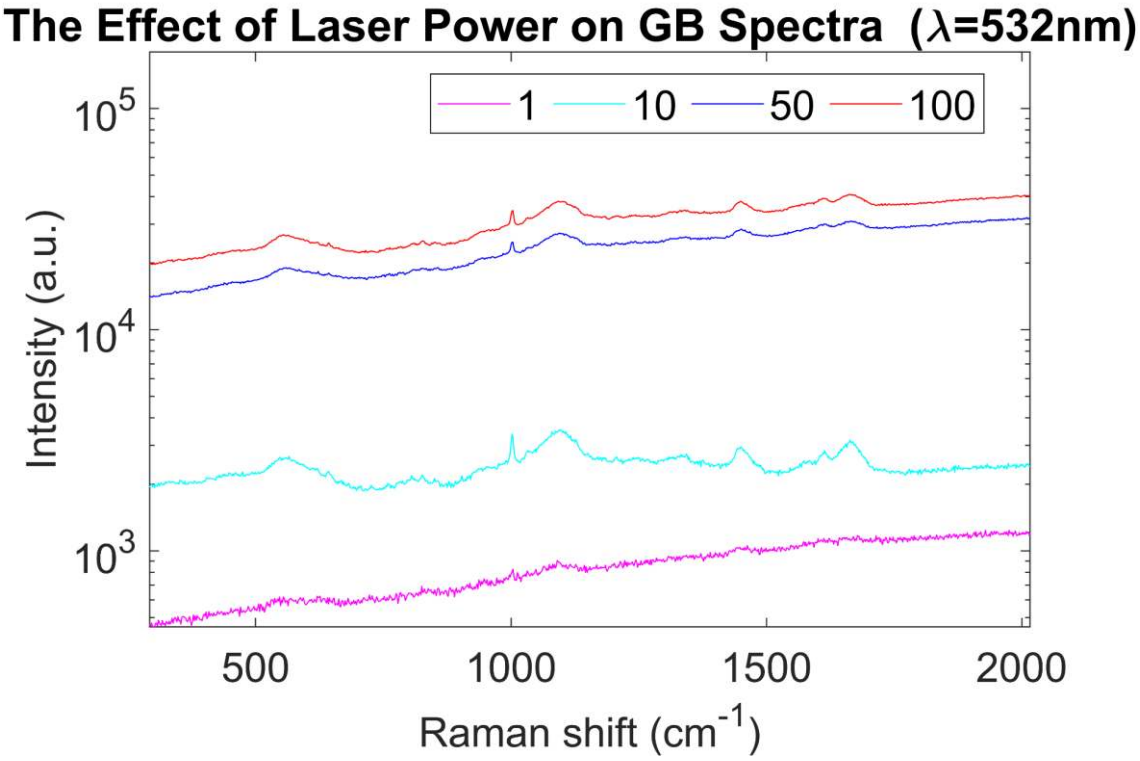


Figure 3.15.The effect of laser power on glioblastoma (GB) spectra on the glass slide (λ=532nm)

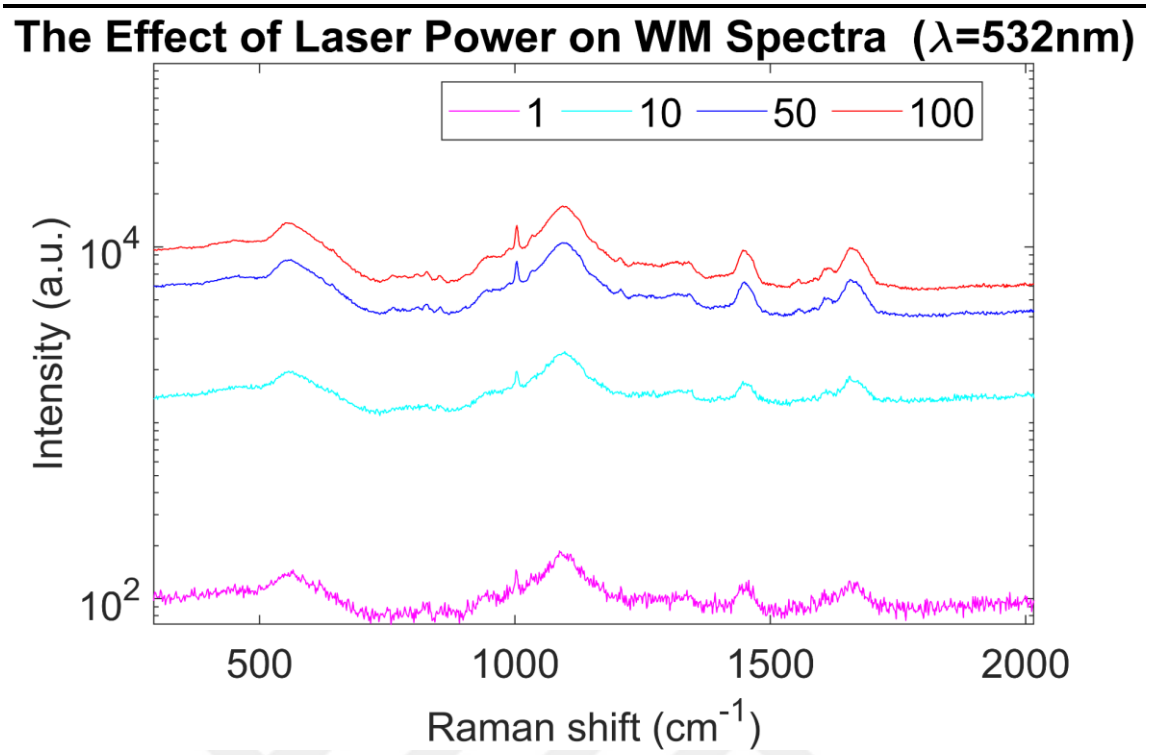


Figure 3.16. The effect of laser power on white matter (WM) spectra on the glass slide (λ=532nm)

Laser acquisition time optimization for 532 nm excitation wavelength

Raman spectra of 10 µm-thick GB and WM tissue sections on both CaF_2 and glass substrates were acquired. The laser power was set to its maximum value as 100% for the measurements in Figures 3.17-3.20 since this power level yields the highest SNR according to the laser power optimization in the previous section (Figures 3.12-3.16). One laser accumulation was applied to the GB and WM tissue sections for each measurement. The laser acquisition time was increased from 1 second to 10, 20, 30, 40, and 50 seconds for GB and WM tissue spectra measurements on both CaF_2 and glass substrates. We observed that the GB spectra saturated the detector when exposures longer than 10 seconds were applied on the CaF_2 substrate and WM spectra saturated the detector when exposures longer than 30 seconds were applied (Figure 3.17 and 3.18). For the tissue sections on the glass substrate, GB spectra saturated the detector when the laser is exposed to the tissue for more than 10 seconds and the WM spectra saturated the detector when the laser is exposed to the tissues for more than 20 seconds (Figure 3.19 and Figure 3.20). Some fluctuations were observed due to the fluorescent background (Figure 3.18). The

results indicate that 10-second accumulation is needed for good spectral contrast (Figures 3.17-3.20).

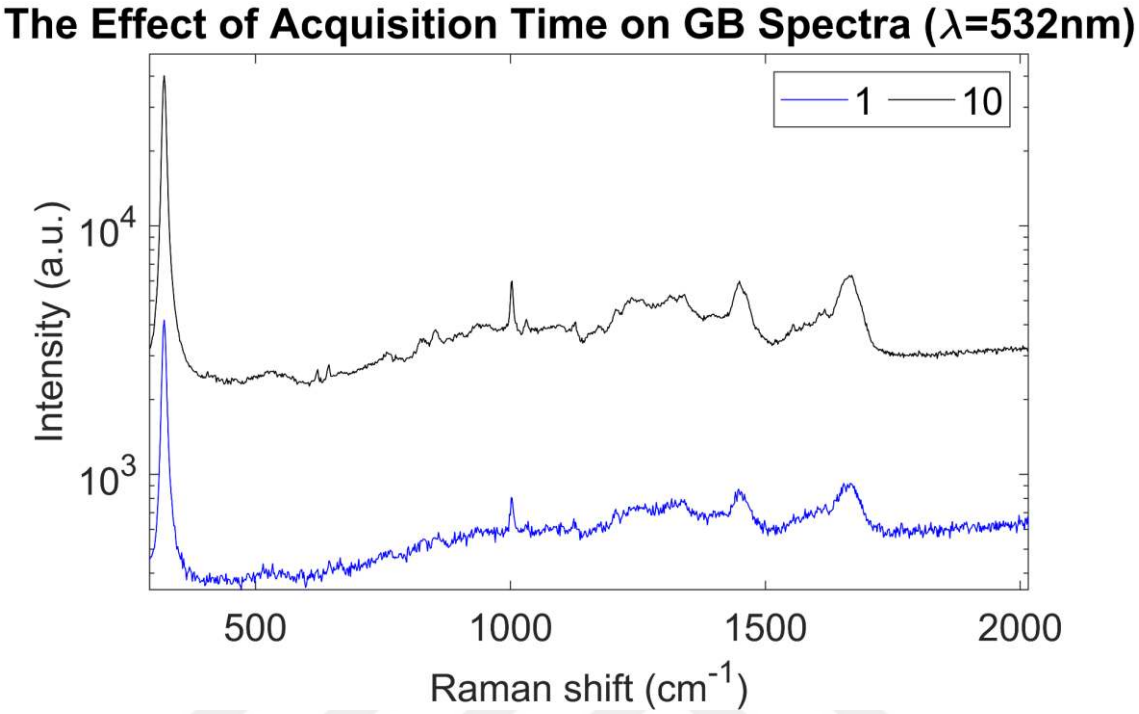


Figure 3.17. The effect of laser acquisition time on glioblastoma (GB) spectra on CaF_2 slide (λ=532nm)

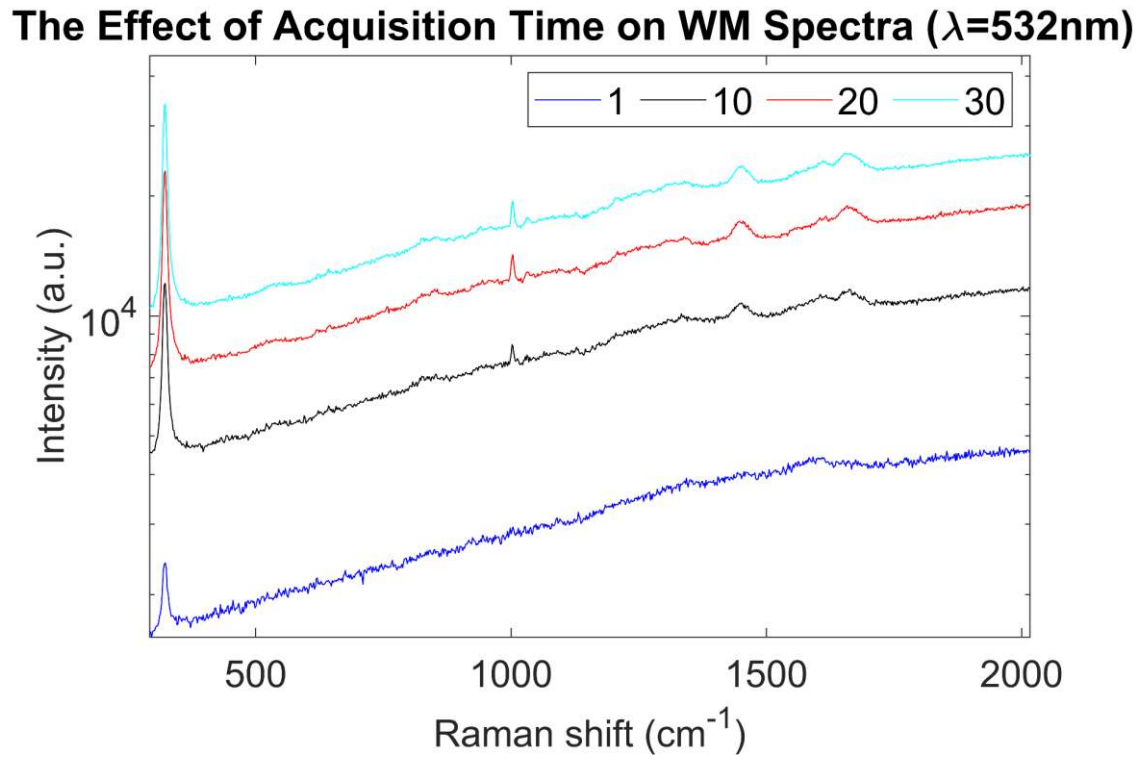


Figure 3.18. The effect of laser acquisition time on white matter (WM) spectra on CaF_2 slide (λ=532nm)

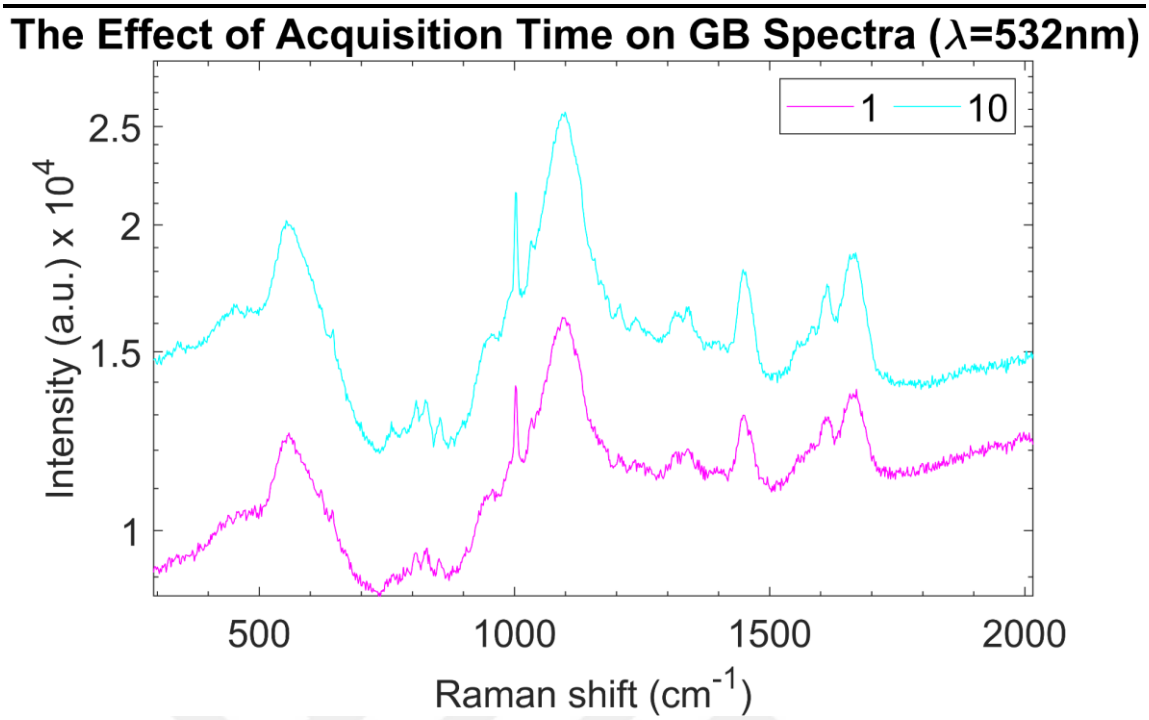


Figure 3.19. The effect of laser acquisition time on glioblastoma (GB) spectra on the glass slide (λ=532nm)

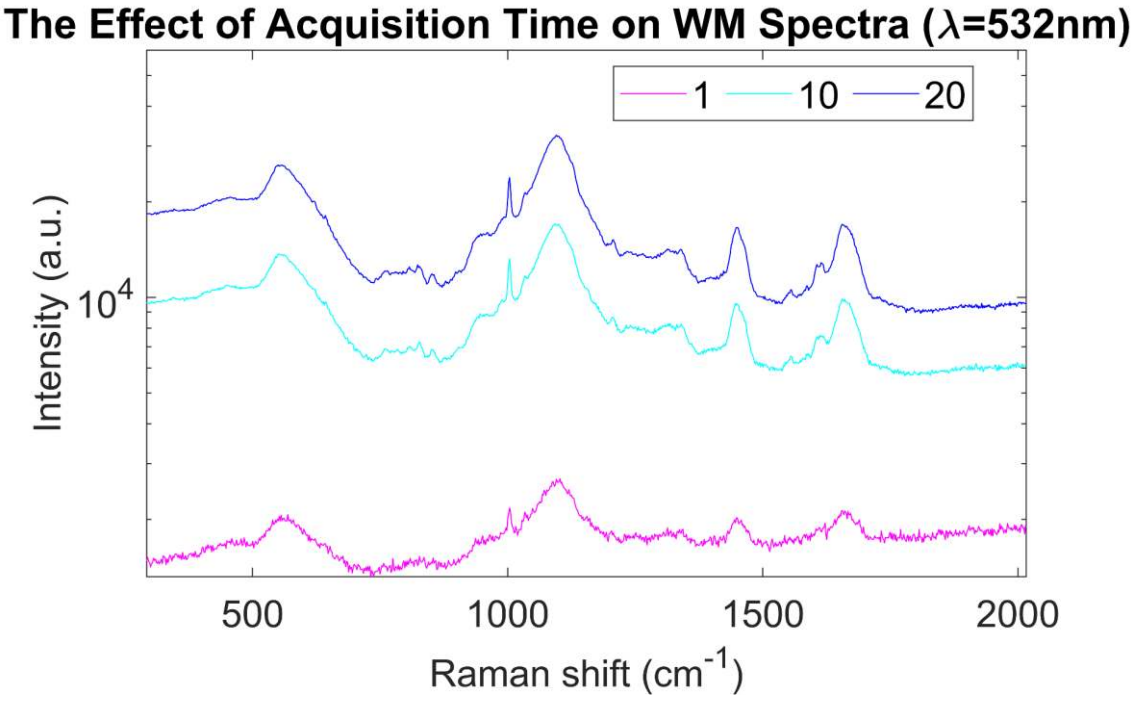


Figure 3.20. The effect of laser acquisition time on white matter (WM) spectra on the glass slide (λ=532nm)

Laser accumulation count optimization for 532 nm excitation wavelength

According to the previously presented optimal laser power and acquisition time parameters, we investigated optimal accumulation count for 532 nm excitation wavelength under the 100% laser power and 10 seconds laser acquisition. The accumulation count is increased 1 from 3, 5, 10, to 20. The spectra acquired from the GB tissue section on CaF_2 substrate saturated the detector when more than 5 accumulation count was applied (Figure 3.21), although no saturation detected on WM tissue section Figure 3.22). On the glass substrate, when more than 1 laser accumulation count is applied, either GB or the WM saturated the spectra (Figure 3.23 and Figure 3.24). Overall, we suggest applying no more than 1 accumulation count as it did not cause any saturation and is adequate for high-SNR spectra and rapid spectral measurements.

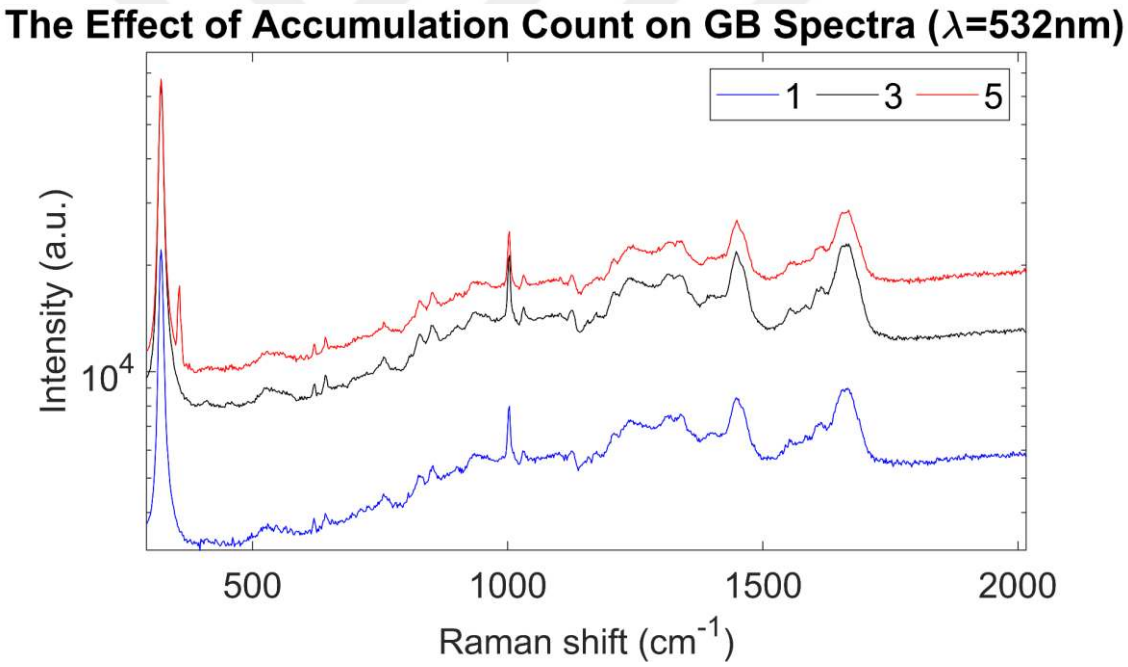


Figure 3.21. The effect of laser accumulation count on glioblastoma (GB) spectra on CaF_2 slide (λ=532nm)

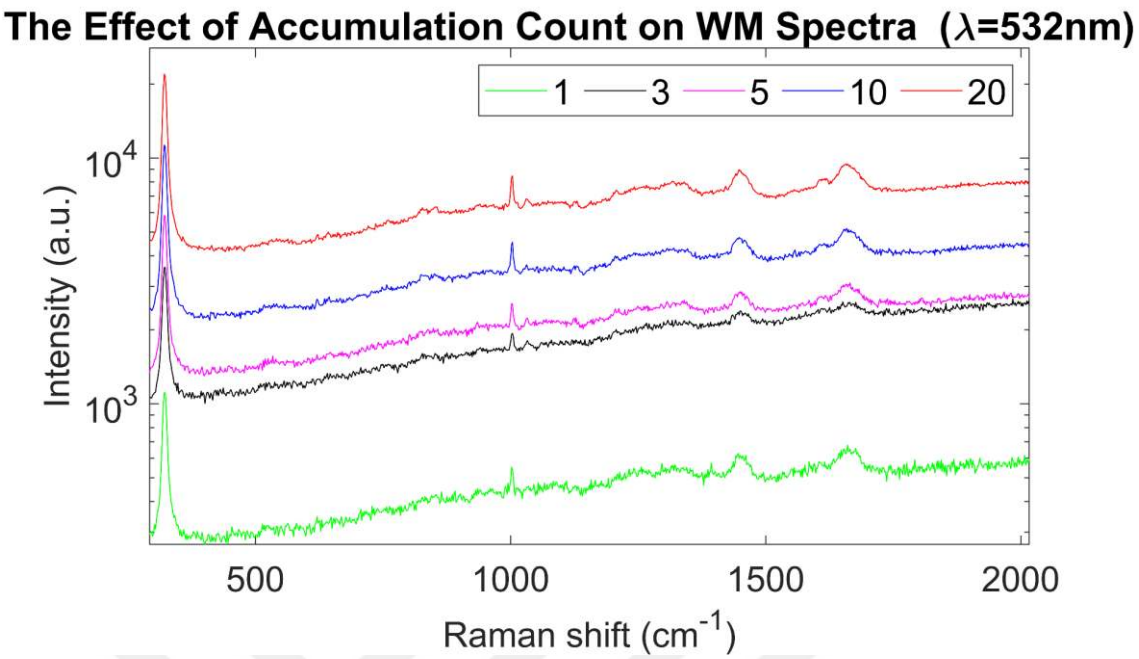


Figure 3.22. The effect of laser accumulation count on white matter (WM) spectra on CaF_2 slide (λ=532nm)

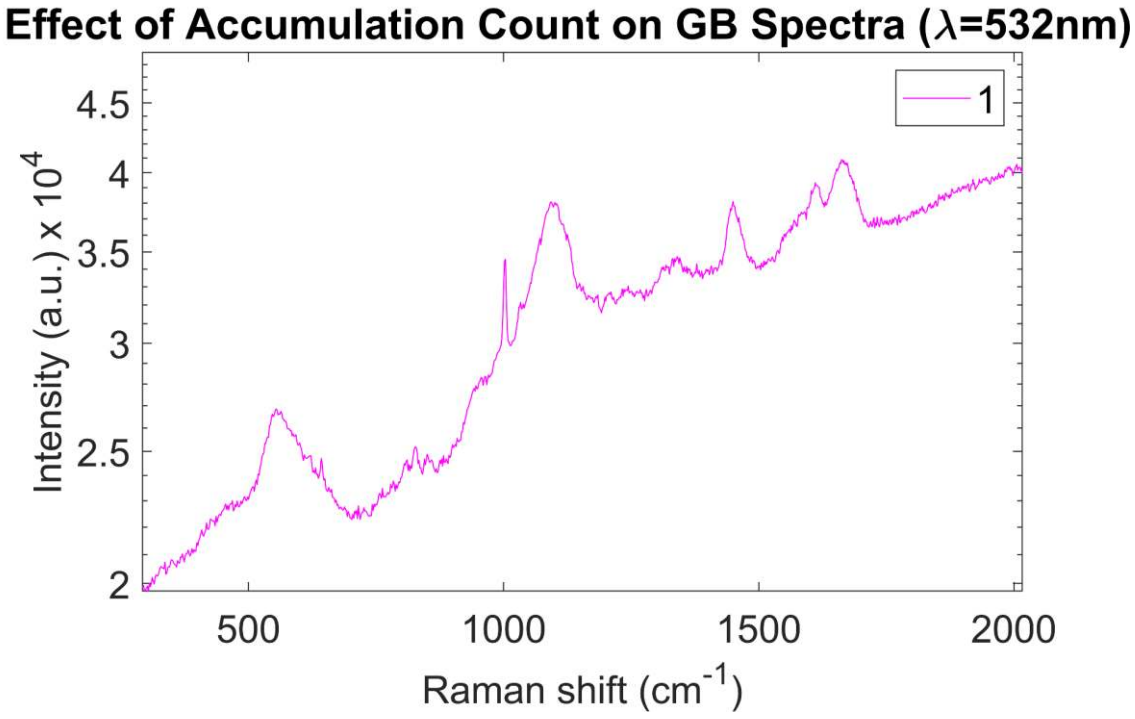


Figure 3.23. The effect of laser accumulation count on glioblastoma (GB) spectra on the glass slide (λ=532nm)

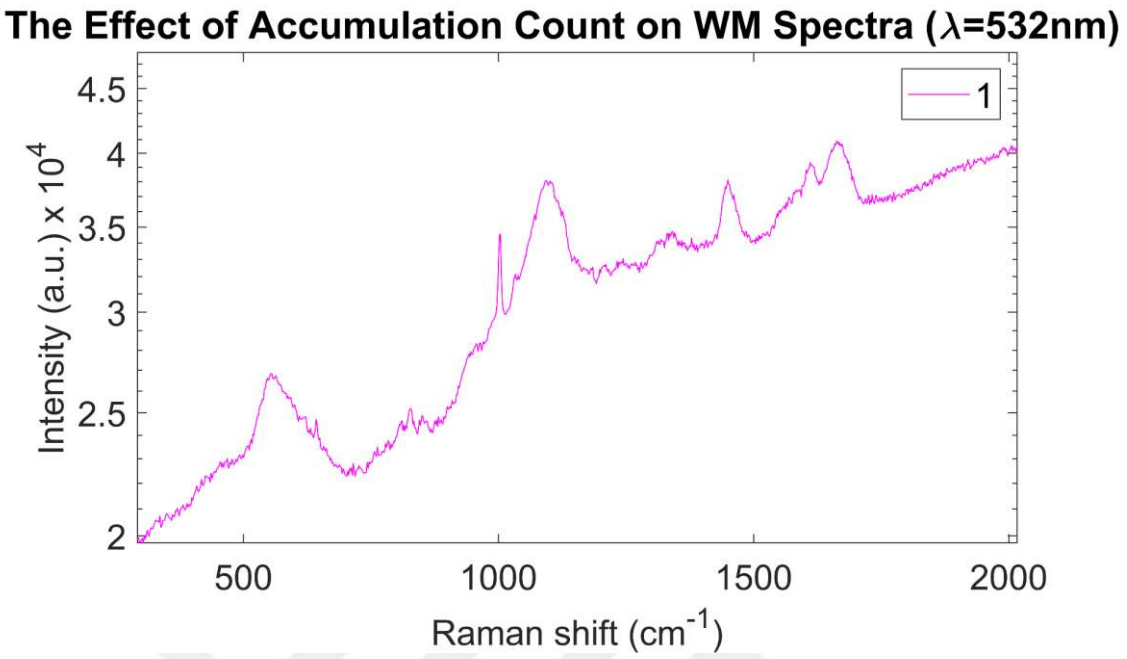


Figure 3.24. The effect of laser accumulation count on white matter (WM) spectra on the glass slide (λ=532nm)

3.3.2 Optimization of Raman Measurement Parameters for 633 nm Excitation Wavelength

Laser power optimization for 633 nm excitation wavelength

Raman spectra of 10 µm-thick GB and WM tissue sections on both CaF_2 and glass substrates were acquired for laser power optimization. Laser power was varied from 1, 10, 50, and 100% (of a maximum of ~2 mW/µm^2 power density), then the spectra were recorded for each power application, same as in the previous section (3.3.1). Raman spectra of GB and WM tissue sections were acquired under 10-second laser acquisition time and 1 laser accumulation count application. The SNR of the spectra increased with the increasing laser power and resulted in the highest value when 100% laser power was applied as its maximum on either CaF_2 or the glass substrates (Figure 3.25-3.28).

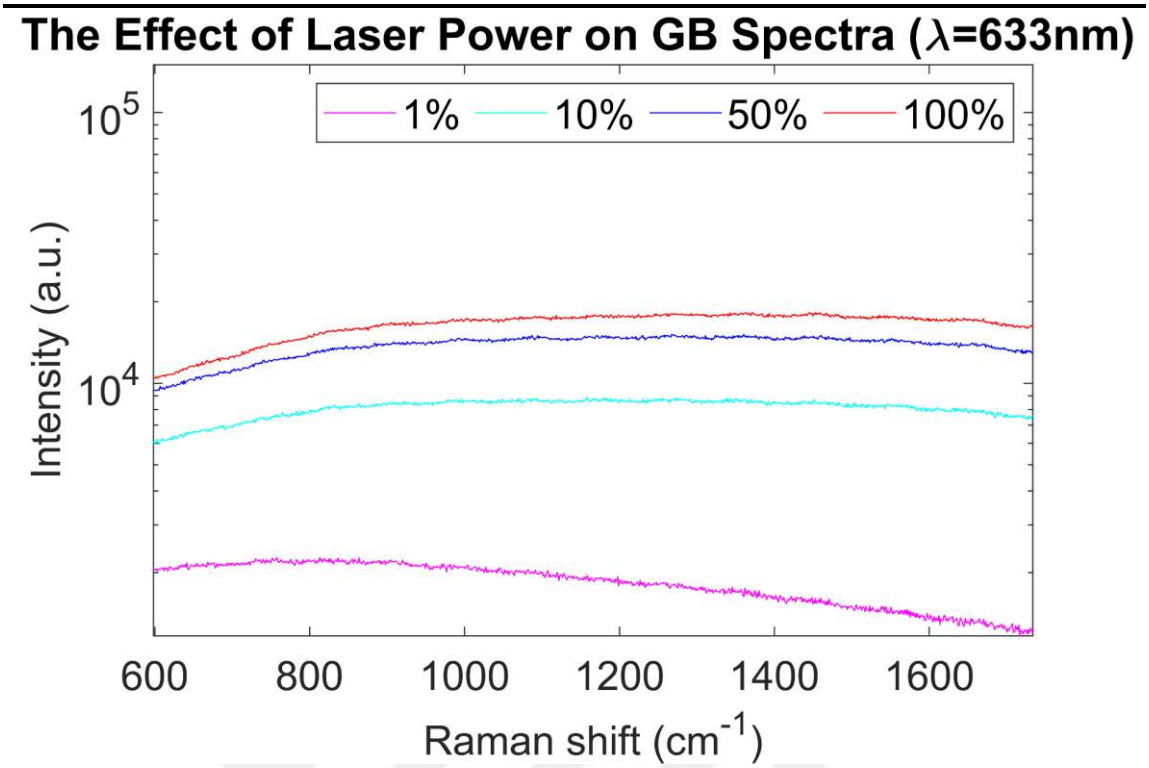


Figure 3.25. The effect of laser power on glioblastoma (GB) spectra on CaF_2 slide (λ=633nm)

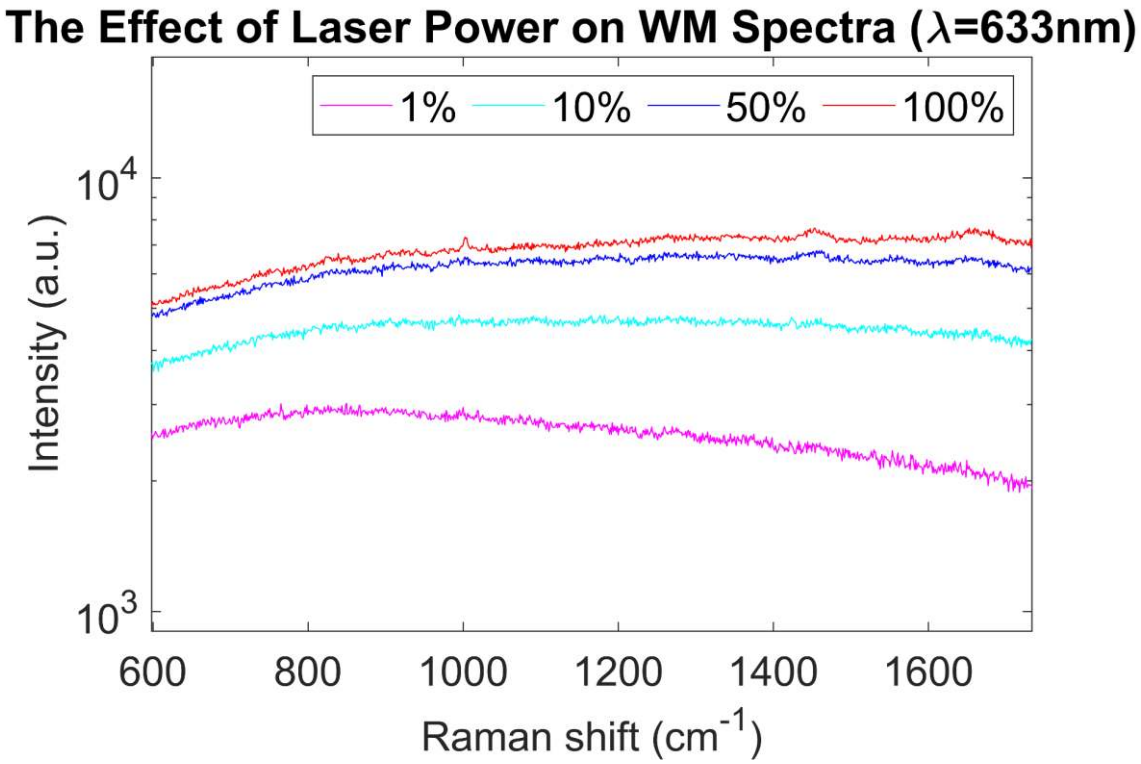


Figure 3.26. The effect of laser power on white matter (WM) spectra on CaF_2 slide (λ=633nm)

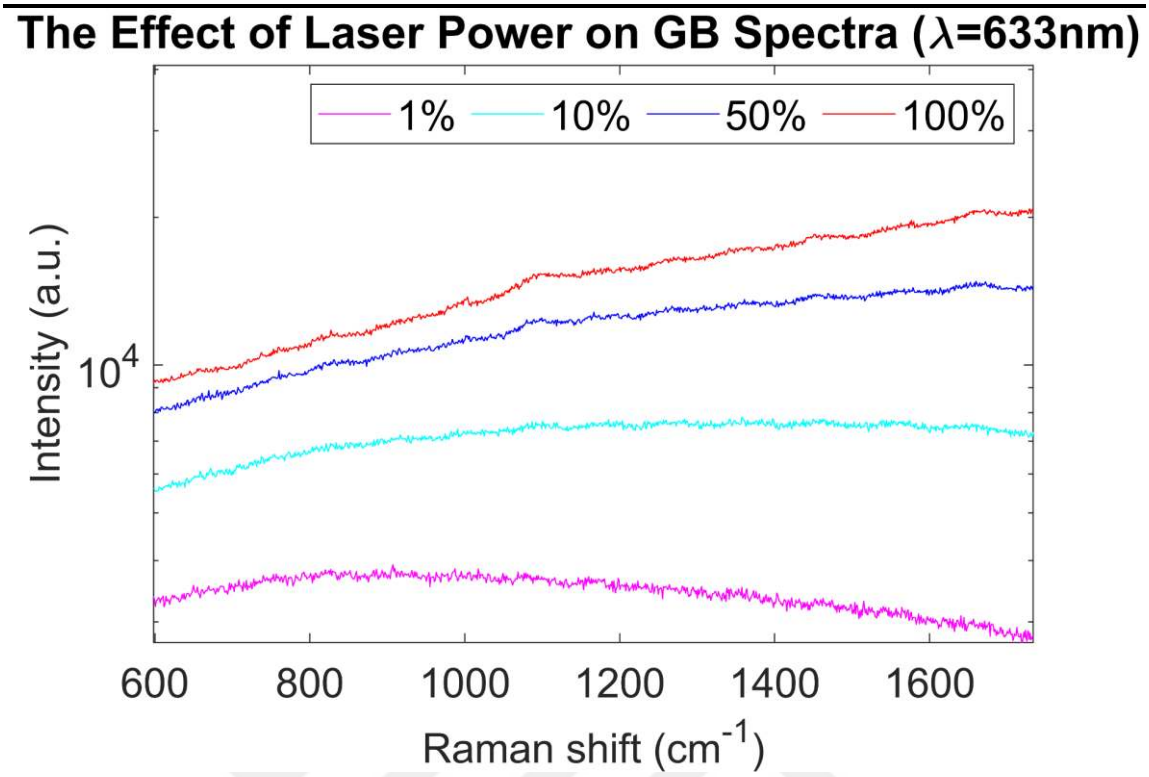


Figure 3.27. The effect of laser power on glioblastoma (GB) spectra on the glass slide (λ=633nm)

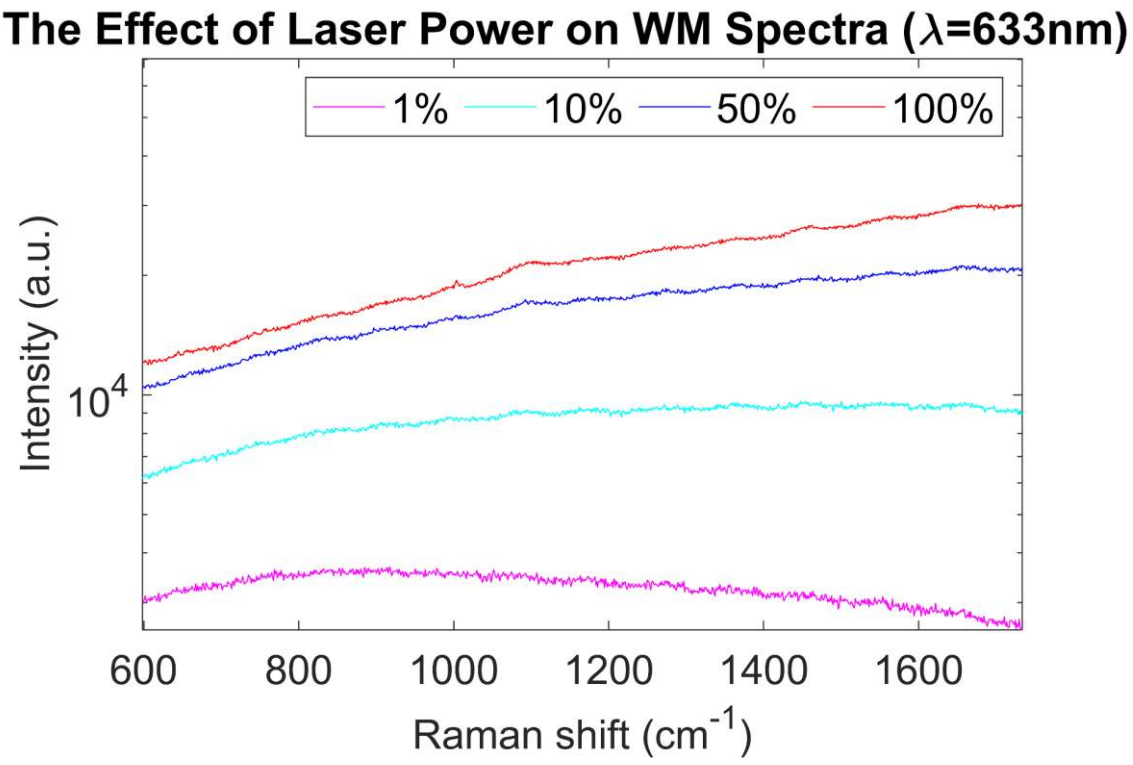


Figure 3.28. The effect of laser power on white matter (WM) spectra on the glass slide (λ=633nm)

Laser acquisition time optimization for 633 nm excitation wavelength

Raman spectra of 10 µm-thick GB and WM tissue sections on both CaF_2 and glass substrates were acquired with full laser power for the measurements in Figures 3.29-3.32 since this power level yields the highest SNR as it is described in the previous section (Figure 3.25-3.28). One laser accumulation count was applied to the GB and WM tissue sections for each measurement. The laser acquisition time was increased from 1 second to 10, 20, 30, 40, and 50 seconds for GB and WM tissue spectra measurements on both CaF_2 and glass substrates. We observed that neither the GB spectra nor the WM spectra saturated the detector in any acquisition time settings between 1-50 seconds for both CaF_2 and the glass slides (Figure 3.29-3.32). The SNR did not significantly increase after 10 seconds of exposure. These results indicate that 10-second accumulation is needed for good spectral contrast and rapid measurement (Figure 3.29-3.32).

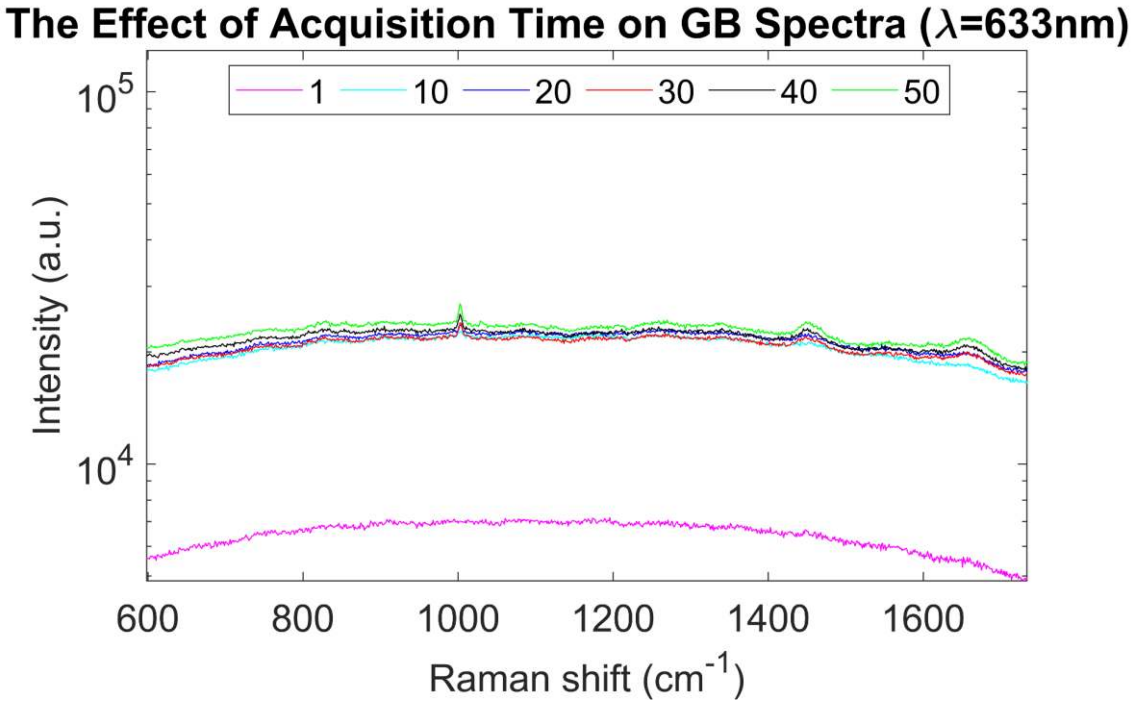


Figure 3.29. The effect of laser acquisition time on glioblastoma (GB) spectra on CaF_2 slide (λ=633nm)

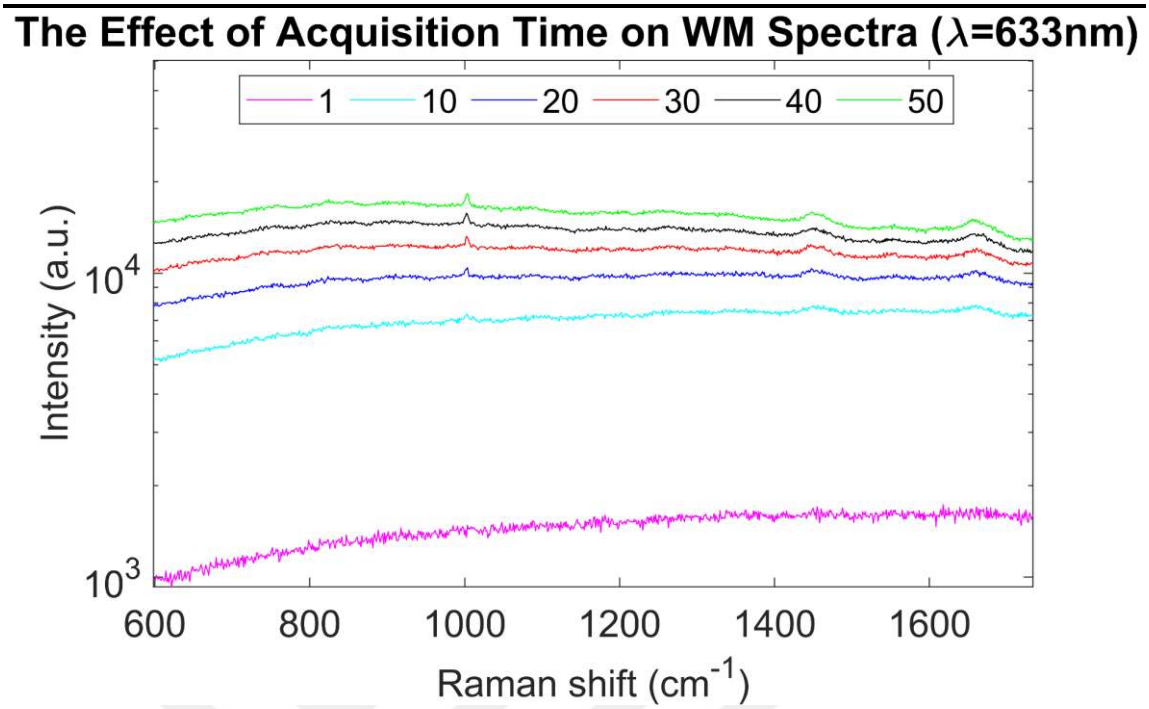


Figure 3.30. The effect of laser acquisition time on white matter (WM) spectra on CaF_2 slide (λ=633nm)

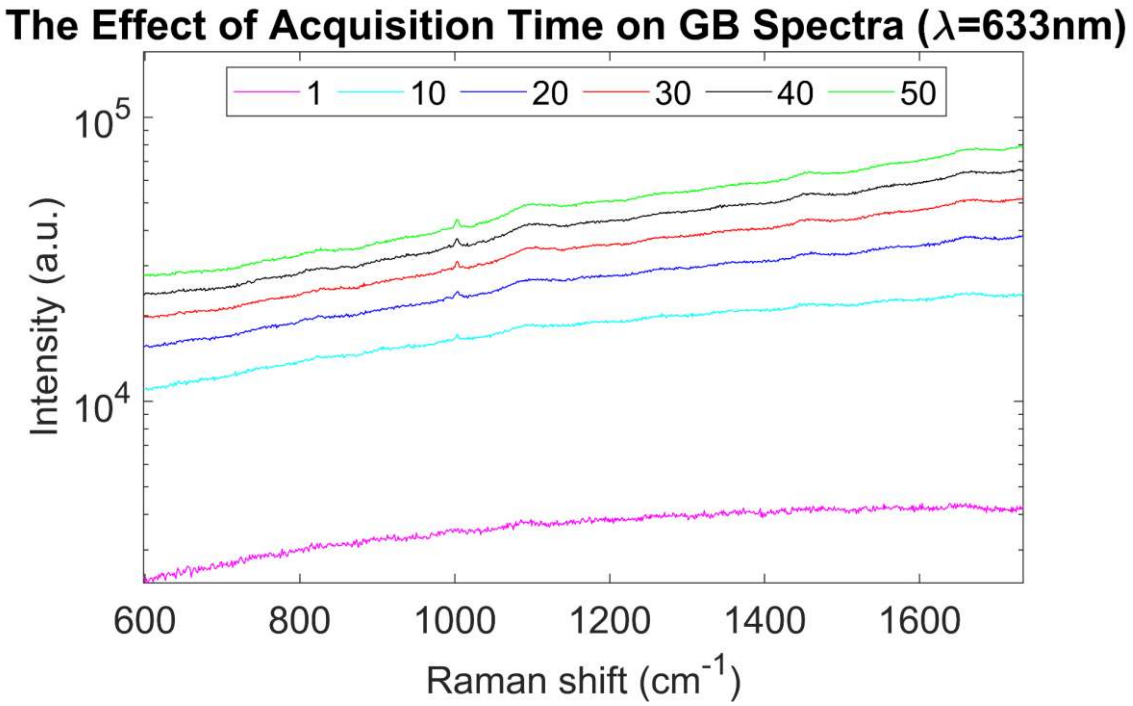


Figure 3.31. The effect of laser acquisition time on glioblastoma (GB) spectra on the glass slide (λ=633nm)

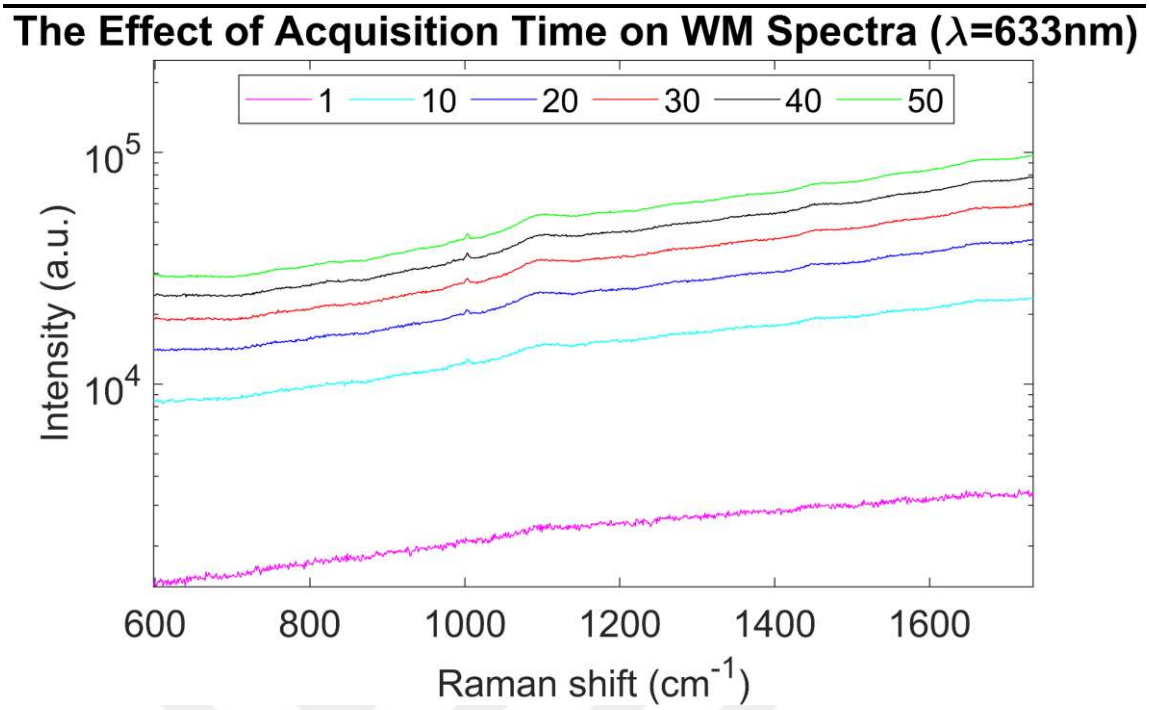


Figure 3.32. The effect of laser acquisition time on white matter (WM) spectra on the glass slide (λ=633nm)

Laser accumulation count optimization for 633 nm excitation wavelength

To investigate the optimal accumulation count, Raman spectra of 10 µm-thick GB and WM tissue sections on both CaF_2 and glass substrates were acquired under full laser power and 10 seconds of laser acquisition (Figure 3.33-3.36) since these parameters yielded the highest SNR as previously presented. (Figure 3.25-32) The accumulation count was varied from 1 to 3, 5, 10, and 20. We suggest applying no more than 3 accumulation counts as the SNR did not significantly change after a point. For rapid measurements, 1 or 3 accumulation count would be adequate for both substrates. The glass substrate showed a higher Raman background than the CaF_2 substrate (Figure 3.33-3.36).

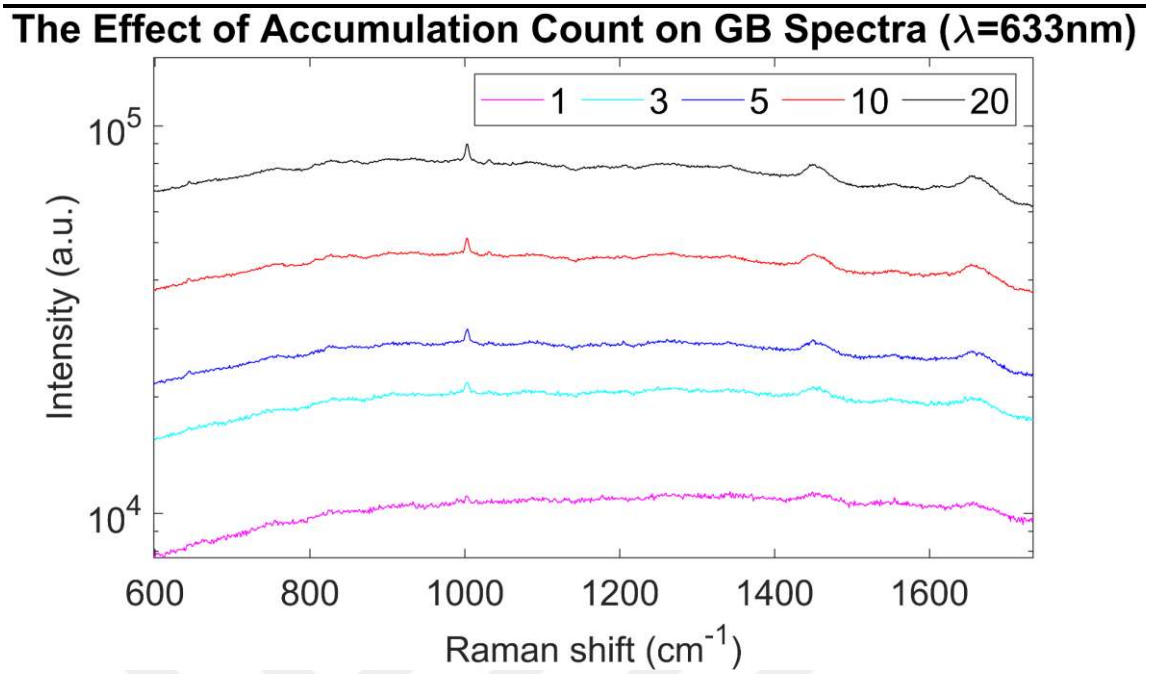


Figure 3.33. The effect of laser accumulation count on glioblastoma (GB) spectra on CaF_2 slide (λ=633nm)

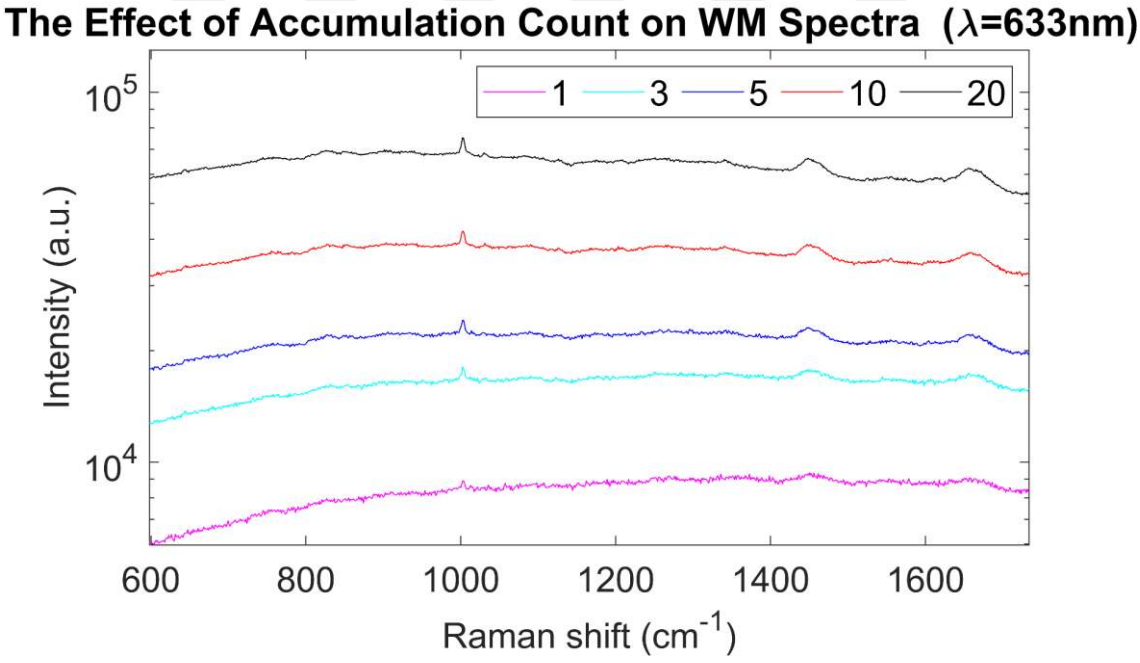


Figure 3.34. The effect of laser accumulation count on white matter (WM) spectra on CaF_2 slide (λ=633nm)

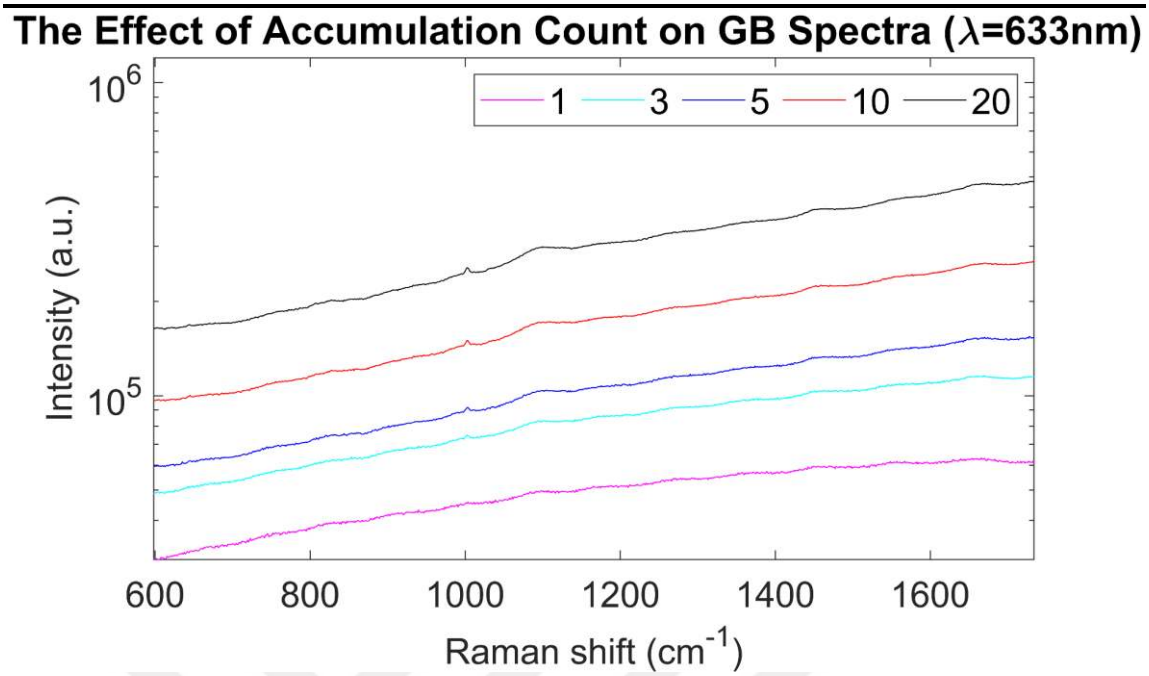


Figure 3.35. The effect of laser accumulation count on glioblastoma (GB) spectra on the glass slide (λ=633nm)

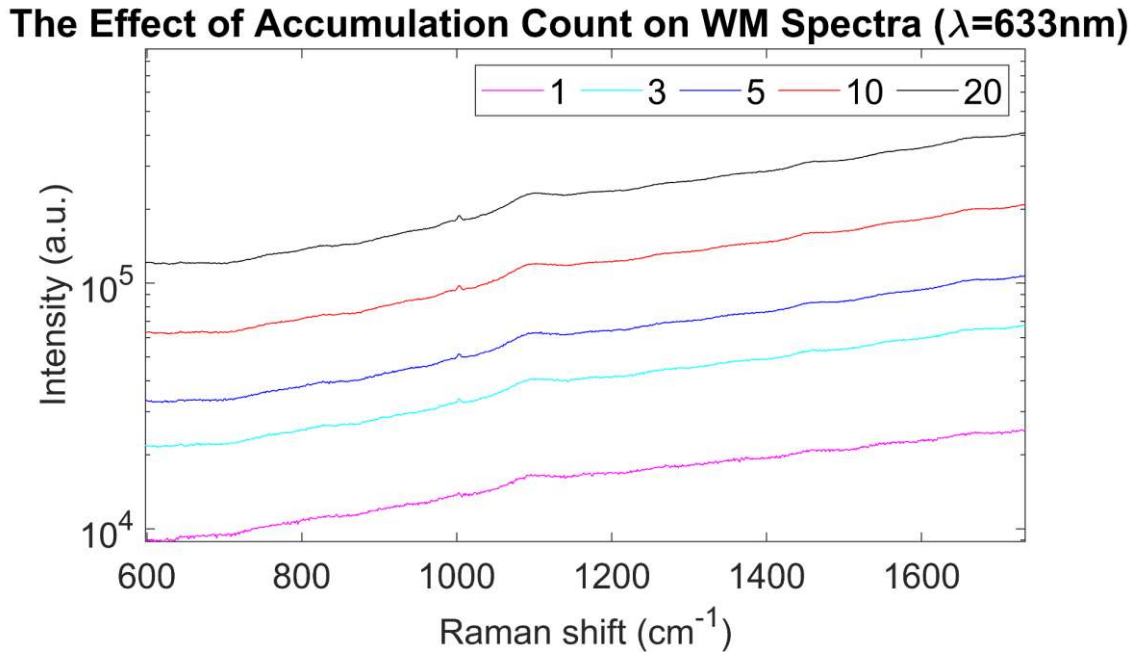


Figure 3.36. The effect of laser accumulation count on white matter (WM) spectra on the glass slide (λ=633nm)

3.3.3 Optimization of Raman Measurement Parameters for 785 nm Excitation Wavelength

Laser power optimization for 785 nm excitation wavelength

The effects of different sets of measurement parameters on GB and WM regions on the same tissue sections are demonstrated using 785 nm excitation wavelength in Figures 3.37-3.48. The Raman spectra were acquired for 10 µm-thick GB and WM sections on both CaF_2 and the glass substrates by increasing the laser power from 1, 10, 50, and 100% (of a maximum of ~12.22 $mW \cdot \mu m^{-2}$ power density). For each of these measurements among Figures 3.37-3.40, 10-second laser acquisition and 1 accumulation count were applied. Increasing the laser power resulted in higher SNR for both GB and WM sections on CaF_2 substrates, however, due to the high Raman background, no significant changes were observed on the spectra acquired from the tissue sections mounted on the glass slide. 10% and lower laser power demonstrates low SNR and requires intensive data analysis for peak detection and discrimination of different types of tissues. 50% and 100% laser power provided the ideal Raman spectra (Figure 3.37-3.40).

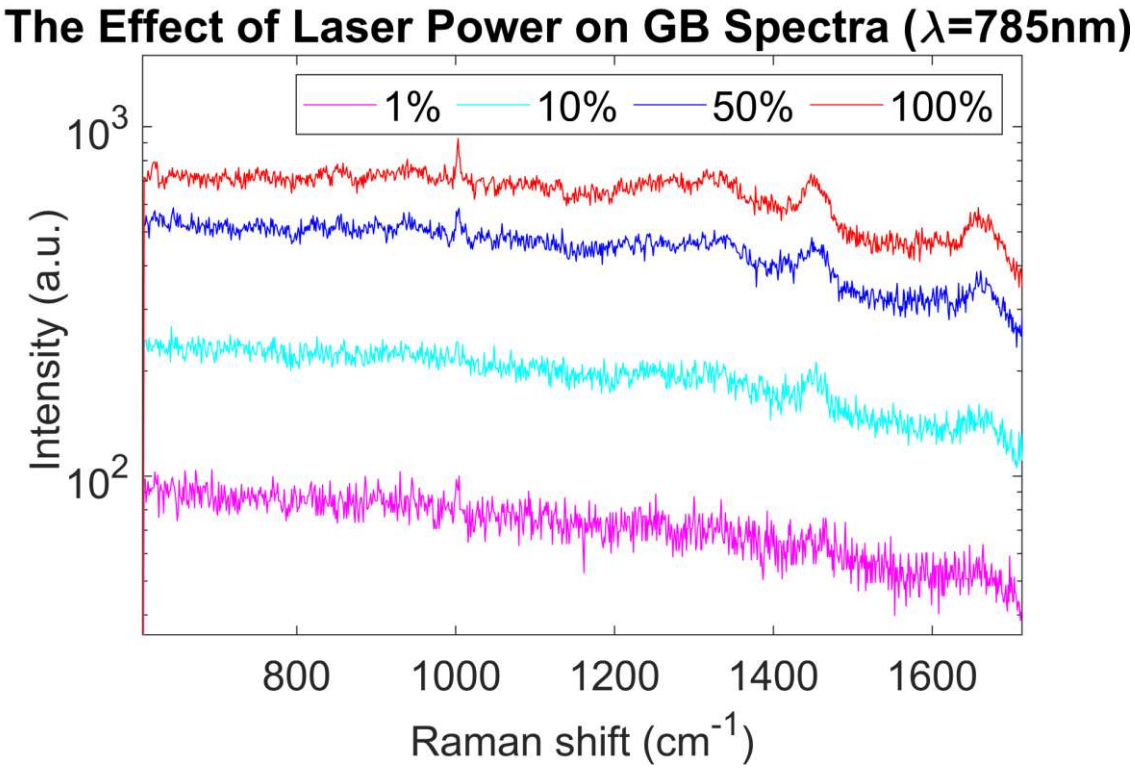


Figure 3.37. The effect of laser power on glioblastoma (GB) spectra on CaF_2 slide (λ=785nm)

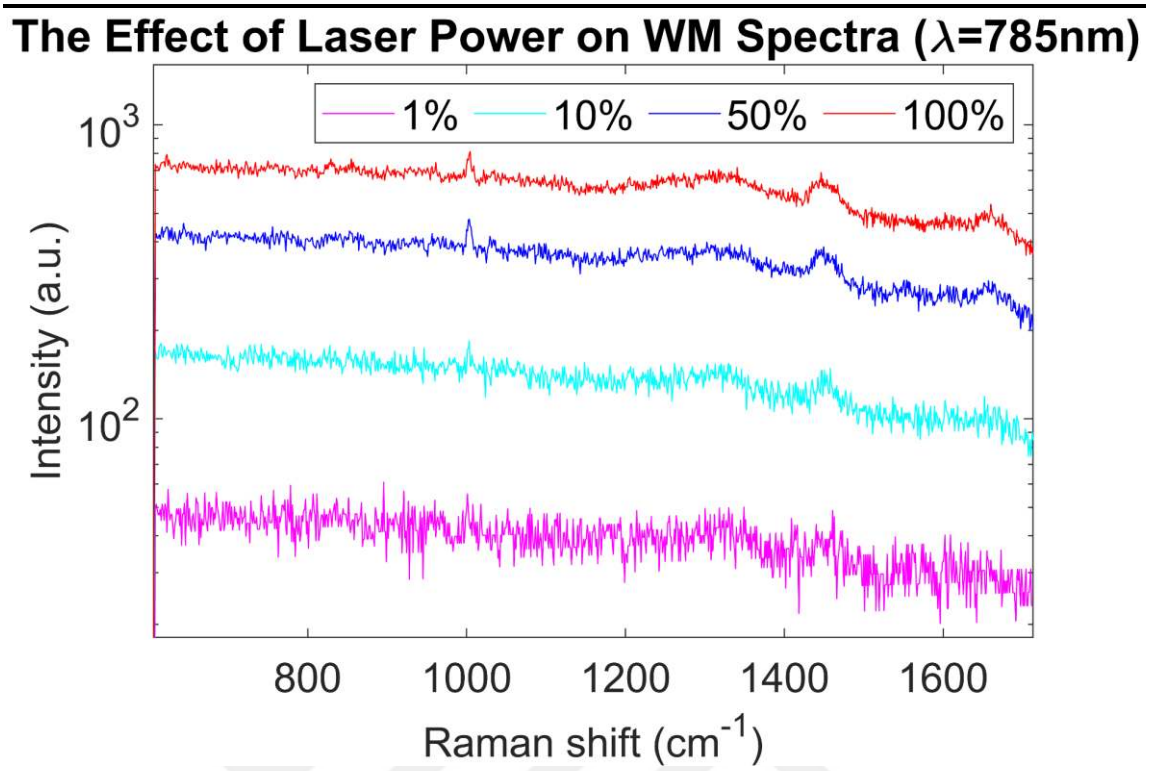


Figure 3.38. The effect of laser power on white matter (WM) spectra on CaF_2 slide (λ=785nm)

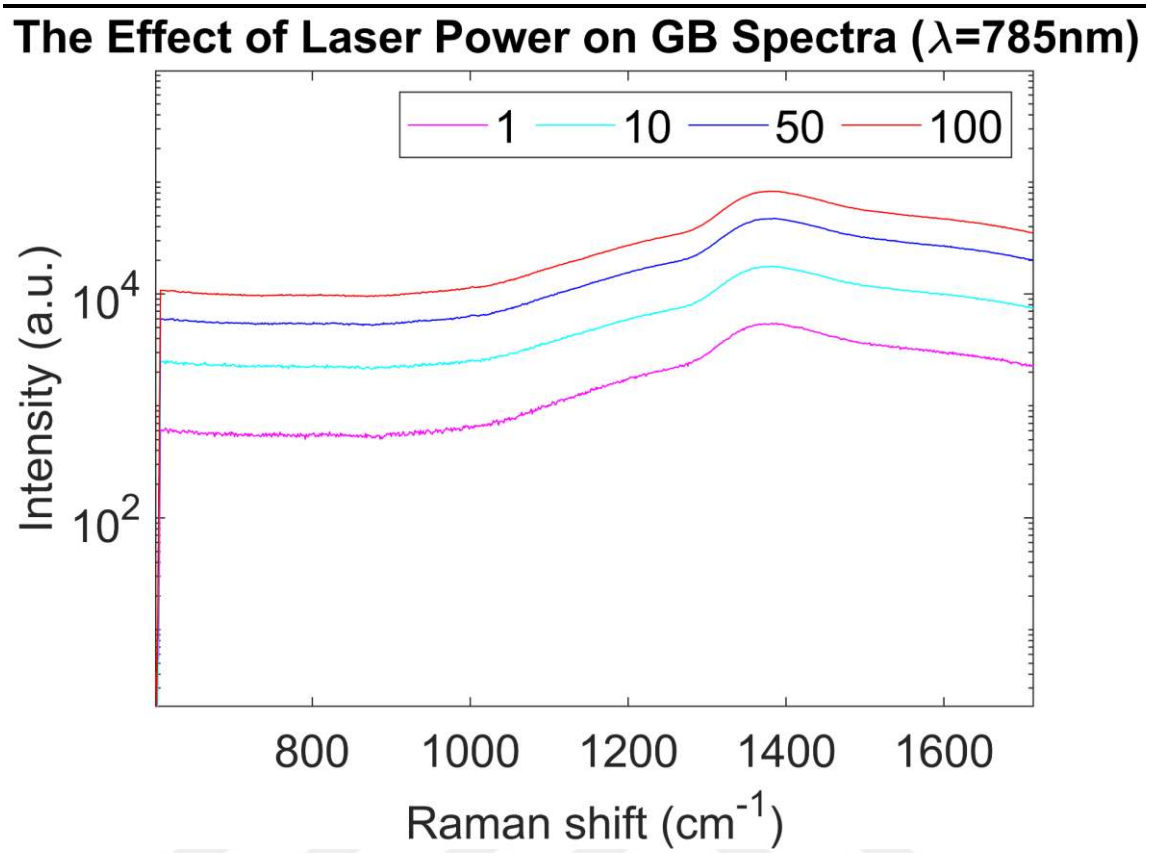


Figure 3.39. The effect of laser power on glioblastoma (GB) spectra on the glass slide (λ=785nm)

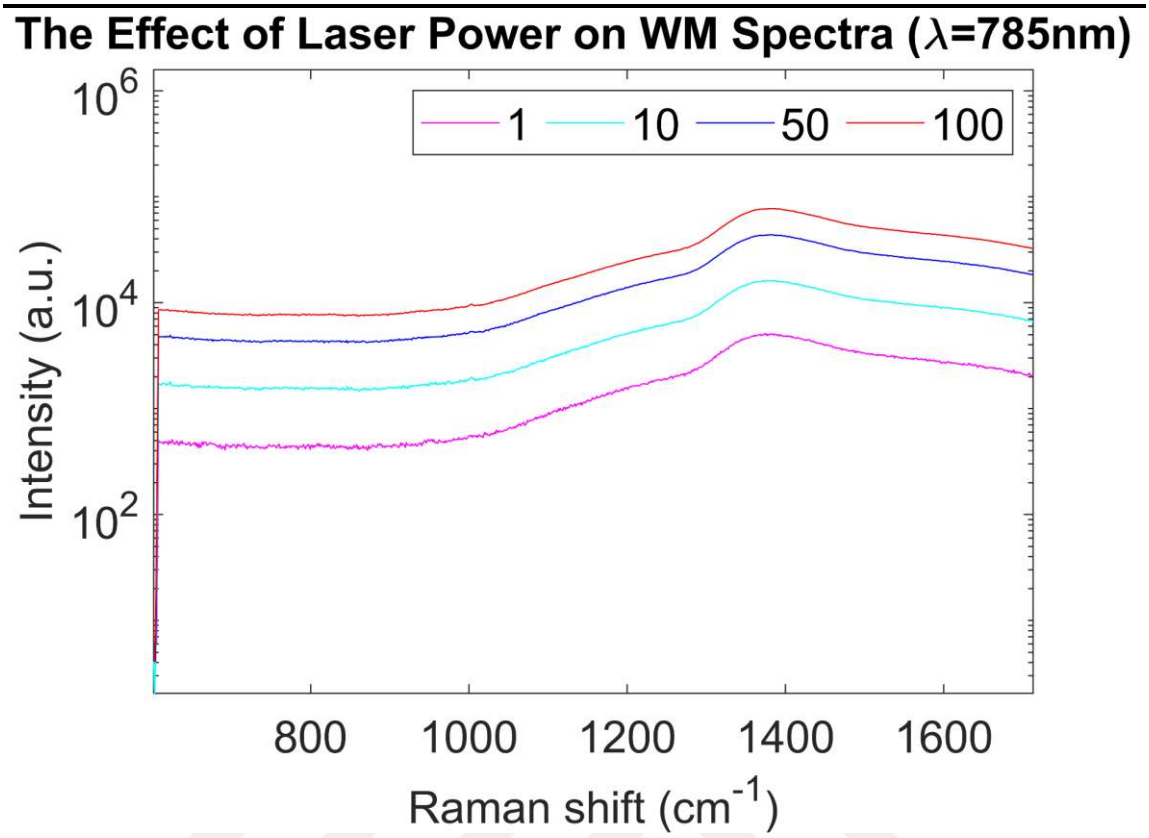


Figure 3.40. The effect of laser power on white matter (WM) spectra on the glass slide (λ=785nm)

Laser acquisition time optimization for 785 nm excitation wavelength

Based on the results in Figures 3.37-3.40, the accumulation count was tested with 100% of laser power. Full laser power and 1 accumulation count were applied for the measurements of the spectra in Figure 3.41-3.44, since this power level and accumulation count yielded the highest SNR and the most rapid measurement, respectively. GB and WM spectra were recorded by the acquisition times were increased from 1 second to 10, 20, 30, 40, and 50 seconds. The spectra acquired from GB and WM tissue sections on the CaF_2 substrates indicated that 10-second accumulation is needed for good spectral contrast. Acquisition time longer than 10 seconds demonstrated better and more stable spectra, lowering the fluorescent background. Above 10 seconds laser acquisition did not show a significant change in the fluorescent background even though those acquisitions maximize the SNR (Figure 3.41 and 3.42). The spectra of GB and WM sections on the glass slide saturated the detector after 10 seconds of laser acquisition. Therefore, we suggest using 10 seconds of laser acquisition for good spectral contrast as fluorescent

background did not significantly change for the tissue sections on CaF_2 substrate (Figure 3.41 and Figure 3.42) and saturated the detector for the tissue sections on the glass substrate (Figure 3.43 and 3.44) in longer laser acquisitions than 10 seconds.

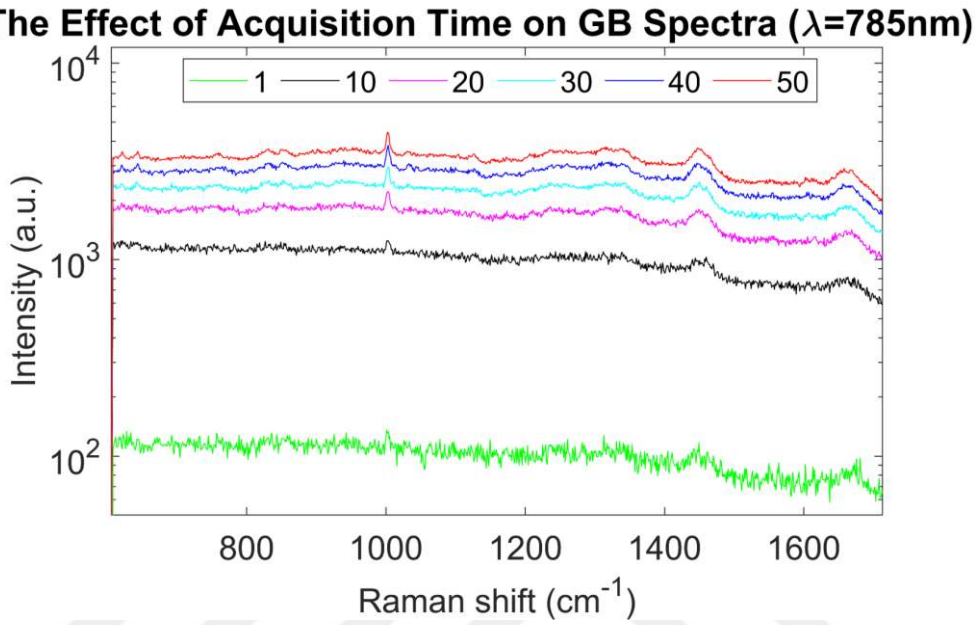


Figure 3.41. The effect of acquisition time on glioblastoma (GB) spectra on CaF_2 slide (λ=785nm)

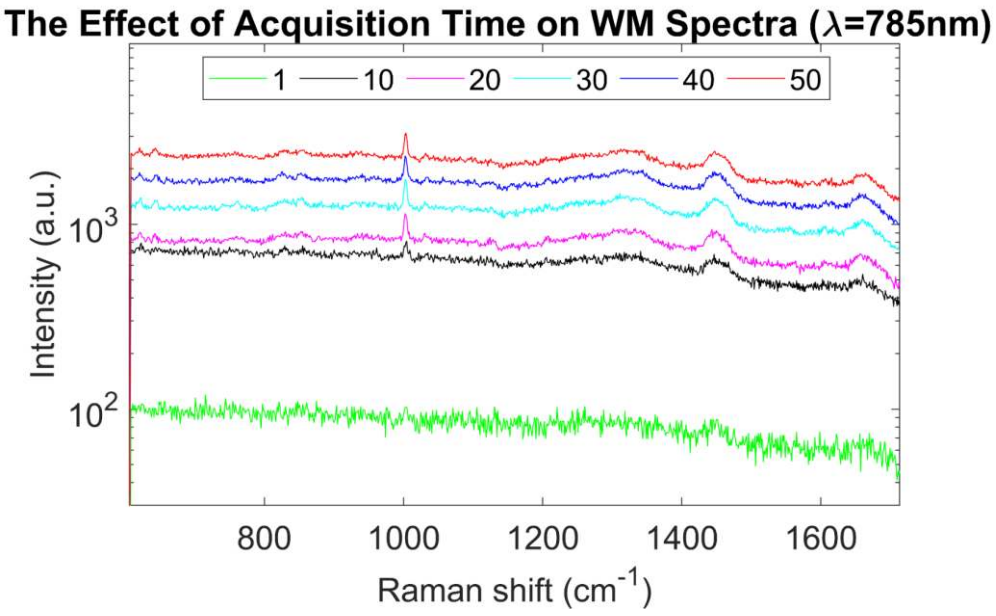


Figure 3.42. The effect of acquisition time on white matter (WM) spectra on CaF_2 slide (λ=785nm)

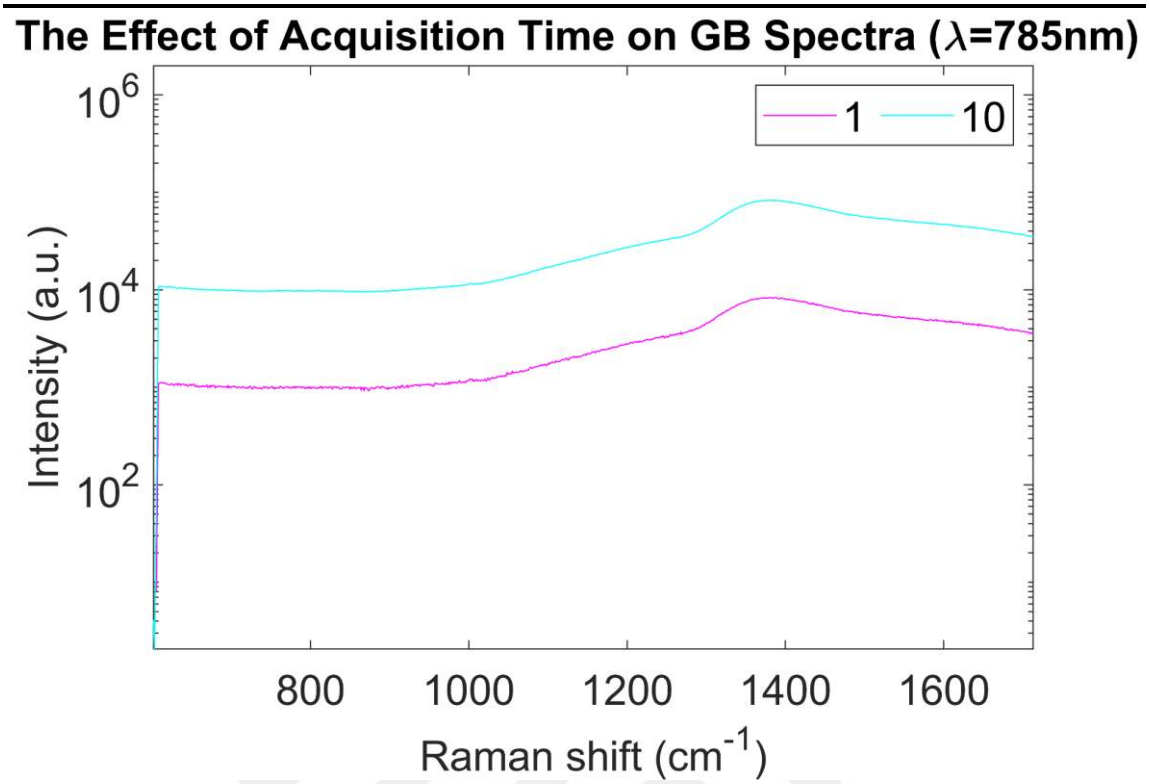


Figure 3.43. The effect of acquisition time on glioblastoma (GB) spectra on the glass slide (λ=785nm)

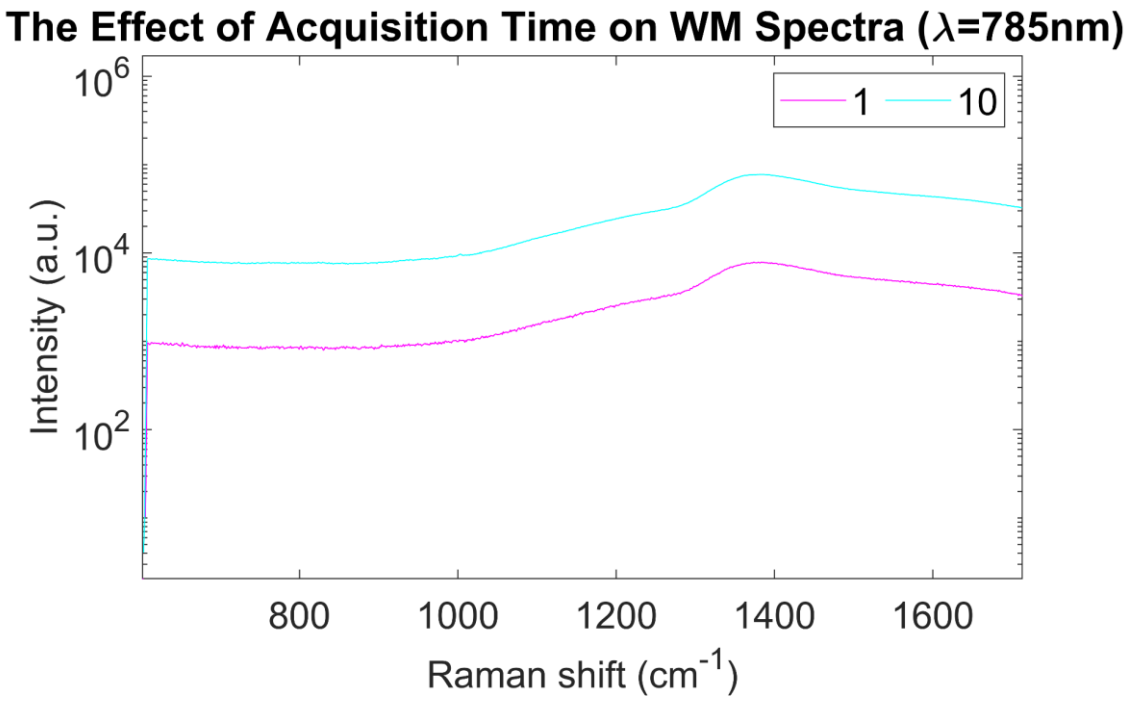


Figure 3.44. The effect of acquisition time on white matter (WM) spectra on the glass slide (λ=785nm)

Laser accumulation count optimization for 785 nm excitation wavelength

Laser accumulation time was optimized for GB and WM spectra by increasing the accumulation count from 1 to 3, 5, 10, and 20 scans, under previously determined optimal parameters: 100% laser power and 10 seconds acquisition time. Increasing accumulation to three counts demonstrated more stable spectral data by reducing the fluorescence background. Since no more than three accumulation count and 10 seconds of acquisition time are needed for high-SNR spectra and rapid spectral measurements for the tissue sections mounted on the CaF_2 substrates, we suggest no more than 3 accumulation counts be applied for rapid measurements with good spectral contrast (Figures 3.45-3.46). For the tissue sections mounted on the glass substrates, the spectra saturated the detector when more than 1 laser accumulation was applied. Therefore, we suggest using 1 accumulation count to avoid any possible detector saturation (Figures 3.47-3.48).

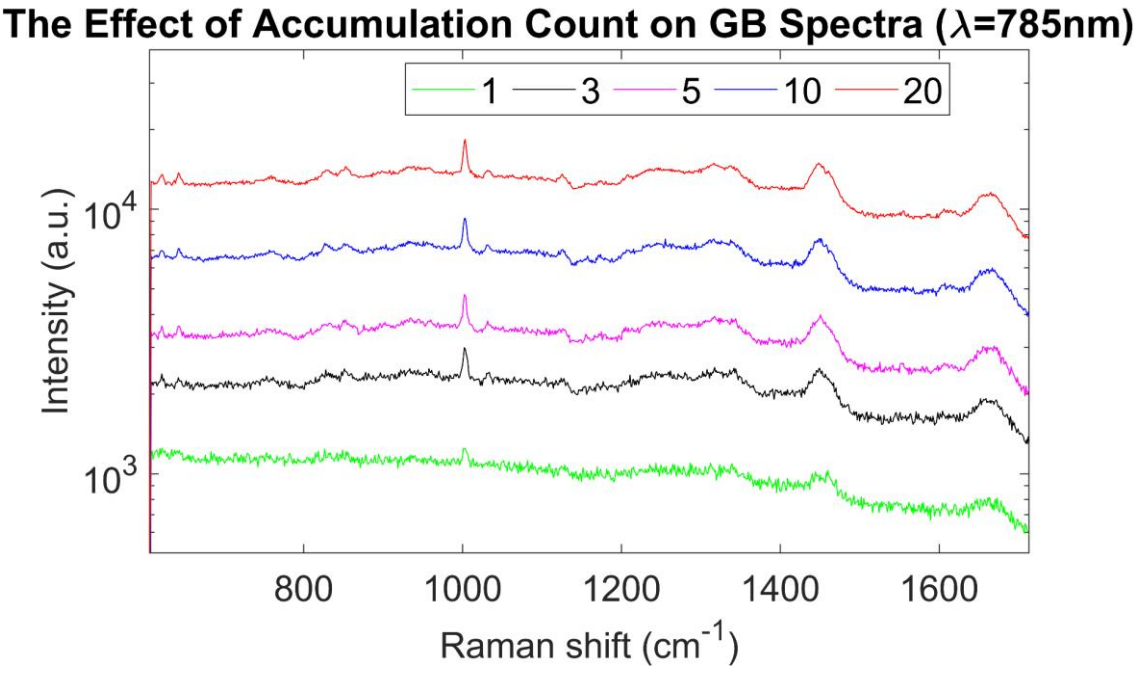


Figure 3.45. The effect of accumulation count on glioblastoma (GB) spectra on CaF_2 slide (λ=785nm)

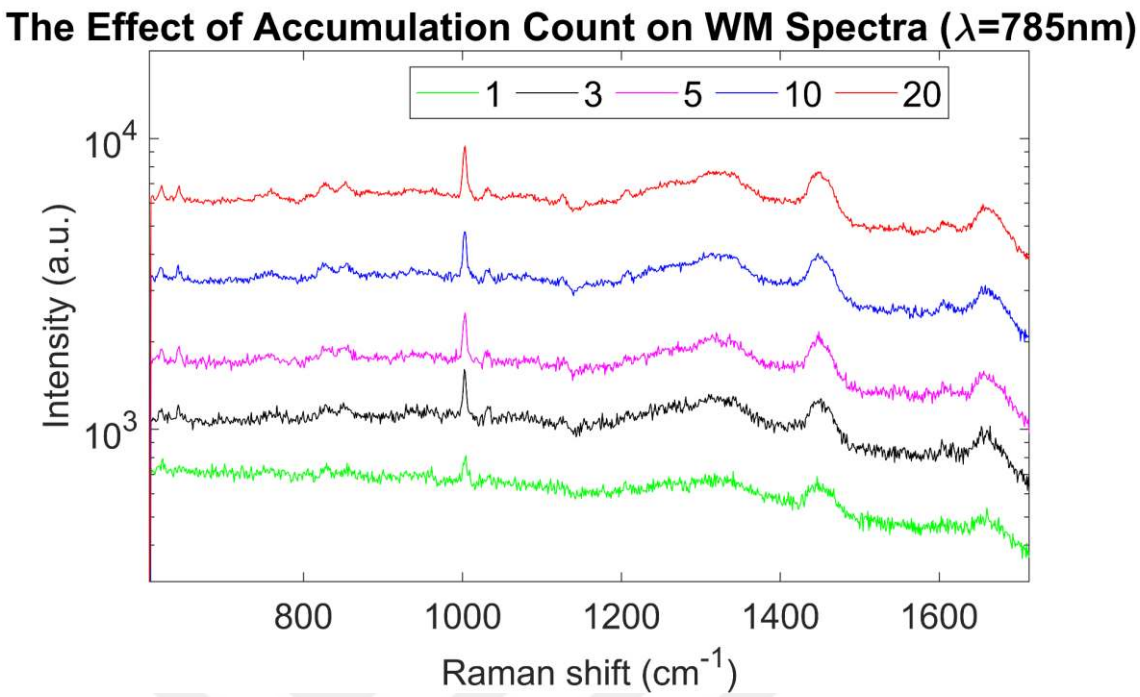


Figure 3.46. The effect of accumulation count on white matter (WM) spectra on CaF_2 slide (λ=785nm)

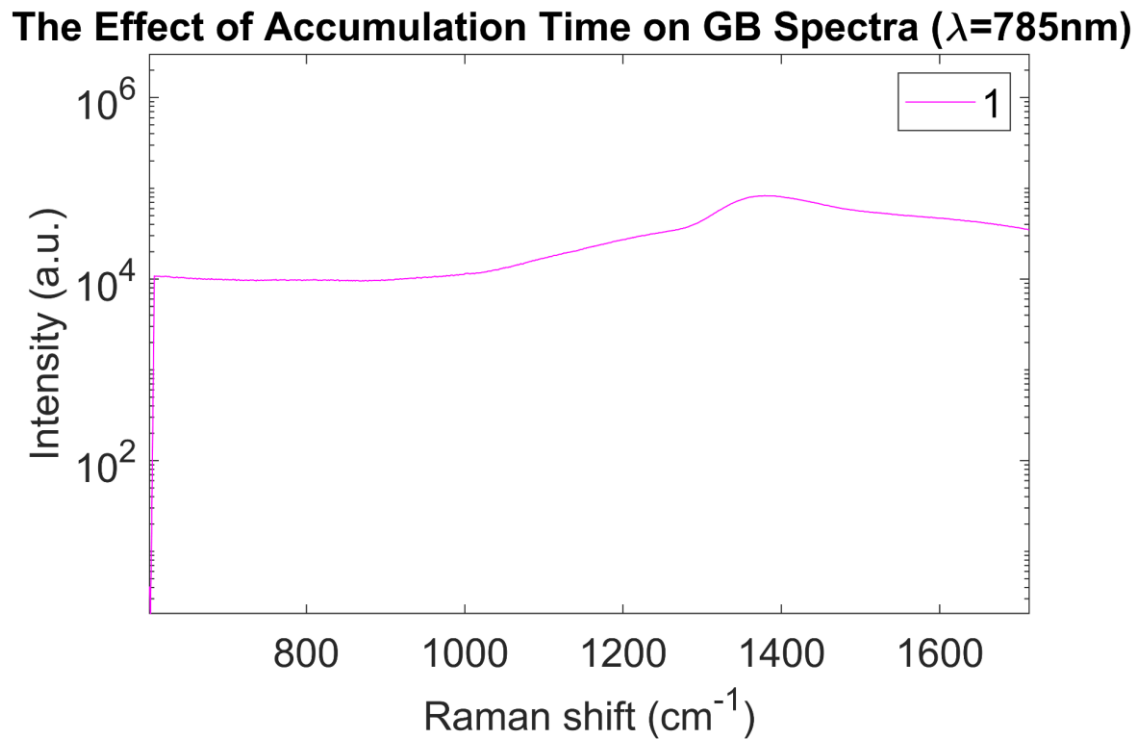


Figure 3.47. The effect of accumulation count on glioblastoma (GB) spectra on the glass slide (λ=785nm)

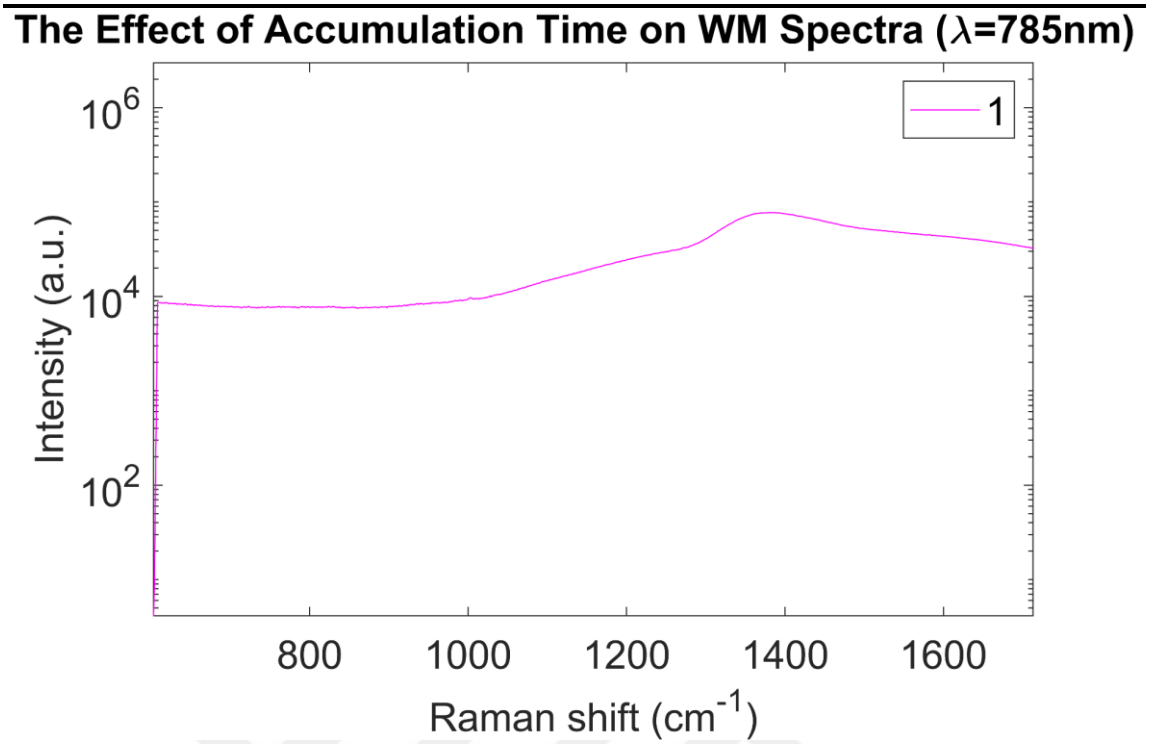


Figure 3.48. The effect of accumulation count on white matter (WM) spectra on the glass slide (λ=785nm)

3.4 The optimized FFPE tissue preparation conditions for Raman Spectroscopy

Optimization of deparaffinization protocol is highly important given that paraffin demonstrates strong Raman peaks that might suppress the tissue spectra or shade the weak narrow tissue-specific (i.e. cancerous or healthy tissue) Raman peaks, and prevent the identification of tissue types. In this section, we developed two deparaffinization protocols (Figure 2.2) using p-Xylene [85] or Clearene™ [84] solvents to achieve the highest SNR Raman spectra from FFPE tissue sections.

In the histopathology laboratories, the most frequently used paraffin clearing agent is p-Xylene, an aromatic hydrocarbon-based solvent [85]. Despite the utility of p-Xylene, its high toxicity (to reproductive or target organs and aquatic toxicity), the risk of carcinogenicity, explosion limit (lower explosion limit is 1.1% and upper explosion limit is 7.0%), and low flash point (25.0 °C in a closed cup) [85] prompted us to substitute p-Xylene with more environmentally benign

and non-hazardous solvents. Thus, we suggest the replacement of p-Xylene with Clearene™, a terpene-based organic solvent. Clearene™ is an effective clearing agent as much as p-Xylene in clearing paraffin residuals. Furthermore, no carcinogenicity was reported related to the Clearene™ use and it has a higher flash point (50.0°C) with less toxicity when compared to Xylene [84, 85]. In addition to these advantages of Clearene™, its cost-effectiveness, and prevention of stain fading features make Clearene™ more advantageous and useful as a deparaffinization agent for the tissue sections in both histopathological or Raman analysis [84].

In this study, we developed a protocol using two clearing agents and tested their viability in retaining the morphological integrity of the tissue, and not giving rise to staining impairments in the tissue sections. We observed that H&E-stained sections retained their morphological integrity and did not result in staining impairments that might prevent the histopathological evaluation of the tissue sections after both Clearene™ and p-Xylene-based deparaffinization methods. It is demonstrated that either p-Xylene or Clearene™ based protocols given in Figure 2.2 can be used for effective deparaffinization with no significant morphological differences in the tissue and structural differences between their effectivities in eliminating paraffin. The deparaffinization protocol that uses Clearene™ preserves the tissue morphology as effectively as the protocol that uses p-Xylene as a clearing solvent. This protocol can be used for the preparation of FFPE tissue sections for RS measurements and the routine histopathological analysis of tissue sections using various staining (i.e. H&E) and light microscopy (Figure 3.49).

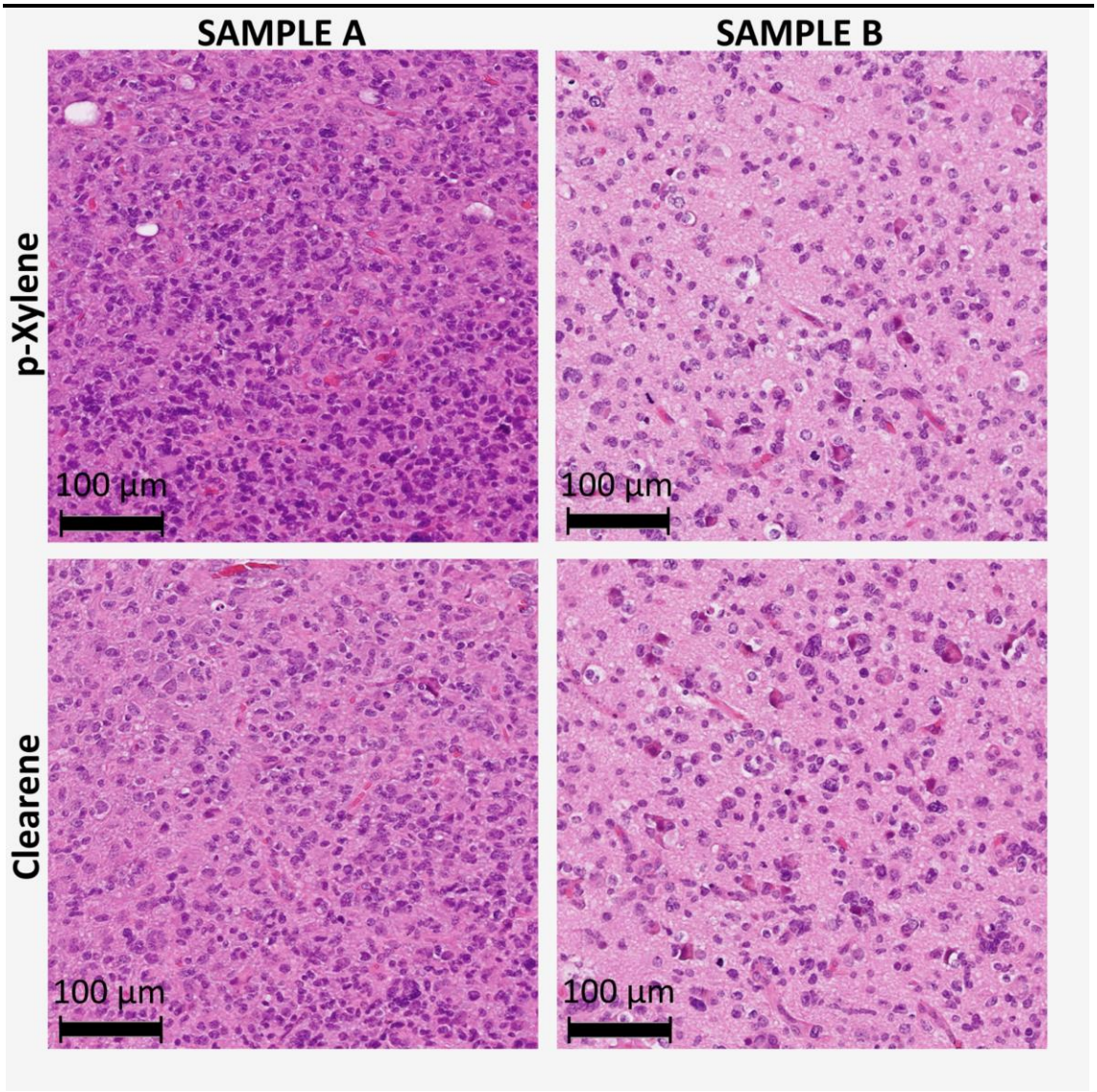


Figure 3.49. The effect of p-Xylene and Clearene™-based deparaffinization methods on tissue morphology. 5 µm-thick two Hematoxylin and Eosin (H&E)-stained glioblastoma sections after deparaffinization with different clearing solvents

After the demonstration of the effectiveness of both p-Xylene and Clearene™ based protocols, two groups of FFPE brain tissue sections were deparaffinized using p-Xylene or Clearene™-based protocols (Figure 2.2) and the Raman spectra of these sections were recorded. The Raman spectra of an FFPE GB block (GB tissue preserved in paraffin block) and an empty paraffin block that does not have any tissue were also recorded. It is observed that the GB spectrum acquired from the FFPE GB block was suppressed by the paraffin spectrum and observed similar with paraffin spectra acquired from the empty paraffin block. The

Raman spectra of deparaffinized GB sections with either p-Xylene or Clearene™ solvents demonstrated that paraffin spectra were removed from GB spectra. The Raman spectra of the tissue sections that were deparaffinized with p-Xylene based protocol and Clearene™ based protocol, paraffin spectra, and paraffin-embedded GB tissue spectra were then compared to test the effectiveness of deparaffinization solvents. According to the Raman measurements, we observed that both protocols eliminated paraffin residuals from the tissue sections (Figure 3.50).

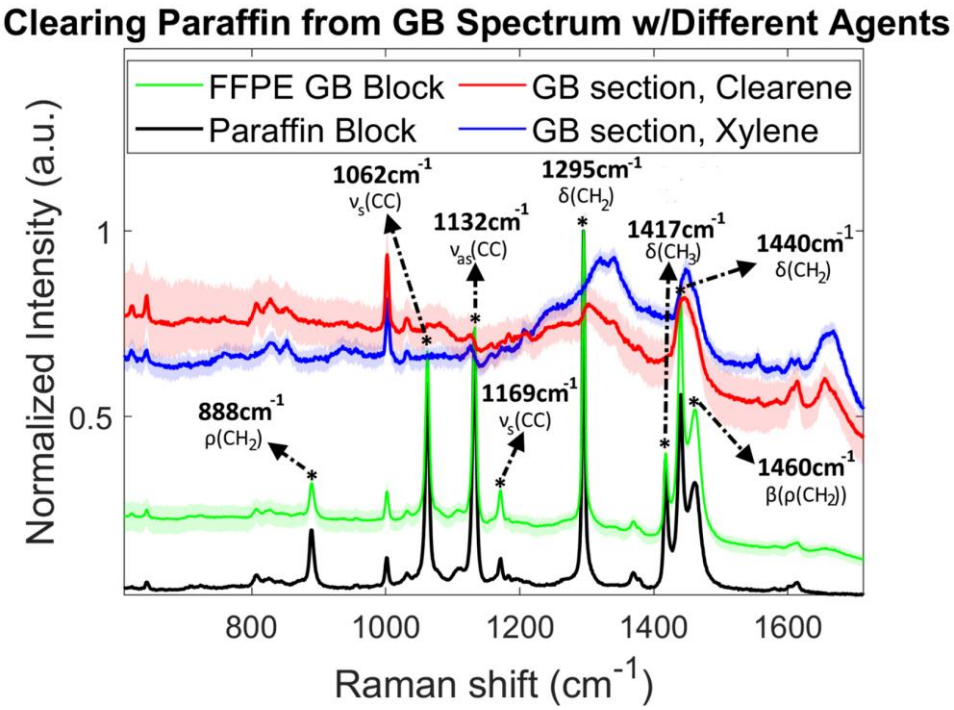


Figure 3.50. Clearing Paraffin from Raman spectrum of glioblastoma (GB) with Different Agents: p-Xylene, Clearene™. The figure shows the comparison of Raman spectra of paraffin, Formalin-Fixed Paraffin-Embedded (FFPE) tissue, deparaffinized FFPE GB tissue sections with Clearene™-based and p-Xylene based protocols. GB spectrum is suppressed by the paraffin spectra acquired from FFPE GB block and paraffin residuals mostly eliminated from the spectra of deparaffinized (w/ p-Xylene and Clearene™-based protocols) GB sections. All spectra were acquired under 100% laser power (100mW input), 10 seconds laser acquisition, and 1 laser accumulation on CaF_2 Raman substrate. (* Asterisk indicates Raman peaks attributed to paraffin in the literature.) Peak assignments are according to references [75, 95-97] (see: Table 3.1).

Table 3.1 Raman shift assignments for paraffin spectrum [75, 95-97]

Raman shift (cm^{-1})	**Raman shift assignments**
888	$\rho(CH_2)$
1062	$\nu_s(CC)$ – CC skeletal stretch
1132	$\nu_{as}(CC)$ – CC stretch
1169	$\nu_s(CC)$ – CC stretch
1295	$\delta(CH_2)$ – CH_2 deformation
1417	$\delta(CH_3)$ – CH_2 deformation
1440	$\delta(CH_2)$ – CH_2 deformation
1460	$\beta(\rho(CH_2))$ – CH_2 bending and rocking

3.5 Discussion

Retrospective Raman spectral analysis of FFPE tissue sections is challenging because of the fluorescent background of the tissue sections, strong spectra of paraffin that suppress the tissue spectra, and highly heterogeneous material composition. To resolve these challenges, minimizing the spectral variability of the tissue sections by developing an effective and reproducible tissue preparation and Raman measurement protocols is necessary. In this thesis, we present an optimal sample preparation and Raman measurement protocol that maximizes the SNR and minimizes the variability between the tissue spectra to better classify these spectra using machine learning algorithms.

3.5.1 Optimized sample preparation protocol

While clearing the FFPE tissue sections from paraffin or formalin residuals, preserving tissue morphology and composition is vital. We demonstrate that organic solvent Clearene^{TM} can be used as a paraffin clearing agent instead of highly toxic Xylene, the most used paraffin clearing agent in histology laboratories, as Clearene^{TM} showed the equivalent performance to Xylene in eliminating paraffin originated Raman peaks from spectra while retaining tissue morphology and tissue originated Raman peaks (Figure 3.49 and Figure 3.50). We suggest the use of Clearene^{TM} as a deparaffinization agent for

retrospective histopathological analysis of archival FFPE tissues due to its safety and low cost.

3.5.2 Optimal Raman spectral measurement protocol

CaF_2 and glass Raman substrates under different 532, 633, and 785 nm excitation wavelengths result in different Raman backgrounds. CaF_2 Raman substrates demonstrated lower Raman background and better spectral quality with higher SNR (Figure 3.5-3.10). On the contrary, the glass substrate resulted in a high background, which prevents the detection of some peaks and the discrimination of tissue spectra more accurately (Figure 3.10). Despite their affordability, spectra were saturated even with the less laser power, laser acquisition time, and acquisition parameters when glass substrates were used. Therefore, CaF_2 would be the more appropriate Raman substrates to reduce fluorescent background in the spectra despite their relatively higher cost [39, 78-80], consistent with the literature.

532 nm excitation wavelength demonstrated high-contrast spectra for brain tissue sections, however, we observed morphological changes on the tissue section in some cases due to thermal degradation of the tissue structure. Therefore, 532 nm lasers may damage the sample and alter their compositions if these lasers are not used with low power and with low acquisition time and accumulation counts. 633 nm lasers triggered the fluorescent background of both tissues and the substrates. Among these three lasers, 785 nm lasers could be the most appropriate option for minimizing fluorescent background, preserving tissue composition by preventing photo-induced damage of the tissues, and maximizing SNR. Also, other near-infrared lasers such as 1064 nm [39] might be used for avoiding fluorescent background. Applying full laser power (100%, 5 $mW \cdot \mu m^{-2}$), 10 seconds laser acquisition time, 1-3 accumulation counts are suggested for high-SNR and rapid spectral measurements that could be appropriate to use in the clinics in the future.

3.6 Conclusions and Insights

We tested and demonstrated the effectiveness of our deparaffinization protocols in preserving the morphological features with our Raman spectra and H&E light microscopy results. Second, we optimized the spectral parameters for analysis of brain tissue by modifying analytical Raman measurement parameters such as minimum optimal tissue

thickness and Raman substrate choice to avoid and eliminate background signal, spectral excitation wavelength parameters, laser exposure time, scanning time, etc. (Aim-1A)

Once the sample preparation and Raman measurement protocols were optimized, tested these optimal parameters and generate specific signals from normal versus abnormal tissue types of CNS (i.e. GB vs white matter) to validate our protocols, allowing the repeatable and reproducible Raman spectral analysis of FFPE brain tissue sections. (Aim-1B)

Chapter 4:
RAMAN SPECTRAL ANALYSIS OF GLIOBLASTOMA AND BRAIN TISSUE CONSTITUENTS

4.1 Raman Spectral Comparison (RS) of Brain Tissue Types

Vibrational and label-free spectroscopic techniques such as RS are promising analytical tools with great potential in identifying the changes in the tissues in real-time with a sub-micron high spatial resolution [105]. RS is a powerful technique for molecular analysis of biological materials as it provides molecular information based on the Raman scattering of photons with bonds of the given sample [39]. In this chapter, we investigated the Raman spectra (and the corresponding molecular peaks) of FFPE GB, NC tissue sections, and normal brain tissue constituent WM, and GM tissue sections, using RS. We obtained Raman spectral profiles of each tissue type (GB, WM, GM, NC) for distinguishing these profiles (Figure 4.1-4.4).

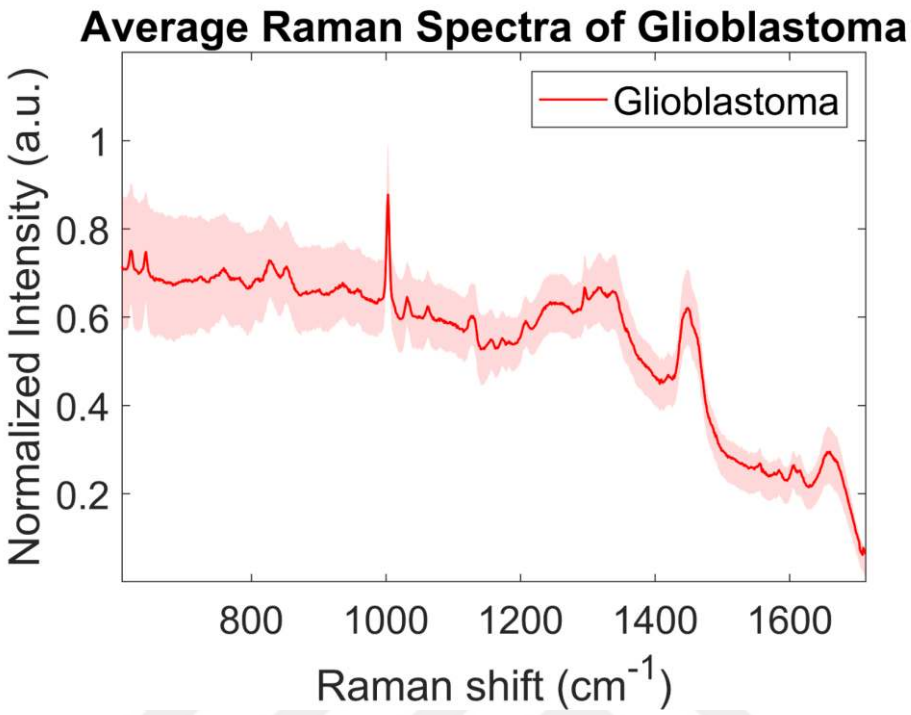


Figure 4.1. Average Spectra of Glioblastoma (GB) (n=38). Map spectra for each patient were acquired from 100 different locations on each GB tissue sections. All spectra were recorded by applying full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. Tissue sections were deparaffinized according to the Clearene^{TM}-based deparaffinization protocol (see: Chapter 2).

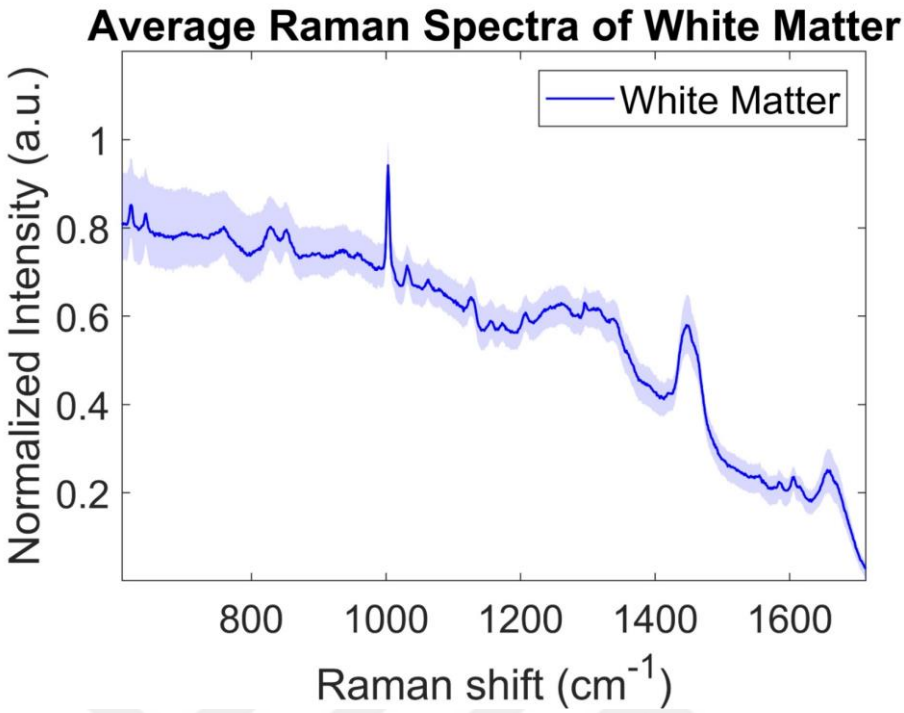


Figure 4.2. Average Spectra of White Matter (WM) (n=18). Map spectra for each patient were acquired from 100 different locations on each WM tissue sections. All spectra were recorded by applying full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. Tissue sections were deparaffinized according to the Clearene^{TM}-based deparaffinization protocol (see: Chapter 2).

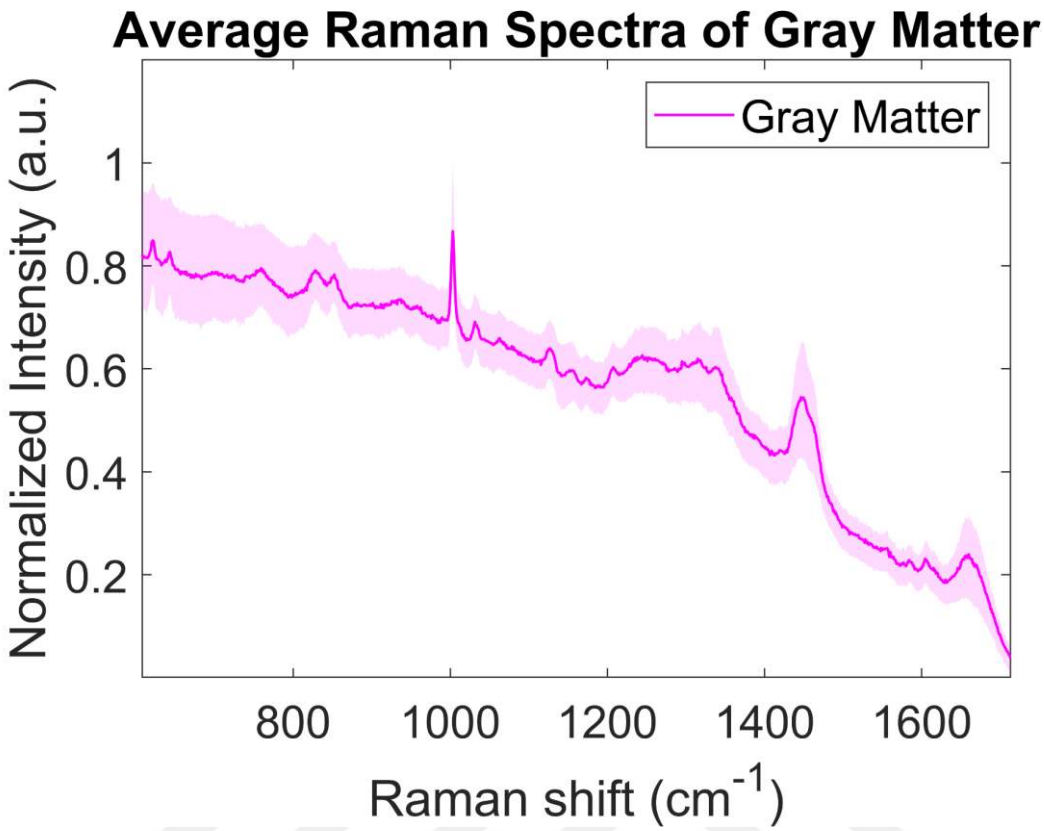


Figure 4.3. Average Spectra of Gray Matter (GM) (n=16). Map spectra for each patient were acquired from 100 different locations on each GM tissue sections. All spectra were recorded by applying full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. Tissue sections were deparaffinized according to the Clearene^{TM}-based deparaffinization protocol (see: Chapter 2).

Brain tissue consists of ~70–83% water, ~7.5–8.5% protein, and ~5–15% lipids, and CNS is composed of low lipid concentration (~5%) in GM and high lipid concentration in (~15%) WM [105]. GM mostly consists of neuronal cell bodies, unmyelinated axons, and dendrites[106-108]. Accordingly, as GM consists of a high amount of nucleus due to the neuronal cell bodies, one may expect to observe more nucleic acid and protein [108] content peaks from the Raman spectra. On the other hand, WM mostly consists of myelinated axons of the neurons[106, 108] and relatively fewer neuronal cell bodies [106] when compared to GM. As myelin sheaths are composed of 75% lipid and 25% proteins [109], it is expected to see more lipid or phospholipid [108] and some specific proteins or amino acids (i.e. myelin oligodendrocyte glycoprotein) that are exclusively expressed on

myelin sheaths and oligodendrocytes [110] and less nucleic acid content in the WM, when compared to GM.

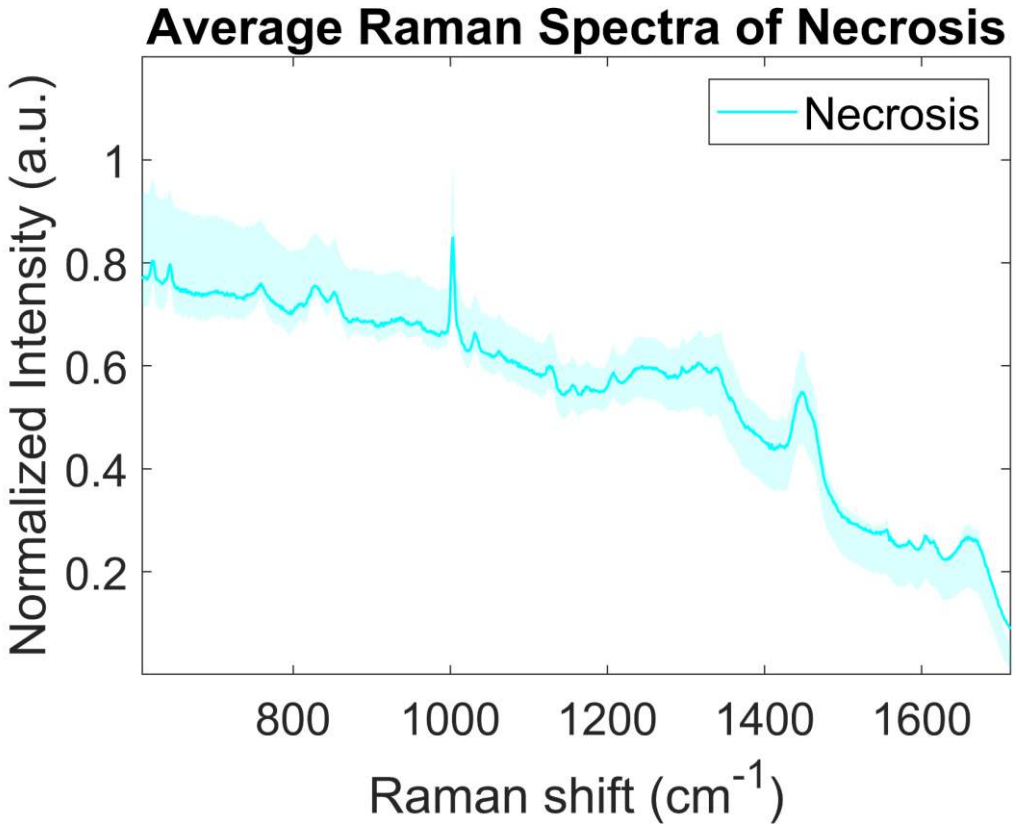


Figure 4.4. Average Spectra of Necrosis (NC) (n=19). Map spectra for each patient were acquired from 100 different locations on each NC tissue sections. All spectra were recorded by applying full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. Tissue sections were deparaffinized according to the Clearene^{TM}-based deparaffinization protocol (see: Chapter 2).

Steroids and cholesterol are the most plentiful lipid types in the brain. Phosphatidylethanolamine, phosphatidylcholine, phosphatidylserine, and phosphatidylinositol are a form of phospholipids and in addition to phospholipids, sphingomyelin, glucocerebroside, ganglioside, and sulfatide are the form of lipids that are present in the brain tissue (Table 4.1). Although the gangliosides are given in the group of 'unidentified' in Table 4.1 since their concentrations are tiny in the total brain lipid content, ganglioside concentration was reported to be six times lower in white matter than gray matter, similar with most of the lipid subtypes [105].

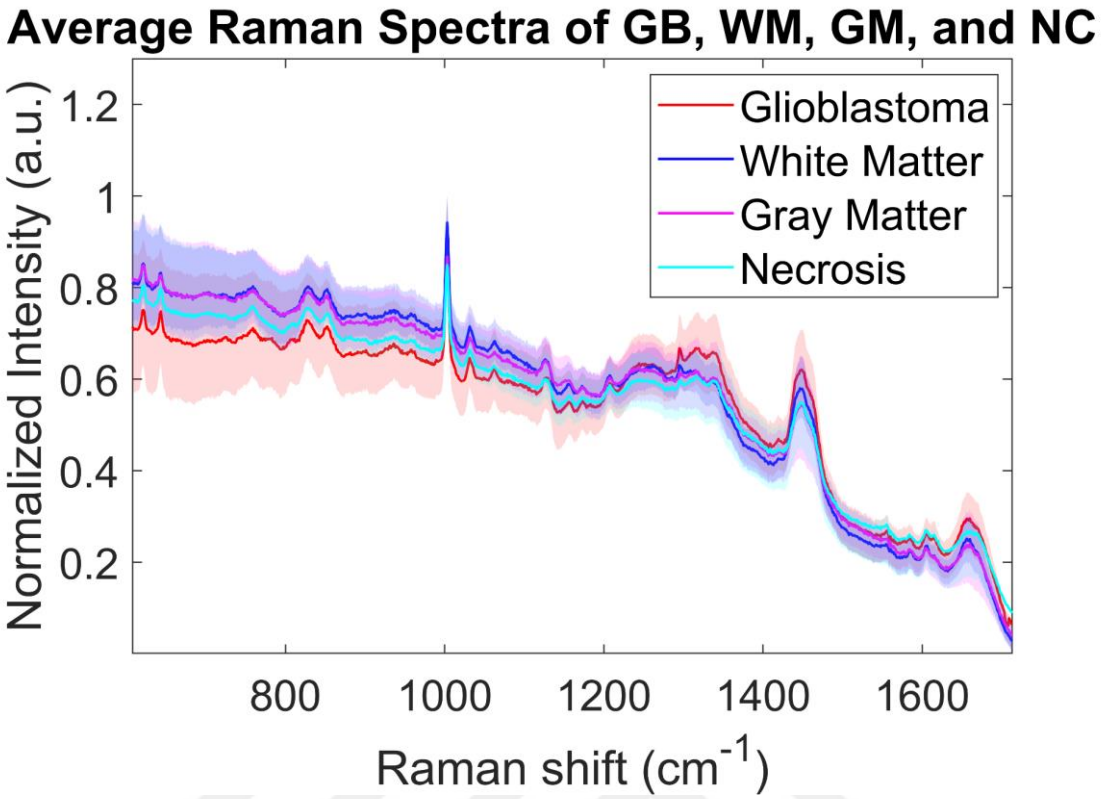


Figure 4.5. Average Raman spectra of glioblastoma (GB), white matter (WM), gray matter (GM), and necrosis (NC) tissue sections. Map spectra for each patient were acquired from 100 different locations on each GB, WM, GM, and NC regions. All spectra were recorded applying full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. (n_{tumor}=38, $n_{whitematter}$=18 $n_{graymatter}$=16, $n_{necrosis}$=19) Sections were deparaffinized according to the Clearene^{TM}-based deparaffinization protocol (see: Chapter 2).

In addition to allowing the distinction of white and gray matter, lipid concentration also has an important role in distinguishing between healthy and neoplastic tissue. As the lipid concentrations are important factors that can assist in the distinguishing of tumors, analysis of lipid content might be used to identify normal tissues from the tumor tissues. The decreasing amount of the total lipid content has been reported while some type of lipids such as phosphatidylcholine increases in the most common primary brain tumors. In gliomas, cholesterol ester concentration has been reported to increase up to a hundred times that in healthy brain tissue [105]. Average brain tissue spectra of GB, WM, GM, and NC were presented in Figure 4.5.

Table 4.1. Lipid content distribution (%) of gray matter and white matter in normal brain tissue. The table is adapted from the following reference [105].

Lipid	**Gray Matter (%)**	**White Matter (%)**
Cholesterol	19.6	26.9
Phosphatidylethanolamine	30.7	19.6
Phosphatidylcholine	25.1	11.8
Galactocerebroside	7.2	18.4
Phosphatidylserine	7.2	9..0
Phosphatidylinositol	3.9	4.0
Sphingomyelin	3.2	4.4
Sulfatide	0.7	5.1
Phosphatidic acid	0.5	0.4
Unidentified	1.9	0.4
SUM	**100**	**100**

4.1.1 Comparison of GB and WM Spectra

Spectra for both GB and WM show similar overall features, which are characterized by nucleic acid, protein, and lipid contents. Those Raman peak intensities change based on their molecule concentrations. Strong Raman shifts were identified at 621, 643, 746-749, 827-829, 852, 937, 957-963, 1003, 1032, 1060-1062, 1126-1131, 1155-1158, 1173, 1205-1209, 1295-1296, 1313-1318, 1336-1339, 1340, 1342, 1409, 1419, 1448-1450, 1458, 1555, 1583-85, 1605-1606, 1615, 1656, 1660, and 1670 cm^{-1} in GB and WM spectra in addition to relatively weak peaks listed with their assignments in Table 4.2 (Figure 4.56).

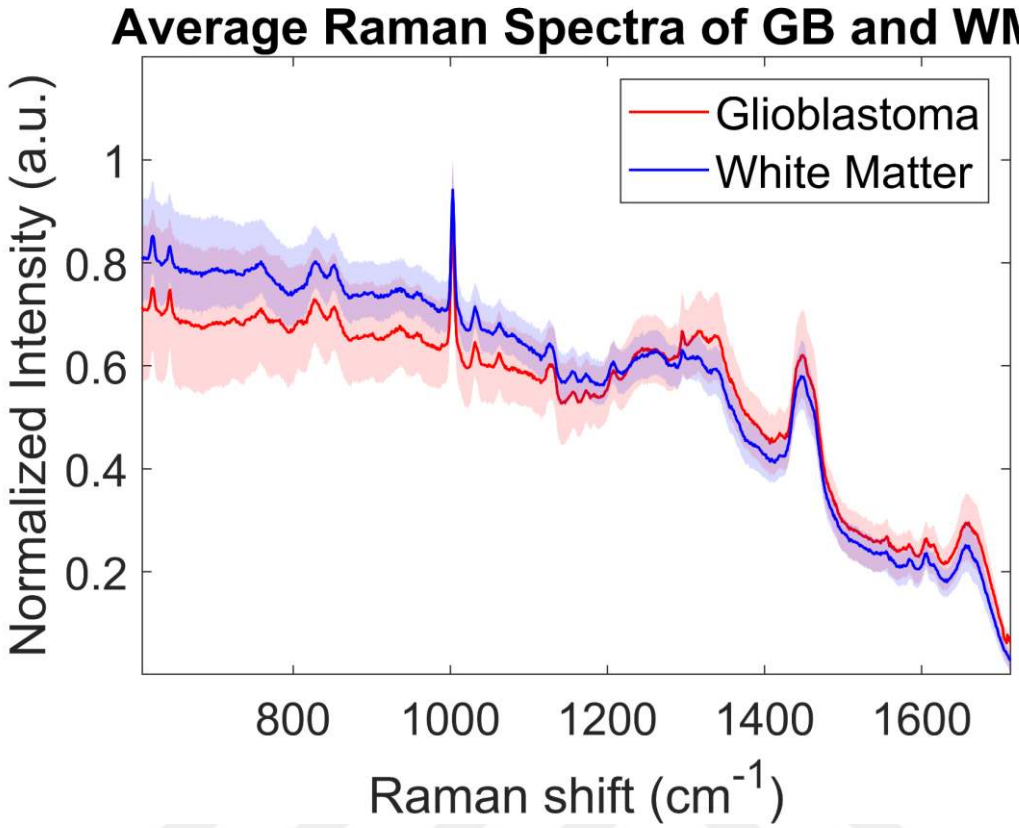


Figure 4.6. Average spectra of glioblastoma (GB) and white matter (WM) tissue sections. Map spectra for each patient were acquired from 100 different locations on each GB, WM, GM, and NC regions. All spectra were recorded applying full laser power (100%), 10 seconds laser acquisition, and one Raman scan on CaF_2 Raman substrate. (n_{tumor}=38, $n_{whitematter}$=18) Sections were deparaffinized according to the Clearene™-based deparaffinization protocol (see: Chapter 2).

Nucleic acid content is expected to be higher than normal tissue and lower lipid content due to the increased nuclear density in tumor tissue. In WM tissue, it is reported that the lipid density is higher than the tumor tissue and the nucleic acid content is less than the tumor tissue [59]. The mean spectrum of GB (n=38) showed lower lipid and cholesterol intensity at 957, 1081-1083, 1126, and 1168 cm^{-1} bands and more nucleic acid and protein intensity at 1313-1318, 1336-1342, 1360, 1419, 1443, 1445, 1446, 1450, 1452, 1458, 1660, 1664, and 1670 cm^{-1} bands than WM (n=18) spectrum (Figure 4.6).

A Raman peak was observed at 725 cm^{-1} in the GB spectrum, which has been attributed to nucleic acid content in the literature [111, 112] (Table 4.2). Broadband at 782 cm^{-1} has been attributed to the vibration of the 5′-C-O-P-O-C3′ molecule and related

with the cytosine at 780 cm^{-1} and thymine at 790 cm^{-1}, depending on the concentration of these DNA nucleic acids [113]. A broadband peak between 1440-1460 cm^{-1} was also observed and the peaks in between have mostly been attributed to protein and lipid content in the literature (Table 4.2). Although one may confuse the 1440 cm^{-1} peak as a paraffin residue, it has been attributed to lipid content in addition to paraffin [105, 112, 114, 115]. The presence of this peak has also been reported in the frozen brain tissue spectrum by Kalkanis et al [59].

Table 4.2. Raman shift assignments for the components of glioblastoma (GB), white matter (WM), gray matter (GM), and necrosis (NC) tissue sections.

Raman shift (cm^{-1})	**Raman shift assignments**
621	C–C twisting mode of phenylalanine (protein assignment) [78, 112, 116, 117]
630	Glycerol [105, 112]
643	C-C twisting mode of tyrosine [78, 112] Tyrosine [118]
725	Adenine (ring breathing modes of the DNA/RNA bases) [111, 112]
739	Ring breathing mode of 5-fluoro-uracil [119] Characteristic peak of Ibuprofen corresponding to C=O [120]
741	7-*cis* configuration of β-Carotene [121]
746-48	Thymine (ring breathing mode of DNA/RNA bases) [111, 112]
749	Tryptophan [118]
752/3/4/5	Symmetric breathing of tryptophan (protein assignment) [47, 112, 118]
755	Symmetric breathing of tryptophan (protein assignment) [78, 112, 116]
757	Symmetric ring breathing in tryptophan (protein assignment) [122]
759	Tryptophan [78, 116, 118, 123]
760	Ring (δ) breathing of tryptophan (protein assignment) [117]
780	Cytosine [113]
782	5'-C-O-P-O-C-3' [113]

790	Thymine [113]
809	O–P–O stretching of the RNA backbone, C–C stretching of proline and hydroxyproline (protein and collagen assignment) [124]
810	Phosphodiester (Z marker) [112, 125]
818	ν_s(C–O–C) vibration [126]
820	Protein [127] Structural protein modes of lesions [112, 128] Proteins (including collagen I) [112]
827	Proteins (Proline, hydroxyproline, and out of plane ring breathing in tyrosine), DNA/RNA (an asymmetric stretch of ν_2 PO_2^-) [112, 118, 124]
828	Out-of-plane ring breathing, tyrosine/O-P-O stretch DNA [78, 112, 116, 117]
829	O–P–O stretching (nucleic acid, tyrosine, and proline) [129] Benzene ring breathing [130]
852	Proline, hydroxyproline [118], Tyrosine ring breathing (protein assignment) [111, 118]
853	Proteins (C–C stretch in proline, collagen assignment), ring breathing of tyrosine (protein assignment), C–O–C stretching of carbohydrates(glycogen and polysaccharide assignment) [122]
937	Proline (collagen type I) [112, 118] Amino acid side chain vibrations of proline and hydroxyproline, as well as a (C-C) vibration of the collagen backbone [112, 118] C-C backbone (collagen assignment) [112, 131] C-C stretching, α-helix (protein assignment) [112, 117] C-O-C glycosides (carbohydrate assignment) [112, 117] Glycogen [112, 122]
957	Hydroxyapatite, carotenoid, cholesterol [78, 112]
963	Unassigned in protein assignments [47, 112] δ(C=O) [132]
1002	Phenylalanine [118] Phenylalanine (collagen assignment) [131]

1003	Proteins (symmetric ring breathing in phenylalanine) [122, 124]
1031	C-O stretching mode [133] Proteins (collagen, C–N stretching in proteins), lipids (phospholipids), carbohydrates (polysaccharides) [122, 124] C–H in-plane bending mode of phenylalanine [78, 116, 122, 124]
1032	Phenylalanine [118]
1060	PO_2^- stretching (DNA/RNA), Chain C-C stretching (carbohydrate assignment) [112] C-O, C-C stretching (carbohydrate assignment) [117] asymmetric C–C stretch (lipid assignment) [134]
1062	Glycosaminoglycan (α-glycan) [135, 136]
1081	Lipid [59] $\nu_1CO_3^{2-}$, $\nu_3PO_4^{3-}$, ν(C-C) skeletal of acyl backbone in lipid (gauche conformation), C–N stretching mode of proteins (and lipid mode to a lesser amount) [112, 118]
1082	Carbohydrate residues of collagen, carbohydrates peak for solutions, nucleic acids [112] Stretching vibration of the bridge C–O–C group in α-lactose [137]
1083	C–N stretching mode of proteins (and lipid mode of a lesser amount) [78, 112, 116]
1125	Symmetric C-C stretch [134]
1126	ν (C-C) skeletal of acyl backbone in lipid (trans conformation) [112, 118] C-N stretching vibration (protein) [111, 112]
1130	Acyl chains [105, 112]
1131	Palmitic acid, fatty acid [105, 112]
1152	ν(C-N), proteins, ν(C-C) carotenoid [47, 112]
1155	C–C and C–N stretching of proteins and carotenoids [78, 112, 116]
1156	C-C and C-N stretching of proteins [111, 112]
1156/7	Carotenoids (absent in normal tissue) [112, 138]
1157	In-plane vibrations of the conjugated =C-C= [112, 139] β-carotene accumulation (C=C stretch mode) [112, 140]

1158	C–C and C–N stretching of proteins [112, 117] ν(C-C) mode of β-carotene [141]
1164	Guanine and Cytosine ring vibrations [113]
1165	ν(=C-C=) of carotenoids [142]
1168	ν(C=C) δ(COH) (lipid assignment), ν(C-C), carotenoid [112, 143]
1173	Guanine, cytosine [112, 125]
1174	Tyrosine [112, 118], phenylalanine, C-H bend (protein) [112]
1180-84	Adenine, guanine, and cytosine [78, 112]
1198	C–N vibrational stretching [144]
1205	Ring-C stretch of Tyrosine [145]
1208	ν(C-C_6H_5), phenylalanine (protein assignment) [47, 112, 143], tryptophan [47, 112] Adenine, Thymine (ring breathing modes of the DNA and RNA bases) [111, 112] Amide III (protein) [111, 112]
1209	Tryptophan and phenylalanine ν(C-C_6H_5) mode [78, 116]
1235-39	Amide III vibration [146]
1285	Amide III [147, 148]
1295	CH_2 twisting mode [149]
1296	CH_2 deformation [75, 112, 146]
1301	CH_2 twisting band (lipid assignment), C-H vibrations [112, 117] Triglycerides (fatty acids) [112, 140]
1313	CH_3CH_2 twisting mode of collagen and lipid [78, 112, 116]
1318	Guanine (ring breathing modes of the DNA and RNA)-C-H deformation (protein) [111, 112] Amide III (α-helix) [112, 114]
1336-39	δ(CH_2), δ(CH_3), twisting mode of collagen [143] CH_3CH_2 wagging and twisting mode of collagen [78, 116], nucleic acid, tryptophan [118] Polynucleotide chain of DNA–purine bases [78, 116] Adenine, Guanine, C-H deformation of proteins [111]
1340	CH_3CH_2 wagging mode of collagen [78, 116]

	Nucleic acid mode [138] Tryptophan [150]
1342	Guanine (DNA and RNA bases), CH deformation (protein and carbohydrate assignment) [112, 117]
1360	Tryptophan, Lysine [150]
1367	$\nu_s(CH_3)$ (phospholipid assignment) [112, 114]
1379	$\delta(CH_3)$ symmetric (lipid assignment) [143]
1409	ν_s COO^- (IgG?) [112, 114]
1419	2'CH_2 scissoring modes (DNA assignment) [151]
1433	CH_2 bending (lipid assignment) [134]
1440	Vibration of CH_2 and CH_3 deformation [112, 127] CH deformation [105] Cholesterol, fatty acid band [105, 112] $\delta(CH_2)$ (lipid assignment) [112, 114] CH_2 bending (lipid assignment) [112, 114, 115] Cholesterol [59]
1443	CH_2 deformation of lipids and proteins [78, 116]
1445	$\delta(CH_2)$, $\delta(CH_3)$ (protein assignment), $\delta(CH_2)$ scissoring (lipid assignment) [112, 143]
1446	CH_2 bending mode of proteins [78, 116]
1448/49	CH_2 and CH_3 deformation [118]
1450	CH_2 bending mode (malignant tissues) [152] CH_2 bending (proteins) [114]
1452	Structural protein modes [127, 128]
1454	CH_2 bending (lipid assignment) [134]
1458	Nucleic acid modes [112, 138] Fatty acid substituents [105]
1463	Fermi interaction $\delta(CH_2)$, & $\Upsilon(CH_2)$ [112, 114] Fatty acid substituents [105]
1555	Amide II (C–N–H valence, deformation vibration of N– H) [153]
1583	C=C bending mode of phenylalanine [112]
1585	ν(C=C) olefinic stretch (protein assignment) [112, 146]

1605	Cytosine (NH_2) [112, 125] $\nu_{as}(COO^-)$ [132] Phenylalanine, tyrosine, C=C (protein assignment) [111, 112]
1606	ν(C=C) of the vinyl group of red blood cells (hemoglobin assignment) [154]
1615	Tryptophan, tyrosine, C=C (protein assignment) [111, 112]
1656	$\nu(C=C)_{cis}$ (phospholipid assignment) [112, 114]
1660	Amide I [122]
1664	Amide I band of proteins, Thymine, Guanine, Cytosine (nucleic acid assignment) [111]
1670	Amide I band of proteins, Thymine, Guanine, Cytosine (nucleic acid assignment) [111]

Table 4.2 shows peaks that can be attributed to multiple different molecules. For instance, peaks originating from $(CH_2)_n$ may belong to paraffin, lipid, nucleic acids, and proteins. Peaks between 1440-60 cm^{-1} have been attributed to protein, lipid, nucleic acid, and paraffin (CH_2 deformation/bending/rocking) content. Similarly, peaks between 1130-1131 cm^{-1} have been attributed to both lipids and paraffin. Hence, instead of subtracting paraffin peaks from the FFPE Raman spectrum, an optimal deparaffinization method could help enhance discrimination of tissue types from Raman spectra. The deparaffinization protocol we proposed eliminates major paraffin peaks within 888, 1062, 1132, 1169, 1295, 1417, 1440, 1460 cm^{-1} (Figure 64), while retaining peaks that can be attributed to protein (i.e. 1340, 1360, 1615 cm^{-1}), nucleic acids (725, 1180-84 cm^{-1}), and lipid (1367, 1379, 1440 cm^{-1}) content (Table 4.2). We preserved both morphology (Figure 3.49) and the chemical content (Figures 3.50-4.6) of the tissue during the deparaffinization process.

4.1.2 Comparison of White Matter and Gray Matter Tissue

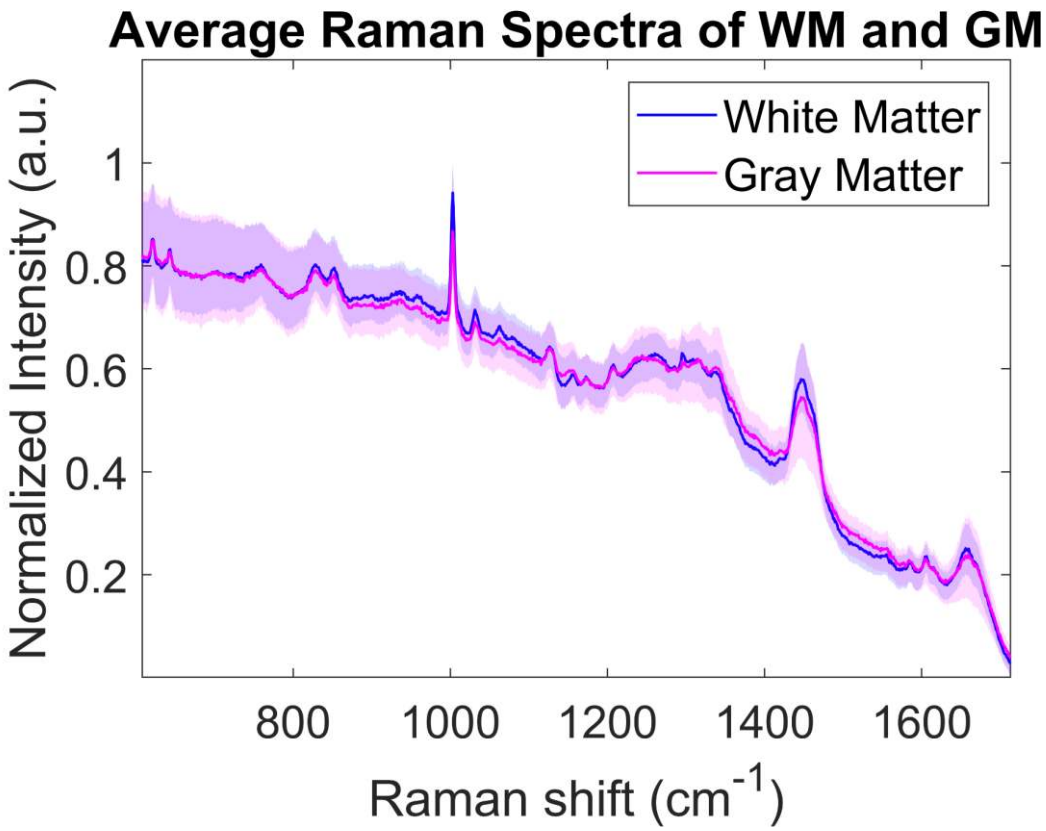


Figure 4.7. Average spectra of white matter (WM) and gray matter (GM) tissue sections. Map spectra for each patient were acquired from 100 different locations on each WM and GM regions per patient. All spectra were recorded applying full laser power (100%), 10 seconds of laser acquisition, and one Raman scan on CaF_2 Raman substrate. ($n_{whitematter}$=18, $n_{graymatter}$=16) Sections were deparaffinized according to the Clearene™-based deparaffinization protocol (see: Chapter 2).

Upon their similar molecular content, GM and WM showed similar Raman spectra with increasing and decreasing intensities of some specific peaks, related with their concentration in the tissue type. Nevertheless, due to the enriched lipid and protein content of the WM originated from the structure of myelin sheaths[106, 108-110], there are some differences in the intensities of the lipid peaks. The peaks at 1060, 1125, 1433, and 1454 cm^{-1} that have been attributed to the lipid content [134] in the literature were observed higher in white matter. Accordingly, there are several peaks observed with higher intensity in WM spectra in the range. Due to the convenience of the spectral range between 1000 to 1200 cm^{-1} in detecting the variations of acyl chains in lipids, the peaks

1085 and 1110 cm^{-1} peaks that have been attributed to hydrocarbon C–C skeletal stretching of lipids [134] were observed and showed higher intensity in WM spectra (Figure 4.7).

Myelin sheaths, 50% of WM also include different myelin proteins that are composed of more than 200 amino acids [155]. The Raman peaks at 820, 827-828, 852-853, 937, 963, 1003, 1030, 1083, 1445-1446 cm^{-1} that have been attributed to different range of proteins or amino acids showed higher intensity at WM spectra. Moreover, due to the neuronal cell bodies and the cell nucleus, GM spectra showed higher intensity in the Raman peaks at 1339-1342, 1419 cm^{-1} that have been attributed to the nucleic acid content (Figure 4.7).

4.1.3 Comparison of Glioblastoma and Necrosis Tissue

Lipid accumulation in the brain could be a necrotic tissue marker [105]. While phospholipids and cholesterols are gradually reduced towards the neoplastic regions, necrotic regions were reported to be rich in plasma protein content [59] (620, 643, 755-760, 827-828, 852-853, 937, 963, 1003, and 1031-1032 cm^{-1}). Despite that, GB is demonstrated higher lipid (1301, 1440, 1443-1445, and 1656 cm^{-1}) and nucleic acid (1208, 1318, 1339, 1419, and 1458 cm^{-1}) content (Figure 4.8) consistent with the literature [59].

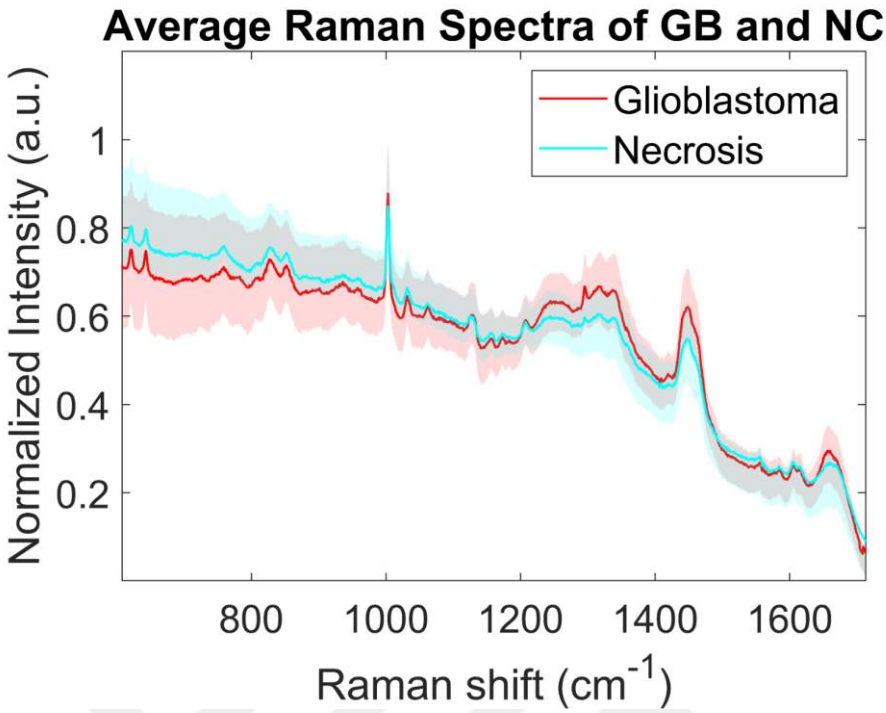


Figure 4.8. Average spectra of glioblastoma (GB) and necrosis (NC) tissue sections. Map spectra for each patient were acquired from 100 different locations on each GB and NC regions per patient. All spectra were recorded applying full laser power (100%), 10 seconds of laser acquisition, and one Raman scan on CaF_2 Raman substrate. (n_{tumor}=38, necrosis=19) Sections were deparaffinized according to the Clearene^{TM}-based deparaffinization protocol (see: Chapter 2).

4.2 Conclusions and Insights

We determine the specific Raman spectral profiles for GB, WM, GM, and the NC. Then, we completed a wide range of literature reviews about the weak, strong, broad, and shoulder peaks that we detect on our measurements and construct a database for analysis of the tissue constituents. Next, we analyzed the specific peaks in the Raman spectral profiles of each tissue type and compared them to each other (Aim-2A).

We constructed a spectral database to develop a digital pathology method as a proof-of-concept that can be used as a reference base for analyzing samples to determine their Raman peak combinations and nature (neoplastic versus non-neoplastic) to distinguish from each other. After collecting the spectral profiles for different tissue types (GB, WM, GM, and NC) (Figure 4.5), we classified the spectra obtained by these four tissue groups using several machine-learning algorithms (Aim-2B) (See: Chapter 5).

Chapter 5:
CLASSIFICATION OF GLIOBLASTOMA USING MACHINE LEARNING MODELS

Refining the optimal substrate type, section thickness, and deparaffinization protocols were necessary to get repeatable spectral measurements by minimizing the differences during the tissue preparation process. Choosing Raman substrate type with a minimum background also helps eliminate the background removal step and could further be useful for our classification of the Raman spectra acquired from GB and WM. In this chapter, we present machine learning classification results of tissue types (Aim-2B). The preprocessing and machine learning model codes are presented in the following reference [156]. The models in this chapter have been developed with support from Numan Batur.

5.1 Pre-processing Methods

GB and WM spectra were classified either with or without pre-processing operations were applied before the classification, using SVM, kNN, and RF models. First, raw data were classified using each model, and given pre-processing method combinations (Table 5.1) were applied to data before classification. The classification results of the data with or without each combination of pre-processing steps were presented in this chapter (Table 5.2-5.4). Mean tissue spectra to which all pre-processing steps (sample wise min-max normalization, Savitzky Golay, PCA, and scaling, within the given order) were applied to, given in Figure 5.1.

Table 5.1. Determination of the pre-processing steps followed before the classification. (Each pre-processing operation was applied with the given order in the line)

None → Spectra were classified without any pre-processing operation was applied.
Normalization → Spectra were classified after the min-max normalization was applied.
Normalization + PCA → Spectra were classified after the sample wise min-max normalization and PCA operations were applied, respectively.
Normalization + Savitzky Golay + PCA → Spectra were classified after the sample wise min-max normalization, Savitzky Golay, and PCA operations were applied.
Normalization + Savitzky Golay + Scaling + PCA → Spectra were classified after the sample wise min-max normalization, Savitzky Golay, PCA, and scaling operations were applied.

5.1.1 Normalization

The normalization step was applied to the tissue spectra to ensure the intensity heights of the spectral data are not detected as features by the classification models in the next step. After normalization of the spectra to the range between 0 and 1, all classification accuracies increased in SVM and RF classifications (Table 5.2, Table 5.3). Unlike SVM and RF classifications, only WM test accuracy increased, in the kNN classification and train and GB test accuracy decreased (Table 4).

5.1.2 PCA

PCA was applied to the spectral data after the normalization step to reduce the dimension of the data. When we give all 1011 Raman shifts in the data as a feature, since each shift cannot distinguish the spectrum type alone, a dimension reduction method such as PCA allows the shifts of maximum variance to be taken as principal components (PCs) and converted into a new feature set. Since the PCs are the features with the maximum variation, our classification accuracies increased in the new feature set (Table 5.2-5.4).

5.1.3 Savitzky Golay Filter

Savitzky Golay filter to denoise the data by polynomial fitting to remove the low-frequency components in our dataset. We determined an optimum window size and polynomial order pair by applying iterative combinations to our data. We applied a 4^{th} order polynomial fitting and 61 data points as the window size for better classification of our data. After the 4^{th}-order polynomial fitting, the first derivative of the resulting signal was taken. When normalization, Savitzky Golay, and PCA were applied in this order, the train and GB accuracies increased in SVM, but the WM test accuracy decreased (Table 5.2). In kNN classification, all accuracy values increased (Table 5.3). Despite that, train and WM test accuracies decreased but the GB accuracy increased in RF classification (Table 5.4).

5.1.4 Scaling

The variations in the intensity among different spectra may cause the standard deviations to vary extensively and bias the models due to different backgrounds, peak intensities, or calibration artifacts. To normalize this deviation to become unit '1' in all features and the mean to be '0', we created a data set in a way that allows us to evaluate all data in an equal plane. By standard scaling, the variance differences between the features were eliminated and the features were not weighted according to each other. When applying normalization; Savitzky Golay, scaling, and PCA was applied in this given order. While train accuracy and WM test accuracy is increasing, GB test accuracy decreased in SVM classification, after all the pre-processing steps were applied in the SVM model (Table 5.2). All accuracy values decreased in the kNN model when all the pre-processing methods were applied (Table 5.3). Train and WM accuracies decreased, and GB test accuracy increased when all pre-processing steps were applied in the RF model (Table 5.4).

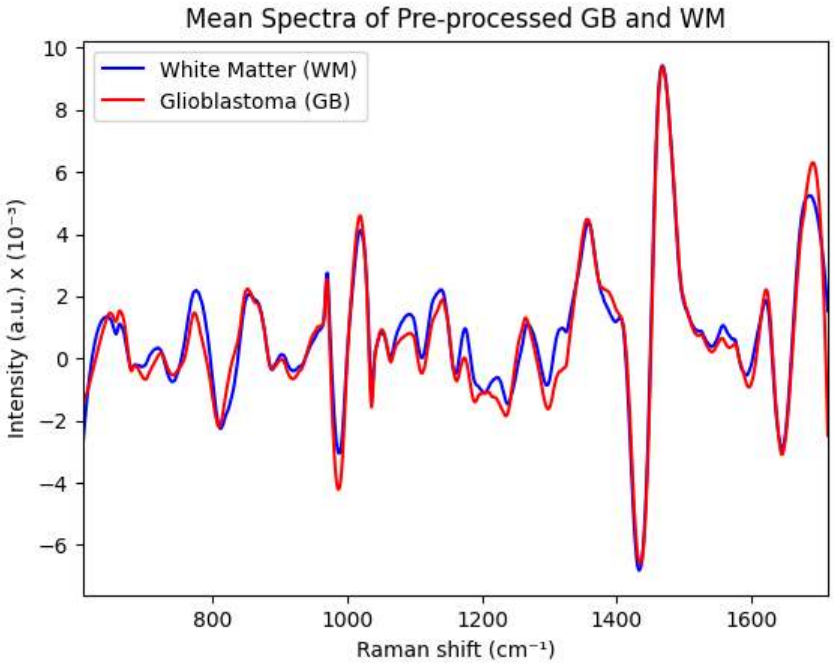


Figure 5.1. Mean Spectra of Pre-processed Glioblastoma (GB) and White Matter (WM)

5.2 Machine Learning Classifications

5.2.1 Support Vector Machine (SVM)

In the SVM model, the 'SupportVectorClassifier' class of 'scikit learn' library was used for the classification of tissue spectra acquired from different tissue types. The 'GridSearchCV' class of the 'scikit learn' library was used for hyperparameter tuning. The optimal number of the 'C' value determined as 1 for all three classification groups (GB vs WM, GB vs WM+GM, and 4-class classification). 'kernel' parameter of the model was set as "rbf" and the other parameters were used in their default settings. The train-test ratio was determined as 80% and 20% to all classification groups, respectively.

For the classification of the groups of GB and WM tissues, 100 spectra from relevant tissue regions per patient were employed from 20 and 18 patients. SVM model resulted in 70-90±1% training, 68-87±1% GB, and 73-90±1% WM test accuracies (Table 5.2). The area under the receiver operating characteristics (ROC) curve was calculated as 0.96 when all pre-processing methods were applied (Figure 5.2). The confusion matrix of GB and WM classification with the SVM model resulted in 314 true negative, 43 false negative, 357 true positive, and 46 false positive values (Figure 5.3).

Table 5.2. Classification Results of Glioblastoma (GB) vs White Matter (WM) Spectra Classification using Support Vector Machine (SVM) Model. 2000 tumor and 1800 white matter tissue spectra were classified within the 80% train and 20% test ratios (n_{GB}=20, n_{WM}=18).

Pre-Processing	**Train Accuracy**	**GB Test Accuracy**	**WM Test Accuracy**
None	69.1±1 %	68.0±1%	73.8±0%
Normalization	80.3±1%	73.0±1%	88.0±1%
Normalization + PCA	85.3±1%	81.5±1 %	91.1±1 %
Normalization + Savitzky Golay + PCA	88.3±1%	88.0±1%	89.7±1%
Normalization + Savitzky Golay + Scaling + PCA	90.0±1%	87.2±1%	90.7±1%

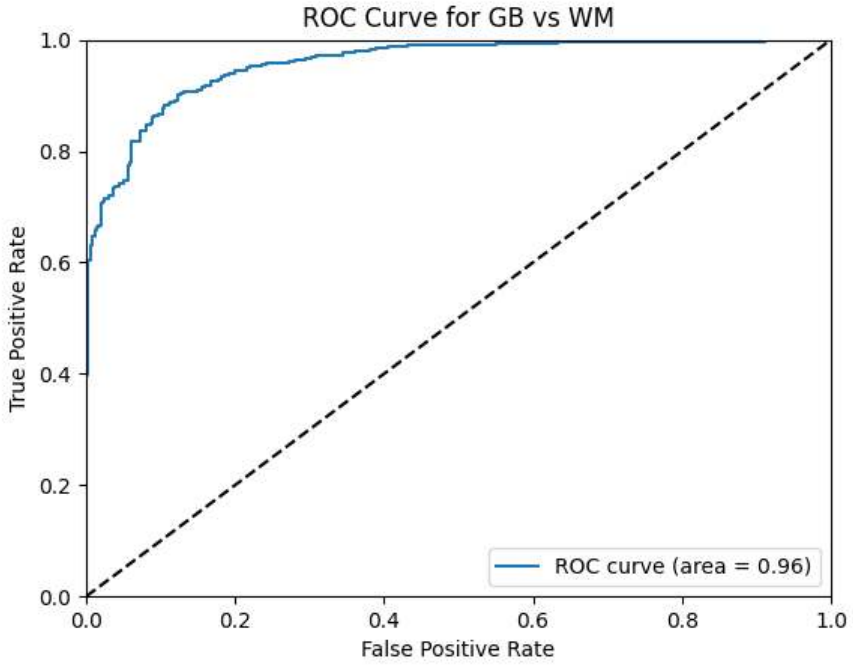


Figure 5.2. Receiver Operator Characteristic (ROC) Curve for Glioblastoma (GB) and White Matter (WM) Classification using Support Vector Machine (SVM) Model. All pre-processing steps (Normalization + Savitzky Golay + PCA + Scaling) were applied. ROC curve area = 0.96.

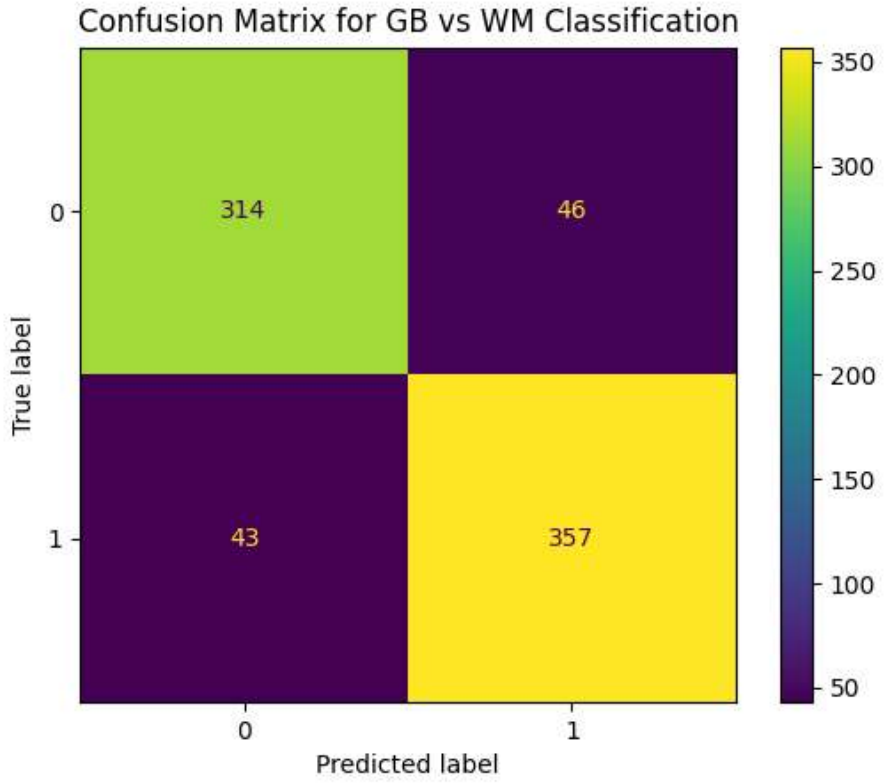


Figure 5.3. Confusion Matrix for Glioblastoma (GB) vs White Matter (WM) Classification. True label (y-axes) represents the actual label (WM:0, GB:1) and predicted label (x-axes) represents the prediction of the Support Vector Machine Model (SVM). True Negative (0-0) = 314, False Negative (1-0) = 43, True Positive (1-1) = 357, False Positive (0-1) = 46.

5.2.2 k-Nearest Neighbors (kNN)

In the kNN model, the 'KNeighborsClassifier' class of 'scikit learn' library was used for the classification of tissue spectra acquired from different tissue types and 'GridSearchCV' class of 'scikit learn' library was used for hyperparameter tuning of the neighbor numbers, and 'k' value. The optimal 'k' number was determined as 7 for GB vs WM classification. The other parameters were set as default. The train-test ratio was determined to be 80% and 20%, respectively.

For the classification of GB and WM, 100 spectra from relevant tissue regions per patient were employed from 20 and 18 patients. The kNN model resulted in 73-82±1% training, 60-77±1% GB and 75-93±1% WM test accuracies (Table 5.3). The area under the ROC curve was calculated as 0.91 when all pre-processing methods were applied

(Figure 5.4). The confusion matrix of GB and WM classification with the kNN model resulted in 333 true negative, 105 false negative, 295 true positive, and 27 false positive values (Figure 5.5).

Table 5.3. Classification Results of Glioblastoma (GB) vs White Matter (WM) Spectra Classification using k-Nearest Neighbors (k-NN) Model. 2000 tumor and 1800 white matter tissue spectra were classified within the 80% train and 20% test ratios. 2000 tumor and 1800 white matter tissue spectra were classified within the 80% train and 20% test ratios (n_{GB}=20, n_{WM}=18).

Pre-Processing	**Train Accuracy**	**GB Test Accuracy**	**WM Test Accuracy**
None	74.1±1%	77.5±1%	75.5±1%
Normalization	73.7±1%	60.7±1%	87.7±1%
Normalization + PCA	79.2±1%	71.2±1%	90.8±1%
Normalization + Savitzky Golay + PCA	82.6±1%	75.2±1%	93.0±1%
Normalization + Savitzky Golay + Scaling + PCA	81.2±1%	75.2±1%	88.0±1%

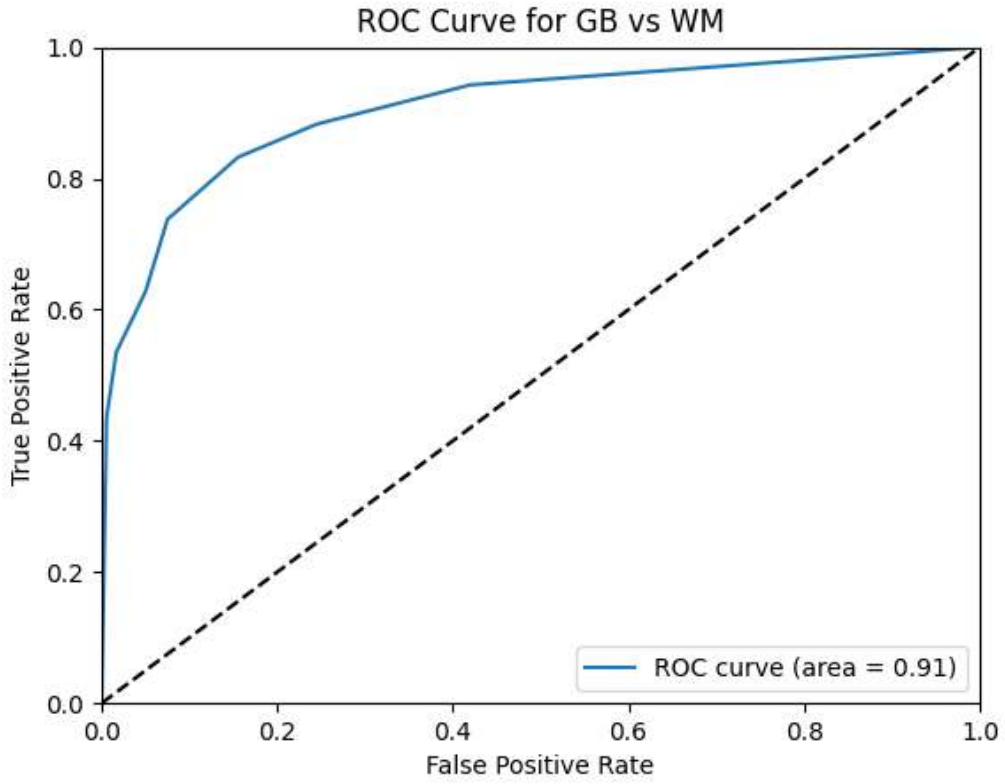


Figure 5.4. Receiver Operator Characteristic (ROC) Curve for Glioblastoma (GB) and White Matter (WM) Classification using k-Nearest Neighbors (kNN) Model. All pre-processing steps (Normalization + Savitzky Golay + PCA + Scaling) were applied. ROC curve area = 0.91.

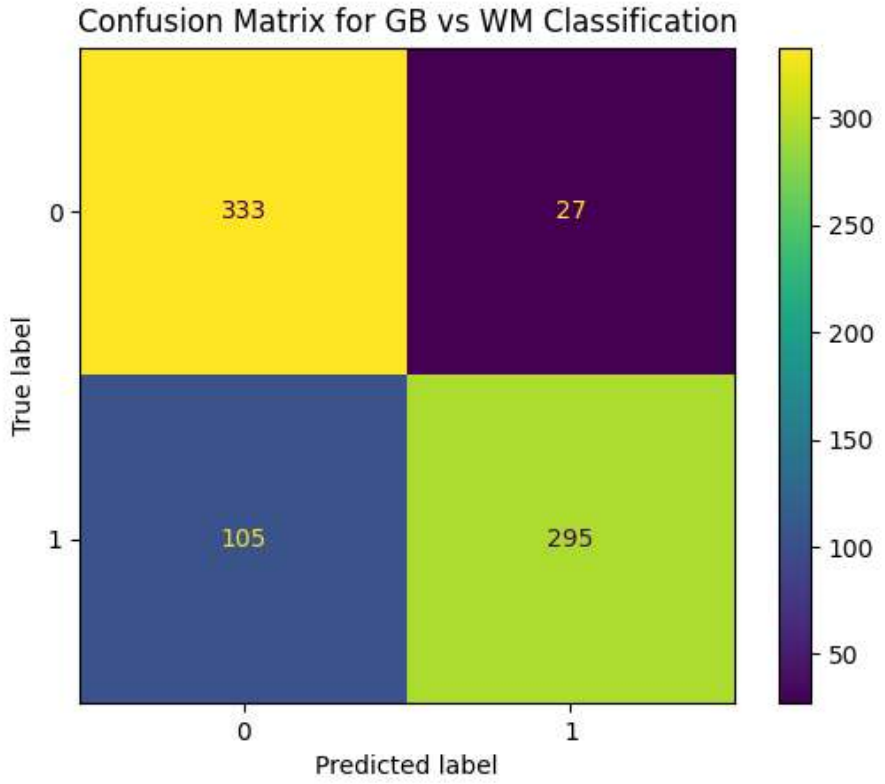


Figure 5.5. Confusion Matrix for Glioblastoma (GB) vs White Matter (WM) Classification. True label (y-axes) represents the actual label (WM+GM:0, GB:1) and predicted label (x-axes) represents the prediction of the k-Nearest Neighbors (kNN). True Negative (0-0) = 333, False Negative (1-0) = 105, True Positive (1-1) = 295, False Positive (0-1) = 27.

5.2.3 Random Forest (RF)

In the RF model, the 'RandomForestClassifier' class of 'scikit learn' library was used for tissue spectra classification and the 'GridSearchCV' class of 'scikit learn' library was used for hyperparameter tuning. The optimal number of the trees (individual classifiers) that were used in the model determined as 50 for GB vs WM classification. The other parameters were set as default. The train-test ratio was determined as 80% and 20%, respectively.

For the classification of the groups of neoplastic (GB) and normal (WM) tissues, 100 spectra from relevant tissue regions per patient were employed from 20 and 18 patients. RF model resulted in 75-88±1% training, 78-91±1% GB and 73-87±1% WM test accuracies (Table 5.4). The area under the ROC curve was calculated as 0.94 when all

pre-processing methods were applied (Figure 5.6). The confusion matrix of GB and WM classification with RF model resulted in 292 true negative, 46 false negative, 354 true positive, and 68 false positive values (Figure 5.7).

Table 5.4. Classification using Model. Classification Results of Glioblastoma (GB) vs White Matter (WM) Spectra Classification using Random Forest (RF) Model. 2000 tumor and 1800 white matter tissue spectra were classified within the 80% train and 20% test ratios (n_{GB}=20, n_{WM}=18).

Pre-Processing	**Train Accuracy**	**GB Test Accuracy**	**WM Test Accuracy**
None	75.4±1%	78.5±1%	73.3±1%
Normalization	81.6±1%	81.2±1%	81.9±1 %
Normalization + PCA	88.5±1%	86.7±1%	87.7±1%
Normalization + Savitzky Golay + PCA	87.4±1%	90.2±1%	82.2±1%
Normalization + Savitzky Golay + Scaling + PCA	86.8±1%	91.0±1%	80.0±1%

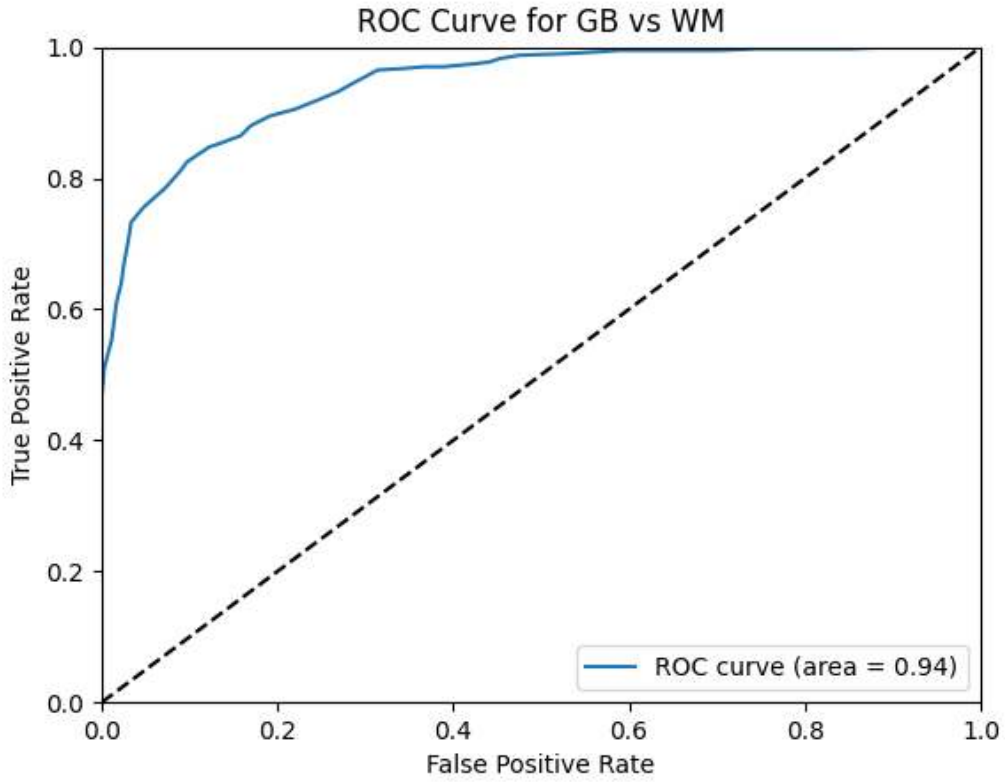


Figure 5.6. Receiver Operator Characteristic (ROC) Curve for Glioblastoma (GB) and White Matter (WM) Classification using Random Forest (RF) Model. All pre-processing steps (Normalization + Savitzky Golay + PCA + Scaling) were applied. ROC curve area = 0.94.

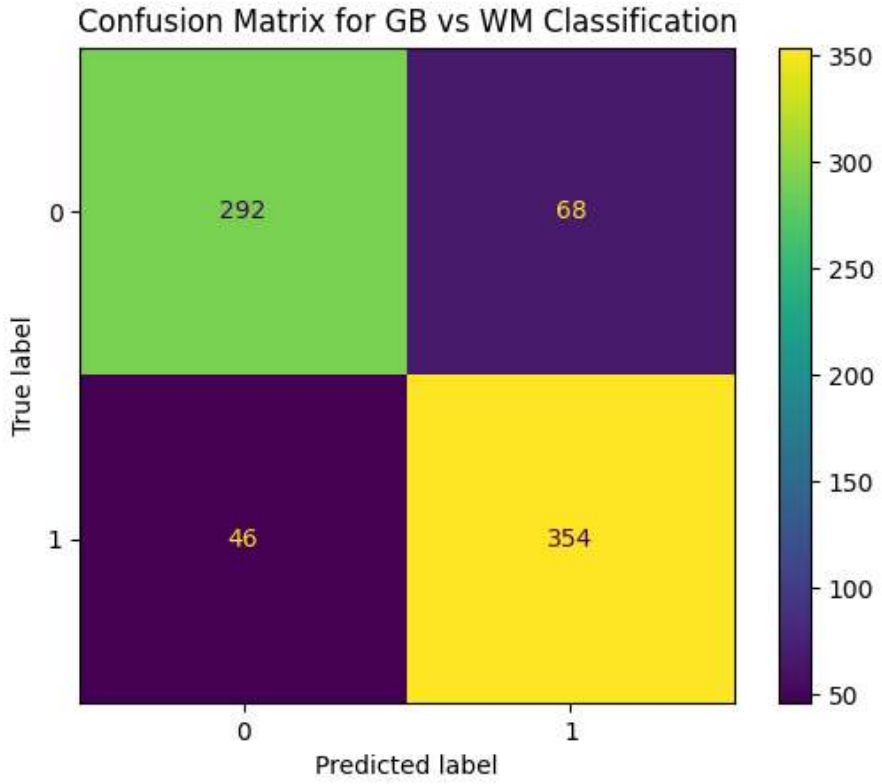


Figure 5.7. Confusion Matrix for Glioblastoma (GB) vs White Matter (WM) Classification. True label (y-axes) represents the actual label (WM:0, GB:1) and predicted label (x-axes) represents the prediction of the Random Forest (RF) Model. True Negative (0-0) = 292, False Negative (1-0) = 46, True Positive (1-1) = 354, False Positive (0-1) = 68.

5.3 Conclusions and Insights

After the optimization of sample preparation and the Raman acquisition protocols, we constructed a spectral library to distinguish the GB spectra from WM using machine learning models. Then we optimized the pre-processing parameters for GB and WM classification and investigate the effects of each pre-processing method on the train and test accuracies. We observed that when the normalization, Savitzky Golay, Scaling, and PCA steps were applied to our dataset in this order, our accuracies increased in the SVM model (Table 5.2). While the first three pre-processing steps yield higher accuracy results in the kNN model, the scaling operation decreased our accuracies (Table 5.3). For the RF model, although normalization and PCA steps increased the accuracies, that Savitzky-

Golay and the scaling operations did not yield higher accuracies for train and WM test accuracies (Table 5.4).

We achieved up to 90.0±1%, 82.6±1%, and 88.5±1% training accuracies with SVM, kNN, and RF models, respectively. We achieved 88.0±1%, 75.2±1%, 91.0±1% GB test and 91.1±1%, 93.0±1%, 87.7±1% WM accuracies using SVM, kNN, and RF models, respectively. This accuracy results may further be improved with the increasing the sample size and refinement of pre-processing and classification models.

We used 2000 GB and 1800 WM spectra with the 80-20% train-test ratio and presented the confusion matrix of each machine learning model (Figure 3, Figure 5, Figure 7).

Chapter 6:
CONCLUSION, ADVANTAGES & LIMITATIONS, AND FUTURE WORK

6.1 Conclusion

We presented a series of protocols presented here for reproducible sample preparation, Raman measurements, and machine learning model-based classification of the measured spectra on FFPE brain tissue sections. Using 10 µm-thick sections, Clearene^{TM} and successive alcohol dipping-based deparaffinization protocols and optimal Raman measurement conditions (785 nm laser with full power, CaF_2 substrates, 1-2 accumulations, and 10 second accumulation times) have been found to yield high SNR spectra. These optimized protocols could allow neuropathologists to do retrospective studies of archival histopathological tissue samples using analytical chemical analysis using RS.

The spectral analysis of each group of tissue was demonstrated by detailed peak analysis using the reference Raman peaks given in the literature. The average spectral profile of GB, WM, GM, and NC was shown in addition to the spectral comparisons and interpretations of GB vs WM, WM vs GM, and GB vs NC. Our peak analysis results discussed in Chapter 4 are consistent with the literature and need to be improved with a larger sample size for more reliable results and a better spectral data library for machine learning classification.

In this study, we presented our results as a proof-of-concept and classified the spectral data that we collected so far from different tissue types. These results were presented based on the tissue spectra acquired from GB and WM tissue sections. 100 GB spectra acquired from each 20 GB and 18 WM tissue sections Collected 2000 GB and 1800 WM spectra were classified within the 80-20% train-test ratio. 1600 GB and 1440 WM spectra were used as a training set and 400 GB and 360 WM spectra were used as a test set.

Three machine learning models (SVM, kNN, and RF) were employed to classify the Raman spectra of GB and WM tissues. GB and WM classification (n_{GB} = 20, n_{WM} = 18,

100 spectra per sample), SVM model resulted in up to 90±1% training, 87±1 % GB, and 90±1% WM test accuracies and 0.96 area under the ROC curve (Table 5.2) (Figure 5.2). kNN model resulted in up to 82±1% training, 77±1% GB, and 93±1% WM test accuracies and 0.91 area under the ROC curve (Table 5.3) (Figure 5.4). RF model resulted in up to 88% training, 91±1% GB and 87±1% WM test accuracies, and 0.94 area under the ROC curve (Table 5.4) (Figure 5.6).

Moreover, we presented the true positive/negative and false positive/negative values with the confusion matrix of each classification model (Figure 5.3, Figure 5.5, and Figure 5.7) As a result of SVM classification, true negative and false negative values were obtained as 314 and 43, and true positive and false positive values were obtained as 357 and 46, respectively (Figure 5.3). As a result of kNN classification, true negative and false negative values were obtained as 333 and 105, and true positive and false positive values were obtained as 295 and 27, respectively (Figure 5.5). As a result of RF classification, true negative and false negative values were obtained as 292 and 46, and true positive and false positive values were obtained as 354 and 68, respectively (Figure 5.7). To conclude, increasing the sample number and refining the pre-processing parameters are needed for better classification accuracies and eventual use of this technique in the clinics to augment neuropathologists in the identification of GB.

6.2 Advantages & Limitations

6.2.1 Advantages

1. The project uses FFPE-based tissues, which allow the use of routine H&E staining methods and pathologist classification of these tissues as the gold standard for comparison with Raman-based tumor identification. As one-to-one overlapping fresh tissue with the gold standard (H&E-stained sections), the use of FFPE tissue has eliminated this challenge. This allowed verification of tissue types analyzed with RS, providing a one-to-one comparison with the gold standard. In addition, the use of FFPE slides allowed us a reproducible analysis of RS and H&E histology in the following section. Since using fresh tissue requires a high number of surgeries in the limited research duration to reach sufficient numbers of patients, excessive pathology archival tissues (FFPE) were selected for this study.

2. The use of FFPE tissues also allowed us the analysis of archival material, providing retrospective studies on a larger sample set in neuropathological tissue banks and archives for spanning a broader population diversity and potentially a broader set of high-grade glioma types for the construction of the spectral database. Moreover, the use of FFPE tissue samples allowed transportation of tissues from one center to another (i.e. UCSF to KUH) with great ease, providing global access to the technology without having to acquire expensive cryogenic equipment (i.e. for storing and transporting fresh tissue sections). As a result, the classification models for RS and microscopy could be tested and developed for prospective studies of high-grade glioma detection in addition to the retrospective studies.

3. The use of FFPE material also allowed us to select the tissue regions of interest such as GB, WM, GM, and NC regions for that study. Besides, we can obtain multiple samples within micrometers from one another that can also be easily compared or overlaid onto prior slides.

4. The development of optimization protocols for RS allowed us to do retrospective studies on the tissue types in neuropathological tissue banks and archives. As a result, the classification models for RS and microscopy might eventually be tested and developed for prospective studies of high-grade glioma detection.

5. The use of Raman spectral analysis allowed us to detect different tissue types and without the use of specialized staining techniques such as immunohistochemistry (IHC) or histochemistry. Raman spectral analysis also helps keep the tissue intact and unstained, allowing the use of the rest of the tissue for other purposes and analysis techniques. It is also possible to analyze the same tissue using other spectral methods such as infrared lasers or Fourier Transform Infrared Spectroscopy (FTIR) to acquire additional characteristics of tissues, providing a finer analysis of the tissues [157].

6.2.2 Limitations

1. The deparaffinization step was required as we used paraffin-embedded tissue sections before proceeding with the spectral measurements and the staining protocol. For the effective and efficient spectral and pathological analysis, FFPE tissue sections must be deparaffinized and rehydrated. Insufficient removal of paraffin may lead to unwanted

spectra under Raman spectroscopic analysis and poor staining process of the tissue sections. The limitations imparted by the ineffective deparaffinization step were abated during the optimization process to enable a maximum SNR. The optimized deparaffinization technique may not be applicable for all desired experiments or different tissue types and maybe a limitation by preventing precise Raman measurements.

2. While we confirmed that both deparaffinization solvents and protocols kept the tumor morphologies intact, we did not fully validate that these deparaffinization steps did not alter the chemical composition of the tumor sections. For an unambiguous confirmation of preservation of the chemical composition of tumor sections, the Raman spectra of the samples before FFPE and after FFPE & deparaffinization must be compared.

2. RS might not be able to provide an extensive analysis at very low molecular concentrations and for nearly identical molecules, some molecules will be too similar in chemical composition but different in their spatial arrangements (i.e. genes and polymorphisms/mutations). However, other molecular and spectral techniques[157] can be used to acquire this information and combine it with the data obtained from RS since the tissue or cells will not be altered through Raman spectroscopic analysis.

3. The study of morphological changes in GBs is challenging due to significant variations in cytological, architectural, and molecular features of these malignant tumors. GBs have numerous molecular profiles and even more variable morphological features. Therefore, it is not possible to define a uniform “GB profile”. However, we aimed to specifically focus on the tissue characteristics (such as necrosis, tumor, healthy tissue, etc) to create a multivariate database that can be used to characterize tissues and distinguish abnormal from normal using machine learning models. This project is a proof-of-concept attempt for the larger goal of developing Raman spectral databases of GB, WM, GM, and NC tissue types.

4. The validation of Raman spectral analysis might be time-consuming and challenging for pathologists since they need to mark all different cell types in the H&E-stained tumor tissue. While this may allow us to limit our goals to only one molecular type of GB, future studies and experiments will further this effort by expanding our observations to a wide range of normal and abnormal samples. Developing informational databases may allow for more efficient analyses of tissues in the future.

5. Since pathologist annotations for tissue section types may be prone to errors, the models are limited by the classification accuracies of pathologists. Multiple board-certified pathologists might need to cross-check each other's annotations to minimize false positives and false negatives.

6.3 Future Work

In this thesis study, we developed an optimal sample preparation protocol (i.e. deparaffinization, thickness optimization) for removing paraffin and formalin residuals from the tissue sections efficiently and effectively to obtain reliable and reproducible Raman signal acquisition while retaining the morphological structure of the tissue samples. Although we estimate that the composition of the tissue sections is not altered significantly, we do not preclude the possibility of unintentionally dissolving weak Raman peaks from low-concentration non-paraffin molecular species during deparaffinization. For a conclusive analysis of whether the composition is fully preserved after deparaffinization, one needs to start with a fresh brain tumor section tissue and measure its initial Raman spectrum. Then, this tissue should be fixed in formalin and embedded in paraffin, and then paraffin should be cleared, and its Raman spectrum should be measured. Finally, these two Raman spectra (before formalin fixation and paraffin embedding and after deparaffinization of FFPE tissue) should be compared. This analysis is reserved for future studies.

The results that we presented as a proof-of-concept in this thesis study will be improved with increasing sample size to obtain more accurate and reliable results by constructing a larger spectral database. We aim to understand whether specific morphological subtypes and molecular patterns can be readily recognized based on spectral analysis without the need for additional wet bench steps, genetic analysis, or histomorphologic evaluation. The collected spectral database of normal, necrotic, and GB tissue profiles will lead us to develop a set of machine learning models as a digital pathology approach that can be used as a decision support system in clinical diagnosis. The results of Raman spectral analyses will also be correlated with an additional and novel spectral methodology such as Surface-Enhanced Raman Spectroscopy (SERS), Infrared (IR) Spectroscopy including Fourier-Transform Infrared (FTIR) Spectroscopy to

obtain molecular spectral analysis with higher sensitivity and complementary spectral insights on the tumor composition.

The cumulative results from our studies are expected to create a comprehensive spectral database that can be used to classify CNS tumors concerning the light microscopic analysis and molecular tests as the gold standard. The use of the advanced spectral analytical method is expected to supplement the existing knowledge about CNS tumors and allow a deeper and more comprehensive characterization of FFPE tissue sections using RS. After further refining Raman spectroscopic analysis of CNR tumors, one may investigate whether it might be possible to use spectral analysis to characterize molecular alterations and save precious time and financial resources by eliminating or complementing expensive genetic tests. RS-based molecular-level profiling of FFPE tissues can also be applied to fresh tissue samples and liquid biopsies as diagnostic tools to improve patient care.

Based on the results of the project, we aim to expand our analytical efforts to all types of GBs by transfer learning as well as other tumors of the CNS to identify unique molecular profiles and to establish a database that will be used to further classify/categorize CNS tumors.

BIBLIOGRAPHY

[1] P. D. Delgado-López and E. M. Corrales-García, "Survival in glioblastoma: a review on the impact of treatment modalities," *Clinical and Translational Oncology,* vol. 18, no. 11, pp. 1062-1071, 2016/11/01 2016, doi: 10.1007/s12094-016-1497-x.

[2] H. Ohgaki *et al.*, "Glioblastoma, IDH-mutant," in *WHO Classification of Tumours of the Central Nervous System*, H. O. David N. Louis, Otmar D. Wiestler, Webster K. Cavenee Ed., (Revised 4th edition). ed. Lyon: International Agency for Research on Cancer (IARC) 69372 Lyon Cedex 08, France, 2016, ch. 1, pp. 52-56.

[3] D. Louis *et al.*, "Glioblastoma, IDH-wild type," in *WHO Classification of Tumours of the Central Nervous System*, H. O. David N. Louis, Otmar D. Wiestler, Webster K. Cavenee Ed., (Revised 4th edition). ed. Lyon: International Agency for Research on Cancer (IARC) 69372 Lyon Cedex 08, France, 2016, ch. 1, pp. 28-51.

[4] D. N. Louis *et al.*, "The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary," *Acta Neuropathologica,* vol. 131, no. 6, pp. 803-820, 2016/06/01 2016, doi: 10.1007/s00401-016-1545-1.

[5] F. Hanif, K. Muzaffar, K. Perveen, S. M. Malhi, and S. U. Simjee, "Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment," (in eng), *Asian Pac J Cancer Prev,* vol. 18, no. 1, pp. 3-9, 2017, doi: 10.22034/APJCP.2017.18.1.3.

[6] W. E. Wild CP, Stewart BW, editors, "World Cancer Report: Cancer Research for Cancer Prevention," *Lyon, France: International Agency for Research on Cancer,* 2020. [Online]. Available: Available from: http://publications.iarc.fr/586.

[7] M. Yao *et al.*, "Cellular origin of glioblastoma and its implication in precision therapy," *Cellular & Molecular Immunology,* vol. 15, no. 8, pp. 737-739, 2018/08/01 2018, doi: 10.1038/cmi.2017.159.

[8] E. Denicolai *et al.*, "Molecular heterogeneity of glioblastomas: does location matter?," *Oncotarget,* vol. 7, no. 1, pp. 902-13, Jan 5 2016, doi: 10.18632/oncotarget.6433.

[9] D. Louis, "WHO classification and grading of tumours of the central nervous system," in *WHO Classification of Tumours of the Central Nervous System*, H. O. David N. Louis, Otmar D. Wiestler, Webster K. Cavenee Ed., R. t. edition). Ed. Lyon: International Agency for Research on Cancer (IARC) 69372 Lyon Cedex 08, France, 2016, pp. 12-13.

[10] D. N. Louis *et al.*, *WHO Classification of Tumours of the Central Nervous System*, (Revised 4th edition). ed. Lyon: International Agency for Research on Cancer (IARC) 69372 Lyon Cedex 08, France, 2016.

[11] R. Hoshide and R. Jandial, "2016 World Health Organization Classification of Central Nervous System Tumors: An Era of Molecular Biology," *World Neurosurg,* vol. 94, pp. 561-562, Oct 2016, doi: 10.1016/j.wneu.2016.07.082.

[12] W. K. Cavenee, D. N. Louis, H. Ohgaki, and O. D. Wiestler, *WHO classification of tumours of the central nervous system*. WHO Regional Office Europe, 2007.

[13] Y. Zhou *et al.*, "Optical biopsy identification and grading of gliomas using label-free visible resonance Raman spectroscopy," *J Biomed Opt,* vol. 24, no. 9, pp. 1-12, Sep 2019, doi: 10.1117/1.JBO.24.9.095001.
[14] Y. Zhou *et al.*, "Human brain cancer studied by resonance Raman spectroscopy," *J Biomed Opt,* vol. 17, no. 11, p. 116021, Nov 2012, doi: 10.1117/1.JBO.17.11.116021.
[15] K. Gajjar *et al.*, "Diagnostic segregation of human brain tumours using Fourier-transform infrared and/or Raman spectroscopy coupled with discriminant analysis," *Anal Methods,* vol. 5, pp. 89-102, Sep 6 2012, doi: 10.1039/C2AY25544H.
[16] P. Chand, S. Amit, R. Gupta, and A. Agarwal, "Errors, limitations, and pitfalls in the diagnosis of central and peripheral nervous system lesions in intraoperative cytology and frozen sections," *J Cytol,* vol. 33, no. 2, pp. 93-7, Apr-Jun 2016, doi: 10.4103/0970-9371.182530.
[17] S. Schipmann, M. Schwake, E. Suero Molina, and W. Stummer, "Markers for Identifying and Targeting Glioblastoma Cells during Surgery," *J Neurol Surg A Cent Eur Neurosurg,* vol. 80, no. 6, pp. 475-487, Nov 2019, doi: 10.1055/s-0039-1692976.
[18] T. Garzon-Muvdi, C. Kut, X. Li, and K. L. Chaichana, "Intraoperative imaging techniques for glioma surgery," *Future Oncol,* vol. 13, no. 19, pp. 1731-1745, Aug 2017, doi: 10.2217/fon-2017-0092.
[19] F. Ohka, A. Natsume, and T. Wakabayashi, "Current trends in targeted therapies for glioblastoma multiforme," *Neurol Res Int,* vol. 2012, p. 878425, 2012, doi: 10.1155/2012/878425.
[20] J. P. Thakkar *et al.*, "Epidemiologic and molecular prognostic review of glioblastoma," *Cancer Epidemiol Biomarkers Prev,* vol. 23, no. 10, pp. 1985-96, Oct 2014, doi: 10.1158/1055-9965.EPI-14-0275.
[21] K. Herrmann *et al.*, "Comparison of visual and semiquantitative analysis of 18F-FDOPA-PET/CT for recurrence detection in glioblastoma patients," *Neuro Oncol,* vol. 16, no. 4, pp. 603-9, Apr 2014, doi: 10.1093/neuonc/not166.
[22] J. Mooney *et al.*, "Current Approaches and Challenges in the Molecular Therapeutic Targeting of Glioblastoma," *World Neurosurg,* vol. 129, pp. 90-100, Sep 2019, doi: 10.1016/j.wneu.2019.05.205.
[23] C. Roder *et al.*, "Maximizing the extent of resection and survival benefit of patients in glioblastoma surgery: High-field iMRI versus conventional and 5-ALA-assisted surgery," *European Journal of Surgical Oncology (EJSO),* vol. 40, no. 3, pp. 297-304, 2014/03/01/ 2014, doi: https://doi.org/10.1016/j.ejso.2013.11.022.
[24] F. K. Albert, M. Forsting, K. Sartor, H. P. Adams, and S. Kunze, "Early postoperative magnetic resonance imaging after resection of malignant glioma: objective evaluation of residual tumor and its influence on regrowth and prognosis," *Neurosurgery,* vol. 34, no. 1, pp. 45-60; discussion 60-1, Jan 1994, doi: 10.1097/00006123-199401000-00008.
[25] O. Solheim, T. Selbekk, A. S. Jakola, and G. Unsgard, "Ultrasound-guided operations in unselected high-grade gliomas--overall results, impact of image quality and patient selection," *Acta Neurochir (Wien),* vol. 152, no. 11, pp. 1873-86, Nov 2010, doi: 10.1007/s00701-010-0731-5.
[26] N. Zhang *et al.*, "Sodium Fluorescein-Guided Resection under the YELLOW 560 nm Surgical Microscope Filter in Malignant Gliomas: Our First 38 Cases

Experience," *BioMed Research International,* vol. 2017, p. 7865747, 2017/10/16 2017, doi: 10.1155/2017/7865747.

[27] W. Stummer, H. Stepp, G. Moller, A. Ehrhardt, M. Leonhard, and H. J. Reulen, "Technical principles for protoporphyrin-IX-fluorescence guided microsurgical resection of malignant glioma tissue," *Acta Neurochir (Wien),* vol. 140, no. 10, pp. 995-1000, 1998, doi: 10.1007/s007010050206.

[28] H. A. Leroy, M. Vermandel, J. P. Lejeune, S. Mordon, and N. Reyns, "Fluorescence guided resection and glioblastoma in 2015: A review," *Lasers Surg Med,* vol. 47, no. 5, pp. 441-51, Jul 2015, doi: 10.1002/lsm.22359.

[29] A. P. Michael, V. L. Watson, D. Ryan, K. R. Delfino, S. V. Bekker, and J. W. Cozzens, "Effects of 5-ALA dose on resection of glioblastoma," *J Neurooncol,* vol. 141, no. 3, pp. 523-531, Feb 2019, doi: 10.1007/s11060-019-03100-7.

[30] -. A. R. G. S. Group, "FIVE-AMINOLEVULINIC ACID FOR FLUORESCENCE-GUIDED RESECTION OF RECURRENT MALIGNANT GLIOMAS: A PHASE II STUDY," *Neurosurgery,* vol. 65, no. 6, pp. 1070-1077, 2009, doi: 10.1227/01.Neu.0000360128.03597.C7.

[31] W. Stummer *et al.*, "5-Aminolevulinic Acid-derived Tumor Fluorescence: The Diagnostic Accuracy of Visible Fluorescence Qualities as Corroborated by Spectrometry and Histology and Postoperative Imaging," *Neurosurgery,* vol. 74, no. 3, pp. 310-320, 2013, doi: 10.1227/neu.0000000000000267.

[32] P. Teixidor *et al.*, "Safety and Efficacy of 5-Aminolevulinic Acid for High Grade Glioma in Usual Clinical Practice: A Prospective Cohort Study," (in eng), *Plos One,* vol. 11, no. 2, p. e0149244, 2016, doi: 10.1371/journal.pone.0149244.

[33] R. K. Hanson, R. M. Spearrin, and C. S. Goldenstein, "Rayleigh and Raman Spectra," in *Spectroscopy and Optical Diagnostics for Gases*, R. K. Hanson, R. M. Spearrin, and C. S. Goldenstein Eds. Cham: Springer International Publishing, 2016, pp. 91-105.

[34] J. R. F. a. K. N. a. C. W. Brown, "Introductory Raman Spectroscopy " *Elsevier,* 2003.

[35] N. Egeland, "Raman spectroscopy applied to enhanced oil recovery research," University of Stavanger, Norway, 2015.

[36] T. J. Johnson *et al.*, "Demonstrated Wavelength Portability of Raman Reference Data for Explosives and Chemical Detection," *International Journal of Spectroscopy,* vol. 2012, p. 297056, 2012/07/05 2012, doi: 10.1155/2012/297056.

[37] M. J. Baker *et al.*, "Developing and understanding biofluid vibrational spectroscopy: a critical review," *Chemical Society Reviews,* 10.1039/C5CS00585J vol. 45, no. 7, pp. 1803-1818, 2016, doi: 10.1039/C5CS00585J.

[38] D. I. Ellis and R. Goodacre, "Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy," *Analyst,* vol. 131, no. 8, pp. 875-85, Aug 2006, doi: 10.1039/b602376m.

[39] H. J. Butler *et al.*, "Using Raman spectroscopy to characterize biological materials," *Nat Protoc,* vol. 11, no. 4, pp. 664-87, Apr 2016, doi: 10.1038/nprot.2016.036.

[40] J. R. Ferraro, K. Nakamoto, and C. W. Brown, "Chapter 6 - Biochemical and Medical Applications," in *Introductory Raman Spectroscopy (Second Edition)*, J. R. Ferraro, K. Nakamoto, and C. W. Brown Eds. San Diego: Academic Press, 2003, pp. 295-324.

[41] P. Crow, A. Molckovsky, N. Stone, J. Uff, B. Wilson, and L. M. WongKeeSong, "Assessment of fiberoptic near-infrared raman spectroscopy for diagnosis of bladder and prostate cancer," *Urology,* vol. 65, no. 6, pp. 1126-30, Jun 2005, doi: 10.1016/j.urology.2004.12.058.

[42] T. J. Harvey *et al.*, "Spectral discrimination of live prostate and bladder cancer cell lines using Raman optical tweezers," *Journal of Biomedical Optics,* vol. 13, no. 6, pp. 1-12, 11/1 2008, doi: 10.1117/1.2999609.

[43] K. Lin, D. L. P. Cheng, and Z. Huang, "Optical diagnosis of laryngeal cancer using high wavenumber Raman spectroscopy," *Biosens Bioelectron,* vol. 35, no. 1, pp. 213-217, May 15 2012, doi: 10.1016/j.bios.2012.02.050.

[44] K. Lin, D. L. P. Cheng, and Z. Huang, "Optical diagnosis of laryngeal cancer using high wavenumber Raman spectroscopy," *Biosensors and Bioelectronics,* vol. 35, no. 1, pp. 213-217, 2012/05/15/ 2012, doi: https://doi.org/10.1016/j.bios.2012.02.050.

[45] W. C. Zúñiga *et al.*, "Raman Spectroscopy for Rapid Evaluation of Surgical Margins during Breast Cancer Lumpectomy," *Sci Rep-Uk,* vol. 9, no. 1, p. 14639, 2019/10/10 2019, doi: 10.1038/s41598-019-51112-0.

[46] H. Abramczyk, J. Surmacki, B. Brożek-Płuska, Z. Morawiec, and M. Tazbir, "The hallmarks of breast cancer by Raman spectroscopy," *Journal of Molecular Structure,* vol. 924-926, pp. 175-182, 2009/04/30/ 2009, doi: https://doi.org/10.1016/j.molstruc.2008.12.055.

[47] Z. Huang, A. McWilliams, H. Lui, D. I. McLean, S. Lam, and H. Zeng, "Near-infrared Raman spectroscopy for optical diagnosis of lung cancer," *Int J Cancer,* vol. 107, no. 6, pp. 1047-52, Dec 20 2003, doi: 10.1002/ijc.11500.

[48] L. Gao *et al.*, "Differential diagnosis of lung carcinoma with coherent anti-Stokes Raman scattering imaging," *Arch Pathol Lab Med,* vol. 136, no. 12, pp. 1502-10, Dec 2012, doi: 10.5858/arpa.2012-0238-SA.

[49] E. M. Kanter *et al.*, "Application of Raman spectroscopy for cervical dysplasia diagnosis," *J Biophotonics,* vol. 2, no. 1-2, pp. 81-90, Feb 2009, doi: 10.1002/jbio.200910001.

[50] U. Utzinger, D. L. Heintzelman, A. Mahadevan-Jansen, A. Malpica, M. Follen, and R. Richards-Kortum, "Near-Infrared Raman Spectroscopy for in Vivo Detection of Cervical Precancers," *Appl. Spectrosc.,* vol. 55, no. 8, pp. 955-959, 2001/08/01 2001. [Online]. Available: http://as.osa.org/abstract.cfm?URI=as-55-8-955.

[51] C. A. Lieber, S. K. Majumder, D. L. Ellis, D. D. Billheimer, and A. Mahadevan-Jansen, "In vivo nonmelanoma skin cancer diagnosis using Raman microspectroscopy," (in eng), *Lasers in surgery and medicine,* vol. 40, no. 7, pp. 461-467, 2008, doi: 10.1002/lsm.20653.

[52] S. K. Teh, W. Zheng, K. Y. Ho, M. Teh, K. G. Yeoh, and Z. Huang, "Near-infrared Raman spectroscopy for early diagnosis and typing of adenocarcinoma in the stomach," *Br J Surg,* vol. 97, no. 4, pp. 550-7, Apr 2010, doi: 10.1002/bjs.6913.

[53] M. S. Bergholt *et al.*, "In vivo diagnosis of gastric cancer using Raman endoscopy and ant colony optimization techniques," (in eng), *International journal of cancer,* vol. 128, no. 11, pp. 2673-2680, 2011/06// 2011, doi: 10.1002/ijc.25618.

[54] A. K. Pandya *et al.*, "Evaluation of pancreatic cancer with Raman spectroscopy in a mouse model," *Pancreas,* vol. 36, no. 2, pp. e1-8, Mar 2008, doi: 10.1097/MPA.0b013e31815a3f1c.

[55] B. Brozek-Pluska, J. Musial, R. Kordek, and H. Abramczyk, "Analysis of Human Colon by Raman Spectroscopy and Imaging-Elucidation of Biochemical Changes in Carcinogenesis," (in eng), *International journal of molecular sciences,* vol. 20, no. 14, p. 3398, 2019, doi: 10.3390/ijms20143398.

[56] L. J. Livermore *et al.*, "Rapid intraoperative molecular genetic classification of gliomas using Raman spectroscopy," *Neurooncol Adv,* vol. 1, no. 1, p. vdz008, May-Dec 2019, doi: 10.1093/noajnl/vdz008.

[57] G. M. Shankar *et al.*, "Rapid Intraoperative Molecular Characterization of Glioma," *JAMA Oncol,* vol. 1, no. 5, pp. 662-7, Aug 2015, doi: 10.1001/jamaoncol.2015.0917.

[58] R. Kast *et al.*, "Identification of regions of normal grey matter and white matter from pathologic glioblastoma and necrosis in frozen sections using Raman imaging," *J Neurooncol,* vol. 125, no. 2, pp. 287-95, Nov 2015, doi: 10.1007/s11060-015-1929-4.

[59] S. N. Kalkanis *et al.*, "Raman spectroscopy to distinguish grey matter, necrosis, and glioblastoma multiforme in frozen tissue sections," *J Neurooncol,* vol. 116, no. 3, pp. 477-85, Feb 2014, doi: 10.1007/s11060-013-1326-9.

[60] D. A. Orringer *et al.*, "Rapid intraoperative histology of unprocessed surgical specimens via fibre-laser-based stimulated Raman scattering microscopy," *Nat Biomed Eng,* vol. 1, 2017, doi: 10.1038/s41551-016-0027.

[61] M. Ji *et al.*, "Rapid, label-free detection of brain tumors with stimulated Raman scattering microscopy," *Sci Transl Med,* vol. 5, no. 201, p. 201ra119, Sep 4 2013, doi: 10.1126/scitranslmed.3005954.

[62] J. W. Chan, D. S. Taylor, and D. L. Thompson, "The effect of cell fixation on the discrimination of normal and leukemia cells with laser tweezers Raman spectroscopy," *Biopolymers,* vol. 91, no. 2, pp. 132-9, Feb 2009, doi: 10.1002/bip.21094.

[63] J. W. Chan, D. S. Taylor, S. M. Lane, T. Zwerdling, J. Tuscano, and T. Huser, "Nondestructive identification of individual leukemia cells by laser trapping Raman spectroscopy," *Anal Chem,* vol. 80, no. 6, pp. 2180-7, Mar 15 2008, doi: 10.1021/ac7022348.

[64] R. E. Kast *et al.*, "Raman molecular imaging of brain frozen tissue sections," (in English), *J Neuro-Oncol,* vol. 120, no. 1, pp. 55-62, Oct 2014, doi: 10.1007/s11060-014-1536-9.

[65] L. J. Livermore *et al.*, "Raman spectroscopy to differentiate between fresh tissue samples of glioma and normal brain: a comparison with 5-ALA–induced fluorescence-guided surgery," *Journal of Neurosurgery,* vol. 1, no. aop, pp. 1-11, 2020.

[66] L. A. Austin, S. Osseiran, and C. L. Evans, "Raman technologies in cancer diagnostics," *Analyst,* vol. 141, no. 2, pp. 476-503, Jan 21 2016, doi: 10.1039/c5an01786f.

[67] M. Jermyn *et al.*, "Intraoperative brain cancer detection with Raman spectroscopy in humans," *Sci Transl Med,* vol. 7, no. 274, p. 274ra19, Feb 11 2015, doi: 10.1126/scitranslmed.aaa2384.

[68] J. Desroches *et al.*, "A new method using Raman spectroscopy for in vivo targeted brain cancer tissue biopsy," *Sci Rep,* vol. 8, no. 1, p. 1792, Jan 29 2018, doi: 10.1038/s41598-018-20233-3.
[69] J. Desroches *et al.*, "Development and first in-human use of a Raman spectroscopy guidance system integrated with a brain biopsy needle," *J Biophotonics,* vol. 12, no. 3, p. e201800396, Mar 2019, doi: 10.1002/jbio.201800396.
[70] M. Ji *et al.*, "Detection of human brain tumor infiltration with quantitative stimulated Raman scattering microscopy," *Sci Transl Med,* vol. 7, no. 309, p. 309ra163, Oct 14 2015, doi: 10.1126/scitranslmed.aab0195.
[71] L. Wang *et al.*, "Evaluation of Raman spectroscopy for diagnosing EGFR mutation status in lung adenocarcinoma," *Analyst,* vol. 139, no. 2, pp. 455-63, Jan 21 2014, doi: 10.1039/c3an01381b.
[72] X. Bi, B. Rexer, C. L. Arteaga, M. Guo, and A. Mahadevan-Jansen, "Evaluating HER2 amplification status and acquired drug resistance in breast cancer cells using Raman spectroscopy," (in eng), *Journal of biomedical optics,* vol. 19, no. 2, pp. 025001-025001, 2014, doi: 10.1117/1.JBO.19.2.025001.
[73] J. Zhao, H. M. Lai, Y. Qi, D. He, and H. Sun, "Current Status of Tissue Clearing and the Path Forward in Neuroscience," *ACS Chemical Neuroscience,* vol. 12, no. 1, pp. 5-29, 2021/01/06 2021, doi: 10.1021/acschemneuro.0c00563.
[74] O. Ibrahim *et al.*, "Improved protocols for pre-processing Raman spectra of formalin fixed paraffin preserved tissue sections," *Analytical Methods,* 10.1039/C6AY03308C vol. 9, no. 32, pp. 4709-4717, 2017, doi: 10.1039/C6AY03308C.
[75] E. O. Faolain *et al.*, "Raman spectroscopic evaluation of efficacy of current paraffin wax section dewaxing agents," *J Histochem Cytochem,* vol. 53, no. 1, pp. 121-9, Jan 2005, doi: 10.1177/002215540505300114.
[76] A. T. Lewis *et al.*, "Mirrored stainless steel substrate provides improved signal for Raman spectroscopy of tissue and cells," *Journal of Raman Spectroscopy,* vol. 48, no. 1, pp. 119-125, 2017.
[77] L. T. Kerr, H. J. Byrne, and B. M. Hennelly, "Optimal choice of sample substrate and laser wavelength for Raman spectroscopic analysis of biological specimen," *Analytical Methods,* 10.1039/C5AY00327J vol. 7, no. 12, pp. 5041-5052, 2015, doi: 10.1039/C5AY00327J.
[78] N. Stone, C. Kendall, J. Smith, P. Crow, and H. Barr, "Raman spectroscopy for identification of epithelial cancers," *Faraday Discuss,* vol. 126, pp. 141-57; discussion 169-83, 2004, doi: 10.1039/b304992b.
[79] C. Krafft, D. Codrich, G. Pelizzo, and V. Sergo, "Raman mapping and FTIR imaging of lung tissue: congenital cystic adenomatoid malformation," *Analyst,* vol. 133, no. 3, pp. 361-71, Mar 2008, doi: 10.1039/b712958k.
[80] D. Bury *et al.*, "Discrimination of fresh frozen non-tumour and tumour brain tissue using spectrochemical analyses and a classification model," *Br J Neurosurg,* vol. 34, no. 1, pp. 40-45, Feb 2020, doi: 10.1080/02688697.2019.1679352.
[81] U. Parlatan *et al.*, "Raman spectroscopy as a non-invasive diagnostic technique for endometriosis," *Sci Rep,* vol. 9, no. 1, p. 19795, Dec 24 2019, doi: 10.1038/s41598-019-56308-y.
[82] S. Koljenovic, L. P. Choo-Smith, T. C. Bakker Schut, J. M. Kros, H. J. van den Berge, and G. J. Puppels, "Discriminating vital tumor from necrotic tissue in

human glioblastoma tissue samples by Raman spectroscopy," *Lab Invest,* vol. 82, no. 10, pp. 1265-77, Oct 2002, doi: 10.1097/01.lab.0000032545.96931.b8.

[83] M. A. Bujang and T. H. Adnan, "Requirements for Minimum Sample Size for Sensitivity and Specificity Analysis," *J Clin Diagn Res,* vol. 10, no. 10, pp. YE01-YE06, Oct 2016, doi: 10.7860/JCDR/2016/18129.8744.

[84] Leica. "Clearene Safety Datasheet, Version 1.0, Leica Biosystems." https://shop.leicabiosystems.com/us/histology-consumables/reagents-solutions/pid-clearene-solvent (accessed May, 2020.

[85] Sigma-Aldrich. "Material Safety Data Sheet, Version 6.3, Merck Milipore, Sigma-Aldrich." https://www.sigmaaldrich.com/MSDS/MSDS/DisplayMSDSPage.do?country=US&language=en&productNumber=296333&brand=SIAL&PageToGoToURL=https%3A%2F%2Fwww.sigmaaldrich.com%2Fcatalog%2Fproduct%2Fsial%2F296333%3Flang%3Den (accessed May, 2020).

[86] F. L. Carson, C. H. Cappellano, and P. American Society for Clinical, *Histotechnology : a self instructional text*. (in English), 2015.

[87] R. Gautam, S. Vanga, F. Ariese, and S. Umapathy, "Review of multidimensional data processing approaches for Raman and infrared spectroscopy," *EPJ Techniques and Instrumentation,* vol. 2, no. 1, p. 8, 2015/06/02 2015, doi: 10.1140/epjti/s40485-015-0018-6.

[88] X. He, Y. Liu, S. Huang, Y. Liu, X. Pu, and T. Xu, "Raman spectroscopy coupled with principal component analysis to quantitatively analyze four crystallographic phases of explosive CL-20," *RSC Advances,* 10.1039/C8RA02189A vol. 8, no. 41, pp. 23348-23352, 2018, doi: 10.1039/C8RA02189A.

[89] H. Adams, B. Carr, and E. Farnell, "Support Vector Machines and Radon's Theorem," *arXiv preprint arXiv:2011.00617,* 2020.

[90] T. Eitrich and B. Lang, "Efficient optimization of support vector machine learning parameters for unbalanced datasets," *Journal of Computational and Applied Mathematics,* vol. 196, no. 2, pp. 425-436, 2006/11/15/ 2006, doi: https://doi.org/10.1016/j.cam.2005.09.009.

[91] H. S. Khamis, K. W. Cheruiyot, and S. Kimani, "Application of k-nearest neighbour classification in medical data mining," *International Journal of Information and Communication Technology Research,* vol. 4, no. 4, 2014.

[92] H. Shee, K. Cheruiyot, and S. Kimani, "Application of k-Nearest Neighbour Classification in Medical Data Mining," vol. 4, 04/04 2014.

[93] C. Nguyen, Y. Wang, and H.-N. Nguyen, "Random forest classifier combined with feature selection for breast cancer diagnosis and prognostic," *Journal of Biomedical Science and Engineering,* vol. 06, pp. 551-560, 01/01 2013, doi: 10.4236/jbise.2013.65070.

[94] K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: from early developments to recent advancements," *Systems Science & Control Engineering,* vol. 2, no. 1, pp. 602-609, 2014/12/01 2014, doi: 10.1080/21642583.2014.956265.

[95] S. Mian, H. Colley, M. Thornhill, and I. Rehman, "Development of a Dewaxing Protocol for Tissue-Engineered Models of the Oral Mucosa Used for Raman Spectroscopic Analysis," *Applied Spectroscopy Reviews,* vol. 49, 10/31 2014, doi: 10.1080/05704928.2014.882348.

[96] F. Lyng and E. Gazi, "Preparation of Tissues and Cells for Infrared and Raman Spectroscopy and Imaging," *Biomedical Applications of Synchrotron Infrared Microspectroscopy: A Practical Approach,* 01/01 2011.
[97] M. A. Puerto and N. M. Balzaretti, "Raman and infrared vibrational modes of tricosane paraffin under high pressure," *Vibrational Spectroscopy,* vol. 75, pp. 93-100, 2014/11/01/ 2014, doi: https://doi.org/10.1016/j.vibspec.2014.10.004.
[98] A. Douplik, G. Saiko, I. Schelkanova, and V. V. Tuchin, "The response of tissue to laser light," in *Lasers for Medical Applications*: Elsevier, 2013, pp. 47-109.
[99] J. Gordon and S. Harman, "A Graduated Cylinder Colorimeter: An Investigation of Path Length and the Beer-Lambert Law," *Journal of Chemical Education,* vol. 79, no. 5, p. 611, 2002/05/01 2002, doi: 10.1021/ed079p611.
[100] G. Wypych, *Handbook of UV degradation and stabilization*. Elsevier, 2020.
[101] N. T. Tsendzughul and A. A. Ogwu, "Chapter 12 - Visible light activated antimicrobial silver oxide thin films," in *Advances in Medical and Surgical Engineering*, W. Ahmed, D. A. Phoenix, M. J. Jackson, and C. P. Charalambous Eds.: Academic Press, 2020, pp. 177-238.
[102] M. Sauer, J. Hofkens, and J. Enderlein, *Handbook of fluorescence spectroscopy and imaging: from ensemble to single molecules*. John Wiley & Sons, 2010.
[103] S. Yang *et al.*, "Laser Wavelength Dependence of Background Fluorescence in Raman Spectroscopic Analysis of Synovial Fluid from Symptomatic Joints," (in eng), *Journal of Raman spectroscopy : JRS,* vol. 44, no. 8, pp. 1089-1095, 2013, doi: 10.1002/jrs.4338.
[104] D. Cebeci, P. Wang, M. A. Alam, R. Pinal, and D. Ben-Amotz, "Photobleaching profile of Raman peaks and fluorescence background," *European Pharmaceutical Review,* vol. 22, pp. 18-21, 12/01 2017.
[105] C. Krafft, L. Neudert, T. Simat, and R. Salzer, "Near infrared Raman spectra of human brain lipids," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy,* vol. 61, no. 7, pp. 1529-1535, 2005/05/01/ 2005, doi: https://doi.org/10.1016/j.saa.2004.11.017.
[106] S. Budday *et al.*, "Mechanical properties of gray and white matter brain tissue by indentation," *Journal of the Mechanical Behavior of Biomedical Materials,* vol. 46, pp. 318-330, 2015/06/01/ 2015, doi: https://doi.org/10.1016/j.jmbbm.2015.02.024.
[107] J. Song *et al.*, "Dysregulation of neuron differentiation in an autistic savant with exceptional memory," *Molecular brain,* vol. 12, no. 1, pp. 1-12, 2019.
[108] A. Beljebbar, S. Dukic, N. Amharref, and M. Manfait, "Ex vivo and in vivo diagnosis of C6 glioblastoma development by Raman spectroscopy coupled to a microprobe," *Analytical and Bioanalytical Chemistry,* vol. 398, no. 1, pp. 477-487, 2010/09/01 2010, doi: 10.1007/s00216-010-3910-6.
[109] R. L. Davis and D. M. Robertson, *Textbook of neuropathology*. Williams & Wilkins, 1985.
[110] F. Di Pauli and T. Berger, "Myelin Oligodendrocyte Glycoprotein Antibody-Associated Disorders: Toward a New Spectrum of Inflammatory Demyelinating CNS Disorders?," (in English), *Frontiers in Immunology,* Review vol. 9, no. 2753, 2018-November-29 2018, doi: 10.3389/fimmu.2018.02753.
[111] J. W. Chan, D. S. Taylor, T. Zwerdling, S. M. Lane, K. Ihara, and T. Huser, "Micro-Raman spectroscopy detects individual neoplastic and normal hematopoietic cells," *Biophysical journal,* vol. 90, no. 2, pp. 648-656, 2006.

[112] Z. Movasaghi, S. Rehman, and I. U. Rehman, "Raman Spectroscopy of Biological Tissues," *Applied Spectroscopy Reviews,* vol. 42, no. 5, pp. 493-541, 2007/09/01 2007, doi: 10.1080/05704920701551530.
[113] D. Guzmán-Embús, M. Orrego Cardozo, and C. Vargas-Hernández, "Genomic DNA characterization of pork spleen by Raman spectroscopy," *J Appl Phys,* vol. 114, no. 19, p. 194704, 2013.
[114] R. J. Lakshmi, V. Kartha, C. Murali Krishna, J. R Solomon, G. Ullas, and P. Uma Devi, "Tissue Raman spectroscopy for the study of radiation damage: brain irradiation of mice," *Radiation research,* vol. 157, no. 2, pp. 175-182, 2002.
[115] S. Koljenović, T. B. Schut, A. Vincent, J. M. Kros, and G. J. Puppels, "Detection of meningioma in dura mater by Raman spectroscopy," *Analytical chemistry,* vol. 77, no. 24, pp. 7958-7965, 2005.
[116] N. Stone, C. Kendall, N. Shepherd, P. Crow, and H. Barr, "Near-infrared Raman spectroscopy for the classification of epithelial pre-cancers and cancers," *Journal of Raman Spectroscopy,* vol. 33, no. 7, pp. 564-573, 2002, doi: 10.1002/jrs.882.
[117] I. Notingher *et al.*, "Discrimination between ricin and sulphur mustard toxicity in vitro using Raman spectroscopy," (in eng), *J R Soc Interface,* vol. 1, no. 1, pp. 79-90, 2004/11// 2004, doi: 10.1098/rsif.2004.0008.
[118] W. T. Cheng, M. T. Liu, H. N. Liu, and S. Y. Lin, "Micro-Raman spectroscopy used to identify and grade human skin pilomatrixoma," *Microscopy research and technique,* vol. 68, no. 2, pp. 75-79, 2005.
[119] J. S. Singh, "IR and Raman spectra with Gaussian-09 molecular analysis of some other parameters and vibrational spectra of 5-fluoro-uracil," *Research on Chemical Intermediates,* vol. 46, no. 5, pp. 2457-2479, 2020/05/01 2020, doi: 10.1007/s11164-020-04101-2.
[120] P. N. Banedar, "A Raman Spectroscopic Study of Solid Dispersions and Co-crystals During the Pharmaceutical Hot melt Extrusion Process," University of Bradford, 2015.
[121] Y. Koyama, I. Takatsuka, M. Nakata, and M. Tasumi, "Raman and infrared spectra of the all-trans, 7-cis, 9-cis, 13-cis and 15-cis isomers of β-carotene: Key bands distinguishing stretched or terminal-bent configurations form central-bent configurations," *Journal of Raman Spectroscopy,* vol. 19, no. 1, pp. 37-49, 1988, doi: 10.1002/jrs.1250190107.
[122] S. O. Konorov, H. G. Schulze, J. M. Piret, R. F. B. Turner, and M. W. Blades, "Evidence of marked glycogen variations in the characteristic Raman signatures of human embryonic stem cells," *Journal of Raman Spectroscopy,* vol. 42, no. 5, pp. 1135-1141, 2011/05/01 2011, doi: 10.1002/jrs.2829.
[123] A.-A. Grosset *et al.*, "Identification of intraductal carcinoma of the prostate on tissue specimens using Raman micro-spectroscopy: A diagnostic accuracy case–control study with multicohort validation," *PLoS medicine,* vol. 17, no. 8, p. e1003281, 2020.
[124] H. G. Schulze, S. O. Konorov, J. M. Piret, M. W. Blades, and R. F. B. Turner, "Label-free imaging of mammalian cell nucleoli by Raman microspectroscopy," *Analyst,* 10.1039/C3AN00118K vol. 138, no. 12, pp. 3416-3423, 2013, doi: 10.1039/C3AN00118K.
[125] A. J. Ruiz-Chica, M. A. Medina, F. Sánchez-Jiménez, and F. J. Ramírez, "Characterization by Raman spectroscopy of conformational changes on guanine–cytosine and adenine–thymine oligonucleotides induced by aminooxy

analogues of spermidine," *Journal of Raman Spectroscopy,* vol. 35, no. 2, pp. 93-100, 2004/02/01 2004, doi: 10.1002/jrs.1107.

[126] K. Wieland *et al.*, "In-Line Ultrasound-Enhanced Raman Spectroscopy Allows for Highly Sensitive Analysis with Improved Selectivity in Suspensions," *Analytical Chemistry,* vol. 91, no. 22, pp. 14231-14238, 2019/11/19 2019, doi: 10.1021/acs.analchem.9b01105.

[127] E. Hanlon *et al.*, "Prospects for in vivo Raman spectroscopy," *Physics in Medicine & Biology,* vol. 45, no. 2, p. R1, 2000.

[128] R. Manoharan *et al.*, "Raman Spectroscopy and Fluorescence Photon Migration for Breast Cancer Diagnosis and Imaging," *Photochemistry and Photobiology,* vol. 67, no. 1, pp. 15-22, 1998/01/01 1998, doi: 10.1111/j.1751-1097.1998.tb05160.x.

[129] E. Lemoine *et al.*, "Feature engineering applied to intraoperative in vivo Raman spectroscopy sheds light on molecular processes in brain cancer: a retrospective study of 65 patients," *Analyst,* vol. 144, no. 22, pp. 6517-6532, Nov 4 2019, doi: 10.1039/c9an01144g.

[130] M. Tsuboi *et al.*, "Raman scattering tensors of tyrosine," (in eng), *Biospectroscopy,* vol. 4, no. 1, pp. 61-71, 1998, doi: 10.1002/(sici)1520-6343(1998)4:1<61::aid-bspy7>3.0.co;2-v.

[131] C. J. Frank, R. L. McCreery, and D. C. B. Redd, "Raman Spectroscopy of Normal and Diseased Human Breast Tissues," *Analytical Chemistry,* vol. 67, no. 5, pp. 777-783, 1995/03/01 1995, doi: 10.1021/ac00101a001.

[132] H. Schulz and M. Baranska, "Identification and quantification of valuable plant substances by IR and Raman spectroscopy," *Vibrational Spectroscopy,* vol. 43, no. 1, pp. 13-25, 2007/01/16/ 2007, doi: https://doi.org/10.1016/j.vibspec.2006.06.001.

[133] W.-P. Hsieh and Y.-H. Chien, "High pressure Raman spectroscopy of H2O-CH3OH mixtures," *Sci Rep-Uk,* vol. 5, no. 1, p. 8532, 2015/02/23 2015, doi: 10.1038/srep08532.

[134] Z. Hu, X. Wang, W. Wang, Z. Zhang, H. Gao, and Y. Mao, "Raman spectroscopy for detecting supported planar lipid bilayers composed of ganglioside-GM1/sphingomyelin/cholesterol in the presence of amyloid-β," *Physical Chemistry Chemical Physics,* 10.1039/C5CP02366A vol. 17, no. 35, pp. 22711-22720, 2015, doi: 10.1039/C5CP02366A.

[135] E. Timchenko, P. Timchenko, L. Volova, S. Pershutkina, and P. Shalkovskaya, "Testing of Raman spectroscopy method for assessment of skin implants," in *Journal of Physics. Conference Series (Online)*, 2016, vol. 769, no. 1.

[136] A. Kunstar *et al.*, "Label-free Raman monitoring of extracellular matrix formation in three-dimensional polymeric scaffolds," (in eng), *J R Soc Interface,* vol. 10, no. 86, p. 20130464, Sep 6 2013, doi: 10.1098/rsif.2013.0464.

[137] B. M. Murphy, S. W. Prescott, and I. Larson, "Measurement of lactose crystallinity using Raman spectroscopy," *Journal of Pharmaceutical and Biomedical Analysis,* vol. 38, no. 1, pp. 186-190, 2005/06/01/ 2005, doi: https://doi.org/10.1016/j.jpba.2004.12.013.

[138] A. Mahadevan-Jansen and R. Richards-Kortum, "Raman spectroscopy for cancer detection: a review," in *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.'Magnificent Milestones and Emerging Opportunities in Medical Engineering'(Cat. No. 97CH36136)*, 1997, vol. 6: IEEE, pp. 2722-2728.

[139] G. Puppels, H. Garritsen, J. Kummer, and J. Greve, "Carotenoids located in human lymphocyte subpopulations and natural killer cells by Raman microspectroscopy," *Cytometry: The Journal of the International Society for Analytical Cytology,* vol. 14, no. 3, pp. 251-256, 1993.

[140] L. Silveira Jr, S. Sathaiah, R. A. Zângaro, M. T. Pacheco, M. C. Chavantes, and C. A. Pasqualucci, "Correlation between near-infrared Raman spectroscopy and the histopathological analysis of atherosclerosis in human coronary arteries," *Lasers in Surgery and Medicine: The Official Journal of the American Society for Laser Medicine and Surgery,* vol. 30, no. 4, pp. 290-297, 2002.

[141] B. S. Ferreira, C. G. de Almeida, M. Le Hyaric, V. E. de Oliveira, H. G. M. Edwards, and L. F. C. de Oliveira, "Raman Spectroscopic Investigation of Carotenoids in Oils from Amazonian Products," *Spectroscopy Letters,* vol. 46, no. 2, pp. 122-127, 2013/02/01 2013, doi: 10.1080/00387010.2012.693569.

[142] S. P. Verma and D. F. Hoelzl Wallach, "Carotenoids as raman-active probes of erythrocyte membrane structure," *Biochimica et Biophysica Acta (BBA) - Biomembranes,* vol. 401, no. 2, pp. 168-176, 1975/08/20/ 1975, doi: https://doi.org/10.1016/0005-2736(75)90301-6.

[143] Z. Huang, H. Lui, D. I. McLean, M. Korbelik, and H. Zeng, "Raman spectroscopy in combination with background near-infrared autofluorescence enhances the in vivo assessment of malignant tissues," *Photochemistry and photobiology,* vol. 81, no. 5, pp. 1219-1226, 2005.

[144] L. Yu, S. Wu, Y. Liu, P. Song, and L. Xia, "Distance-regulating surface plasmon catalyzed coupling reaction of p-nitrophenyl disulfide," *RSC advances,* vol. 8, no. 62, pp. 35646-35650, 2018.

[145] Y. Xie, D. Zhang, G. K. Jarori, V. J. Davisson, and D. Ben-Amotz, "The Raman detection of peptide tyrosine phosphorylation," *Analytical Biochemistry,* vol. 332, no. 1, pp. 116-121, 2004/09/01/ 2004, doi: https://doi.org/10.1016/j.ab.2004.05.052.

[146] E. Ó Faoláin *et al.*, "A study examining the effects of tissue processing on human tissue sections using vibrational spectroscopy," *Vibrational Spectroscopy,* vol. 38, no. 1, pp. 121-127, 2005/07/29/ 2005, doi: https://doi.org/10.1016/j.vibspec.2005.02.013.

[147] H. Ishizaki, P. Balaram, R. Nagaraj, Y. V. Venkatachalapathi, and A. T. Tu, "Determination of beta-turn conformation by laser Raman spectroscopy," (in eng), *Biophysical journal,* vol. 36, no. 3, pp. 509-517, 1981, doi: 10.1016/S0006-3495(81)84749-2.

[148] S. Chen, L. Dong, M. Yan, Z. Dai, C. Sun, and X. Li, "Rapid and sensitive biomarker detection using molecular imprinting polymer hydrogel and surface-enhanced Raman scattering," *Royal Society open science,* vol. 5, no. 1, p. 171488, 2018.

[149] Y. Jin, A. P. Kotula, A. R. Hight Walker, K. B. Migler, and Y. J. Lee, "Phase-Specific Raman Analysis of n-Alkane Melting by Moving-Window Two-Dimensional Correlation Spectroscopy," (in eng), *Journal of Raman spectroscopy : JRS,* vol. 47, no. 11, pp. 1375-1384, 2016, doi: 10.1002/jrs.4967.

[150] J.-E. Clément, A. Leray, A. Bouhelier, and E. Finot, "Spectral pointillism of enhanced Raman scattering for accessing structural and conformational information on single protein," *Physical Chemistry Chemical Physics,* 10.1039/C6CP06667D vol. 19, no. 1, pp. 458-466, 2017, doi: 10.1039/C6CP06667D.

[151] G. J. Thomas Jr *et al.*, "Polarized Raman spectra of oriented fibers of A DNA and B DNA: anisotropic and isotropic local Raman tensors of base and backbone vibrations," *Biophysical journal,* vol. 68, no. 3, pp. 1073-1088, 1995.

[152] K. E. Shafer-Peltier *et al.*, "Raman microspectroscopic model of human breast tissue: implications for breast cancer diagnosis in vivo," *Journal of Raman Spectroscopy,* vol. 33, no. 7, pp. 552-563, 2002.

[153] P. Timchenko, E. Timchenko, L. Volova, D. Dolgushkin, V. Boltovskaya, and O. Frolov, "Raman spectroscopy method for the evaluation of bone bioimplants made using the «lyoplast» technology from cadaveric and in vivo resected bone tissue," in *Journal of Physics: Conference Series*, 2018, vol. 1038, no. 1.

[154] B. R. Wood, B. Tait, and D. McNaughton, "Micro-Raman characterisation of the R to T state transition of haemoglobin within a single living erythrocyte," *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research,* vol. 1539, no. 1, pp. 58-70, 2001/05/28/ 2001, doi: https://doi.org/10.1016/S0167-4889(01)00089-1.

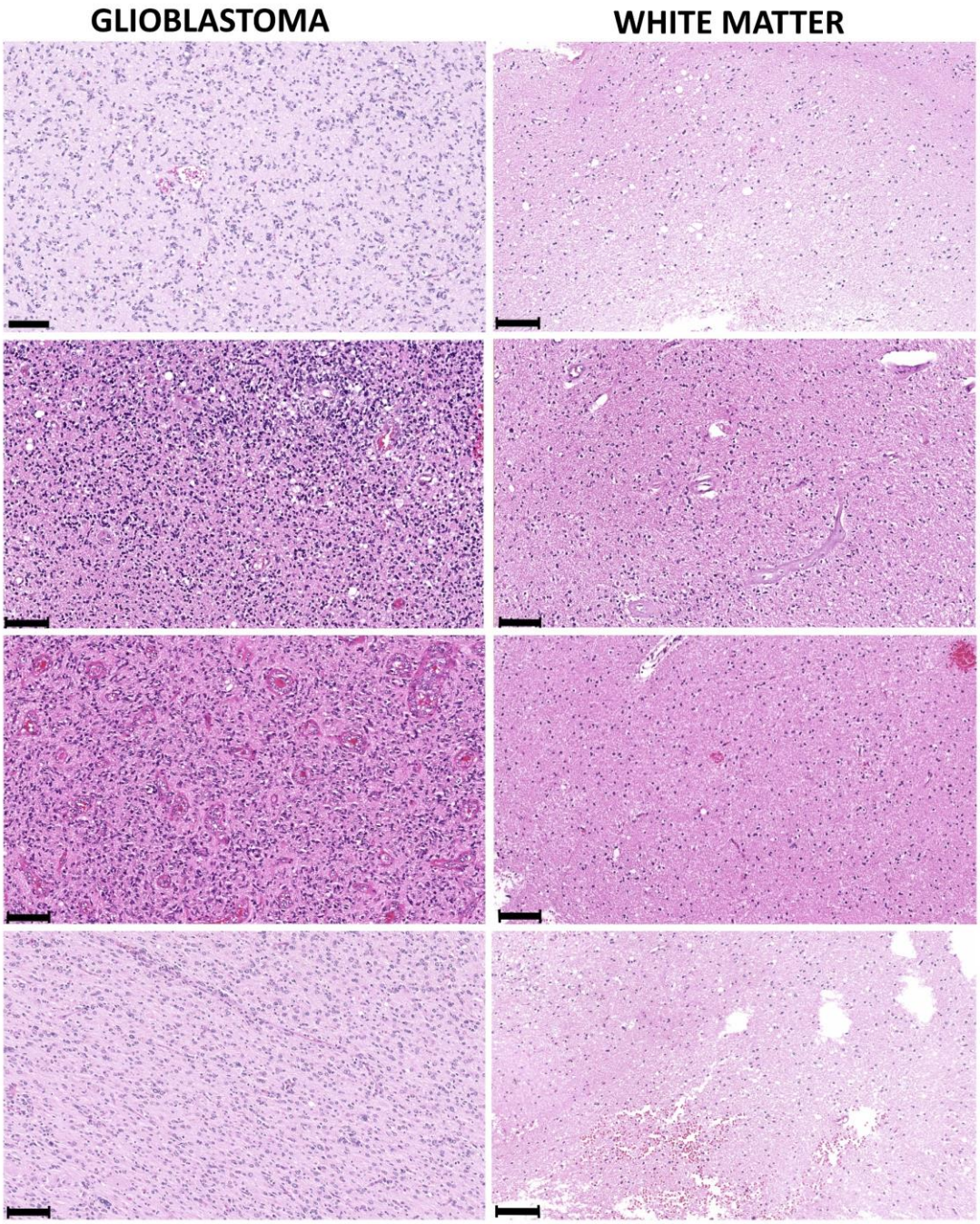
[155] A. G. Tzakos *et al.*, "Structure and function of the myelin proteins: current status and perspectives in relation to multiple sclerosis," *Current medicinal chemistry,* vol. 12, no. 13, pp. 1569-1587, 2005.

[156] N. a. T. Batur, Hulya. "Glioblastoma Raman Spectra Classification using SVM, KNN, and RF." https://github.com/HulyaTorun/Glioblastoma-Raman-Spectra-Classification (accessed.

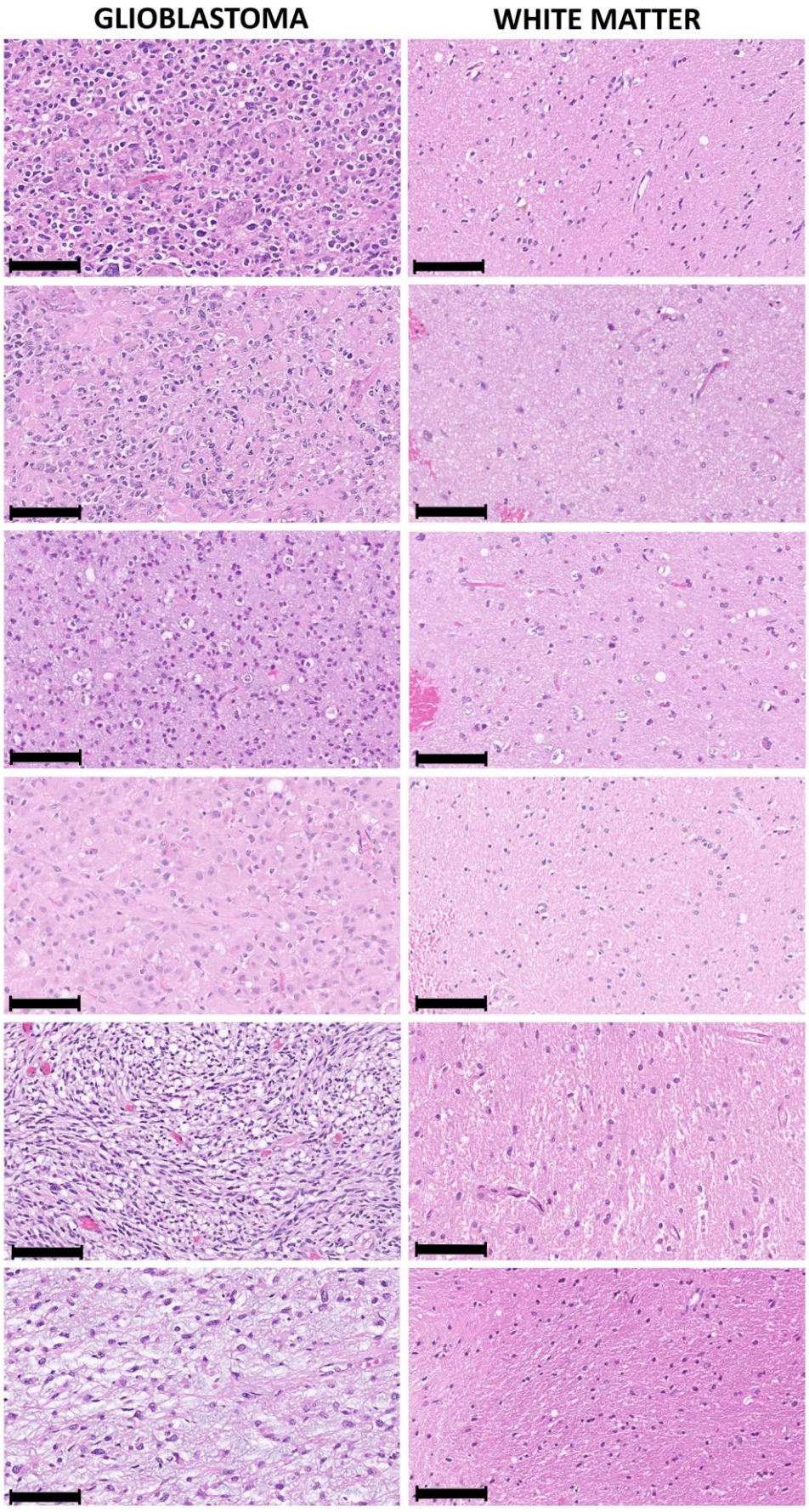
[157] V. Zohdi, D. R. Whelan, B. R. Wood, J. T. Pearson, K. R. Bambery, and M. J. Black, "Importance of tissue preparation methods in FTIR micro-spectroscopical analysis of biological tissues: 'traps for new users'," (in eng), *Plos One,* vol. 10, no. 2, pp. e0116491-e0116491, 2015, doi: 10.1371/journal.pone.0116491.

Appendix A: Histomorphologic Demonstration of GB and WM Tissue sections used in this study.

Scalebars = 100µm



Appendix A: Histomorphologic Demonstration of GB and WM Tissue sections used in this study.



Appendix A: Histomorphologic Demonstration of GB and WM Tissue sections used in this study.