T.C.
DOKUZ EYLUL UNIVERSITY
IZMIR INTERNATIONAL
BIOMEDICINE AND GENOME
INSTITUTE

DEVELOPING NOVEL TARGETED THERAPIES TOWARDS HIGH GRADE GLIOMA (HGG) BY USING OMICS DATA INTEGRATION APPROACHES

FADİME ÖZTOPRAK

MOLECULAR BIOLOGY AND GENETICS
PhD THESIS

İZMİR-2023

TEZ KODU: DEU.IBG.PhD-2016850030

T.C.
DOKUZ EYLUL UNIVERSITY
IZMIR INTERNATIONAL
BIOMEDICINE AND GENOME
INSTITUTE

DEVELOPING NOVEL TARGETED THERAPIES TOWARDS HIGH GRADE GLIOMA (HGG) BY USING OMICS DATA INTEGRATION APPROACHES

MOLECULAR BIOLOGY AND GENETICS
PhD THESIS

FADİME ÖZTOPRAK
Supervisor: Asst. Prof. Yavuz OKTAY
Co-Supervisor: Assoc. Prof. Zerrin IŞIK

İZMİR-2023

TEZ KODU: DEU.IBG.PhD-2016850030

Dokuz Eylül University Izmir International Biomedicine and Genome Institute Department of Genomics and Molecular Biotechnology, Molecular Biology and Genetics PhD program student Fadime Oztoprak has successfully completed the PhD thesis entitled **"Developing novel targeted therapies towards high grade glioma (HGG) by using omics data integration approaches"** on date of ../../….

HEAD OF THE COMMITTEE

Asst. Prof. Yavuz OKTAY

MEMBER

Assoc. Prof. Gökhan KARAKÜLAH

MEMBER

Asst. Prof. Cihangir YANDIM

MEMBER

Assoc. Prof. Gizem ÇALIBAŞI KOÇAL

MEMBER

Assoc. Prof. Evren KOBAN BAŞTANLAR

TABLE OF CONTENTS

ACKNOWLEDGEMENT **v**

ABSTRACT **vii**

ÖZET **ix**

LIST OF TABLES **xiii**

LIST OF FIGURES **xv**

ABBREVIATIONS **xix**

1. INTRODUCTION **1**

2. LITERATURE REVIEW **6**

2.1. Overview of Gliomas 6

2.2. Overview of LGG and GBM 6

2.3. Incidence Rate and Survival for LGG and GBM 7

2.4. Clinical Presentation of LGG and GBM 8

2.5. Intrinsic Risk Factors for Gliomas and GBM 9

2.6. Subtypes of LGG or GBM Based on Genetic Mutations 10

2.6.1. IDH Wild Type and IDH Mutant Gliomas or GBM 11

2.6.2. 1p/19q codeletion 12

2.6.3. TERT Promoter Mutation 12

2.6.4. EGFR Amplification 13

2.6.5. PTEN Mutation 13

2.7. Transcriptomic Profiling for LGG 14

2.8. DNA Methylation Profiling for LGG 15

2.9. Multi-Omics Data Integration for Biological Insights 16

3. MATERIALS AND METHODS **19**

3.1. Transcriptomics Data Collection and Preprocessing 19

3.2. Construction of the Weighted Gene Co-expression Network 20

3.3. Identification of Clinically Significant Modules 20

3.4. Functional Enrichment Analysis 21

3.5. Differential Gene Expression Analysis 21

3.6. Drug-gene Interactions 22

3.7. Data Collection and Preprocessing for DNA Methylation 22

3.8. Differential Gene Expression Analysis with GTex data 24

3.9. Differentially Methylated Region Analysis 24

3.10. Integration of DEGs and DMGs 25
3.11. Network Analysis 25
3.12. Perturbation Profiling 26
4. RESULTS 27
4.1. Pre-processing of TCGA-LGG RNASeq Dataset and Construction of Weighted Gene Co-expression Networks 27
4.2. Identification of Clinically Significant Modules 29
4.3. Enrichment Analysis for Modules of Interest 33
4.4. Differentially Expressed Genes (DEGs) 38
4.5. Drug-Gene Interactions 40
4.6. Data Collection and Preprocessing for DEG Analysis and DNA Methylation Data 41
4.7. Differential Gene Expression Analysis with GTex Data to Be Integrated with DNA Methylation Data 42
4.8. Differentially Methylated Region Analysis (CpG sites) 43
4.9. IDH mut pTERTmt - Quality Control 44
4.10. IDH mut pTERTwt - Quality Control 46
4.11. IDHwt pTERTmt - Quality Control 47
4.12. IDHwt pTERTwt - Quality Control 49
4.13. Integration of Methylation and Expression data 51
4.14 Network Analysis and EpiFactors 55
4.15. Perturbation Profiling 58
5. DISCUSSION 61
5.1. Subtype Specific Co-expression Modules and Their Enrichment Analysis 62
5.2. DEGS from Module of Interest 64
5.3. Interaction of Identified Gene Signatures with Known Drugs 66
5.4. Methylation Data Integration with Transcriptomics Data 67
5.5. Network Construction Based on MDGs + Modules of Interest and Essential Epifactor Analysis 69
6. CONCLUSION 73
7. REFERENCES 75
CURRICULUM VITAE 102

LIST OF TABLES

Table 4.1 : KEGG pathway enrichment of the clinically significant modules 33

Table 4.2 : Common pathways between the IDH mutant (IDH mut) greenyellow module and IDH wild-type (IDH wt) blue module 35

Table 4.3 :IDH wild-type blue module pathway-enriched DEGs 39

Table 4.4 : Drug-gene interaction table for IDH wild-type blue module pathway-enriched DEGs 40

Table 4.5 : The number of differentially expressed genes in different LGG subtypes 43

Table 4.6 :Differentially Methylated CpG sites (DMC) results 51

Table 4.7 : Methylation Driven Genes (MDGs) summary table 52

Table 4.8 : Enrichment analysis of Venn Diagram genes 53

Table 4.9 : Network statistics on epigenetic data with MDG nodes, edges, neighbors and protein complexes. 55

Table 4.10 : CHRONOS dependency scores of 66 diffuse glioma models 59

LIST OF FIGURES

Figure 3.1 : Module construction and clinically significant module detection by WGCNA 20

Figure 4.1a :Sample dendrogram of IDH mutant samples to detect outliers 27

Figure 4.1b :Sample dendrogram of IDHwt samples to detect outliers 28

Figure 4.2a : Hierarchical clustering of genes with module colors for IDH mut samples. 29

Figure 4.2b : Hierarchical clustering of genes with module colors for IDH wt samples 29

Figure 4.3a : Module-Trait relationship plot for LGG IDH mut samples 31

Figure 4.3b : Module-Trait relationship plot for LGG IDH wt samples 32

Figure 4.4a : Selected common KEGG pathway schemes of the IDH wt pTERT- blue module and IDH mut pTERT+ greenyellow module ~ Glutamatergic synapse 37

Figure 4.4b : Selected common KEGG pathway schemes of the IDH wt pTERT- blue module and IDH mut pTERT+ greenyellow module ~ GABAergic synapse 37

Figure 4.4c : Selected common KEGG pathway schemes of the IDH wt pTERT- blue module and IDH mut pTERT+ greenyellow module ~ Retrograde endocannabinoid signaling 38

Figure 4.5 :Sample clustering of normal samples 42

Figure 4.6a : Sample dendrogram for 356743 probes in IDH mt

pTERT mt group 44
Figure 4.6b :Multidimensional Scaling plot for 1000 most variable positions

~ IDH mt pTERT mt 45
Figure 4.6c :Heatmap for Top 1000 variable CpGs ~ IDH mt pTERT mt 45
Figure 4.7a :Sample dendrogram for 356743 probes in IDH mt pTERT wt 46
Figure 4.7b :Multidimensional Scaling plot for 1000 most variable positions

~ IDH mt pTERT wt 46
Figure 4.7c :Heatmap for Top 1000 variable CpGs ~ IDH mt pTERT wt 47
Figure 4.8a :Sample dendrogram for 356743 probes in IDH wt pTERT mt 47

Figure 4.8b :Multidimensional Scaling plot for 1000 most variable positions

~ IDH wt pTERT mt 48
Figure 4.8c :Heatmap for Top 1000 variable CpGs ~ IDH wt pTERT mt 48
Figure 4.9a :MDS plot of IDH wt pTERTwt samples 49
Figure 4.9b :Sample dendrogram for 356743 probes in IDH wt pTERT wt 49

Figure 4.9c :Multidimensional Scaling plot for 1000 most variable positions

~ IDH wt pTERT wt 50

Figure 4.9d :Heatmap for Top 1000 variable CpGs ~ IDH wt pTERT wt 50

Figure 4.10 :Venn diagram of methylation driven genes 52

Figure 4.11a : IDH mt pTERT mt - Protein complex focused subnetwork 56

Figure 4.11b : IDH mt pTERT wt - Protein complex focused subnetwork 57

Figure 4.11c : IDH wt pTERT mt - Protein complex focused subnetwork 57

Figure 4.11d : IDH wt pTERT wt - Protein complex focused subnetwork 58

ABBREVIATIONS

AC	Adenylate Cyclase
AMPA	Alpha-amino-3-hydroxy-5-methylisoxazole4-propionic acid
AASTR	Anaplastic Astrocytoma
ASTR	Astrocytoma
β	Beta
BMIQ	Beta Mixture Quantile dilation
CCLE	Cancer Cell Line Encyclopedia
CGGA	Chinese Glioma Genome Atlas
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAVID	Database for Annotation, Visualization, and Integrated Discovery
DEA	Differential gene expression analysis
DEGs	Differentially Expressed Genes
DMC	Differentially Methylated CpGs
DMGs	Differentially Methylated Genes
DIPGs	Diffuse Intrinsic Pontine Gliomas
DGB	Drug Gene Budger
DGI	Drug-Gene Interaction
E	Eigengene
FC	Fold Change
GABRA3	Gamma-amino butyric acid receptor alpha subunit 3
GEO	Gene Expression Omnibus
GDC	Genomic Data Commons
GTEx	Genotype-Tissue Expression
GBM	Glioblastoma Multiforme
GRA1	Glutamate ionotropic receptor AMPA type 1
HUGO	Human Genome Organisation
ICGC	International Cancer Genomics Consortium
IDH	Isocitrate Dehydrogenase
KEGG	Kyoto Encyclopedia of Genes and Genomes
LGG	Lower-Grade Glioma
MDGs	Methylation Driven Genes
MDS	Multidimensional Scaling
METABRIC	Molecular Taxonomy of Breast Cancer International Consortium

MDS	Multidimensional scaling
mt *or* mut	Mutant
NCBI	National Center for Biotechnology Information
NSCs	Neural Stem Cells
NGS	Next Generation Sequencing
ODG	Oligodendroglioma
OmicsDI	Omics Discovery Index
PIP2	phosphatidylinositol-4,5-bisphosphate
PIP3	phosphatidylinositol-3,4,5-trisphosphate
PKA	Protein Kinase A
QC	Quality Control
RNA	Ribonucleic Acid
RPKM	Reads Per Kilobase per Million
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
STN	Solid Tissue Normal
SE	Summarized Experiment
TCGA	The Cancer Genome Atlas
FDR	The False Discovery Rate
pTERT	TERT promoter
UUID	Universally Unique Identifier
WGCNA	Weighted Gene Co-expression Network Analysis
wt	Wild Type

ACKNOWLEDGEMENT

I didn't know that so many unpredictable events would happen during this time when I started my PhD program. I'm thrilled to have reached the finish line! I want to thank the amazing people who stood by me and helped me navigate this life-changing journey. Their support and guidance were instrumental in helping me earn my doctorate, even in the face of unexpected challenges.

I extend my foremost gratitude to my PhD supervisors, Asst. Prof. Yavuz Oktay and Assoc. Prof. Zerrin Işık, for their consistent support, guidance, and encouragement, which proved invaluable throughout the entire research process. Their unwavering presence and depth of knowledge have played a crucial role in forming my academic growth from the early phases of developing my research proposal to the final submission of my thesis.

Besides my supervisors, I would like to thank my thesis committee members Assoc. Prof. Gökhan Karakülah and Asst. Prof. Cihangir Yandım for their contributions and comments. They have challenged me with constructive criticisms maturing me and my research. My genuine thanks also goes to Dr. Evangelia Petsalaki for giving me the opportunity of fellowship in EMBL-EBI for a research experience in a dynamic environment.

My sincere thanks also go to Oktay Neurogenomics Lab members in IBG especially, Ece Sönmezler Adalı, Işık Esin Kıroğlu, Melike Tetik, Burcu Ekinci, Aykut Kuruoğlu, Güneş Şayan Can, Ebru Bakır, Tutku Yaraş and Dr. Semra Hız for the times of work and friendship. I also want to thank Paula Weidemüller and other members of the Petsalaki research group for their companionship in my journey in EBI.

I would like to thank Dr. Aadil Javed, Dr. Vanessa Segovia Aksu, Dr. Güneş Özhan, Dr. Hasan Buğra Çoban for their continued support and friendship during my PhD journey. I would also like to thank to my friends Melike Köse, Gamze Yurdunkulu, Çağrı Buldu, Necati Kaan Kutlu and Dr. Ayşegül Çifçi for their nourishing and refreshing companionship.

I would like to acknowledge the Scientific and Technological Research Institution of Turkey (TUBITAK - BIDEB 2211-C), Council of Higher Education (CoHE - YÖK 100/2000), and KYK doctoral scholarship programs for supporting me and my research financially. I would also like to acknowledge TUBITAK ULAKBIM, High Performance and Grid Computing Center (TRUBA resources) for supporting me

with computational power (computational analyses performed in this thesis were partially performed at TRUBA).

Last but not least I want to express my deepest thankfulness to my family members Doğan Öztoprak, Elife Öztoprak, Derya Öztoprak and Şeyma Öztoprak for believing in me and their never ending financial and emotional support during my long and arduous journey.

Fadime ÖZTOPRAK

ABSTRACT

Developing Novel Targeted Therapies Towards High Grade Glioma (HGG) by Using Omics Data Integration Approaches

Gliomas are the brain tumors that develop in glial cells and present severe challenges based on intertumoral heterogeneity associated with different subtypes, further leading to poor prognosis and outcomes for patients. This study was conducted to utilize the transcriptomics and DNA methylation datasets available to researchers to arrive at conclusions that can be utilized for either screening novel targets or use already established drugs that can target the specific gene signatures associated with low grade gliomas (LGGs), which develop into high grade gliomas (HGGs) or Glioblastoma multiforme (GBM). We identified co-expression modules and their associated pathways for specific subtypes of LGG IDH mut pTERT-, IDH mut pTERT+, IDH wt pTERT-, and IDH wt pTERT+. We constructed co-expression modules based on these subtypes and found common and different enriched pathways as synapse pathways and immune-related pathways, respectively. We further explored the differentially expressed genes (DEGs) and found a gene signature of upregulated *GNG12* and downregulated *PLCB1, GRIA2, GABRA3, and GNAL* after mapping DEGs on our co-expression modules of interest. This gene signature was included in our drug-gene interaction analysis, leading us to 4 drugs (Vemurafenib, Vanadium Pentoxide, Imatinib, and Cisplatin,) that can target 4 out of 5 genes. Therefore, we recommend exploring the synergistic effects of the combination of these drugs against low and high grade gliomas. We also integrated transcriptomics and DNA methylation data to develop networks including epigenetic factors that can be targeted in a subtype specific manner. Our analysis revealed that *PRMT5* can be used as a target irrespective of the LGG subtype and *WEE1* is a specific target for IDH wt regardless of pTERT status. The specific chemical inhibitors of these targets being available could facilitate translation of our findings into preclinical settings.

Keywords: Glioblastoma, Lower Grade Glioma, Transcriptomics, Co-expression Analysis, DNA Methylation, Bioinformatics Pipeline, Drug-Gene Interaction

ÖZET

Omics Veri Entegrasyon Yaklaşımlarını Kullanarak Yüksek Dereceli Gliomaya (HGG) Yönelik Yeni Hedefli Tedavilerin Geliştirilmesi

Gliomalar, glial hücrelerde gelişen beyin tümörleridir ve farklı alt tiplerle ilişkili intertümörel heterojenite tedaviyi zorlaştırıcı bir faktör olup, hastalar için kötü prognoza ve sonuçlara yol açar. Bu çalışma, araştırmacıların kullanımına açık olan transkriptomik ve DNA metilasyon veri setlerinden yararlanılarak yüksek dereceli gliomalar (HGG) / glioblastoma multiforme'ye (GBM) dönüşme potansiyeline sahip düşük dereceli gliomalar (LGG) ile ilişkili spesifik gen imzalarını hedef alabilen önceden belirlenmiş ilaçların kullanılmasını veya yeni terapi hedeflerinin belirlenmesi amaciyla gerçekleştirilmiştir. Bu amaçla, LGG IDH mut pTERT-, IDH mut pTERT+, IDH wt pTERT- ve IDH wt pTERT+'nin spesifik alt tiplerinde ortak ekspresyon modüllerini ve bunlarla ilişkili yolakarı belirledik. Bu alt tiplere dayanarak ortak gen ifade modülleri oluşturduk ve sırasıyla sinaps yolakları ve bağışıklıkla ilgili yolaklar gibi ortak ve farklı zenginleştirilmiş yolaklar bulduk. Diferansiyel olarak eksprese edilen genleri (DEG'ler) daha ileri düzeyde araştırdık ve DEG'leri ilgili ortak ekspresyon modüllerinde haritaladıktan sonra ifadesi artmış *GNG12* ve ifadesi azalmış *PLCB1, GRIA2, GABRA3 ve GNAL*'den oluşan bir gen imzası belirledik. Bu gen imzasının ilaç-gen etkileşimi analizimize dahil edilmesi bizi 5 genden 4'ünü hedef alabilen 4 ilaca (Vemurafenib, Vanadyum Pentoksit, Imatinib ve Cisplatin) yönlendirdi. Bu nedenle, bu ilaçların kombinasyonunun düşük ve yüksek dereceli gliomalara karşı sinerjistik etkilerinin araştırılmasını öneriyoruz. Ayrıca alt tipe özgü bir şekilde hedeflenebilecek epigenetik faktörleri içeren ağlar geliştirmek için transkriptomik ve DNA metilasyon verilerini entegre ettik. Analizimiz, *PRMT5'*in LGG alt tipinden bağımsız bir hedef olarak kullanılabileceğini ve *WEE1'*in, pTERT durumuna bakılmaksızın IDH wt için spesifik bir hedef olduğunu ortaya çıkardı. Bu hedeflerin spesifik kimyasal inhibitörlerinin mevcut olması bulgularımızı klinik öncesi çalışmalarda kullanım açısından faydalı kılmaktadır.

Anahtar Kelimeler: Glioblastoma, Düşük Dereceli Glioma, Transkriptomik, Ko-ekspresyon Analizi, DNA Metilasyonu, Biyoinformatik, İlaç-Gen Etkileşim

1. INTRODUCTION

Glioblastoma, also known as glioblastoma multiforme (GBM) is an extremely aggressive type of brain tumor initiating in the glial cells of the brain. These cells

inherently are involved in maintaining the structural stability of the neurons, which are the fundamental operational units of the nervous system (Hanif et al., 2017). The complex nature of Glioblastoma raises several difficulties in the selection of treatment modalities according to the attributes of the surroundings of the brain as well (Akter et al, 2021). The issues leading to such difficulties include the phenomena of immunosuppression, challenges within the drug development, the recurrence of the disease, location of tumor and its functional impact, the tumor microenvironment, limited treatment options, resistance to available therapeutic options, heterogeneity of the tumor, blood brain barrier, and invasive nature of the tumor. In order to deal with aforementioned hurdles, the expertise of a wide range are necessitated including oncologists, neurologists, researchers, and medical experts (Wu et al, 2021). For enhancing the treatment outcomes for the patients, several elements including inventive treatments such as immunotherapies and precision drugs along with progression in understanding the genetic foundations and molecular signatures of Glioblastoma are required (Ganipineni et al, 2018)..

Integrative omics analysis encompasses the understanding and merging of the data obtained from different advanced biological domains including metabolomics, proteomics, transcriptomics, and genomics among others. This methodology provides a unique and deep perspective on the profiles of biological systems towards specific scenarios which has the capacity to empower the research to further elucidate the intricate biological systems and significant understanding of entities involved in different diseases (Nice et al, 2018). The integrative omics analysis gives rise to the construction of prognostic models, annotation of genome functionalities, facilitation in exploration of novel drugs, tailored medical approaches, investigations in disease mechanisms, and identification of molecular biomarkers (Misra et al, 2019). The strategy has specially been utilized in helping clinicians and researchers in untangling the biological phenomena to further our understanding of the diseases and to incorporate that information in patient well being as the goal (Karczewski & Snyder, 2018).

The comprehensive study of Ribonucleic acid (RNA) molecules in tissues or cells is called transcriptomics, which has been considered indispensable for identifying the various subsets of different cancers based on expression of genes i.e. transcripts (Maniatis et al, 2021). The subsets obtained from gene expression data forming

clusters can lead to insights informing the diversity of individual cancers, which can further help in individualizing the therapeutic strategies efficiently (Valdes-Mora et al, 2018). The detailed applications of transcriptomics are exploration of gene expression profiling and its association with targeted therapies, functional enrichment analysis, differential gene expression analysis, unsupervised clustering and integration with other omics data for patient characterization (Supplitt et al, 2021). The sophisticated computational techniques along with the gene expression patterns obtained from transcriptomics analysis facilitate the researchers in diagnosis, treatment strategies, empowering comprehension, and investigate the molecular diversity of different cancers based on pinpointing the subsets within heterogeneous cancers such as Glioblastoma (Fan et al, 2020).

There are other types of omics data which have been used for exploring the biological systems informing the gene expression profiles further in different scenarios such as methylation status of DNA molecules. The methyl groups are attached to DNA molecules which inherently store the genetic information and constitute the chromosomes. This modification of DNA can govern changes in structure of chromatin and influences the accessibility of molecular machinery such as RNA polymerases to DNA molecules, preventing the cellular apparatus from initiating the process of transcription or gene expression (Bock, 2012; Ramsey et al., 2010). The epigenetic regulation of the genome can be delineated from the datasets related to DNA methylation along with the mechanistic details of genetic pathways, subtypes, biomarkers, and gene expression. Therefore, this modification of DNA which can be identified on genomic scale utilizing various platforms holds significant importance in unveiling gene activity and subsequently the molecular signatures arising from the activity or inactivity of various genes in states such as normal or diseased conditions (Horvath et al, 2013). The tailored strategies can be formulated on the plausible therapeutic targets obtained from the epigenomic data (Li et al, 2015).

The molecular changes occurring across different cancer types along with the characterization of the genetic data from cancer patients are the major objectives of The Cancer Genome Atlas (TCGA), which is the pioneering project undertaken to study cancer genomes (Tomczak et al, 2015). The extensive datasets from TCGA projects contain a plethora of data involving the whole cancer genomes, transcriptomes, epigenomes, and proteomes along with the information obtained from the clinical features of the patients. The clinical application and molecular

investigations on cancers at the molecular levels have been impacted profoundly upon utilizing these datasets as the cancer biology field has been revolutionized due to deep knowledge obtained from TCGA datasets (Wang et al, 2016). More recent applications of TCGA datasets involve the determination of survival related markers and application of deep learning methods to predict the drug response (Nicolle et al, 2019; Shen et al, 2023).

Since glioblastoma possesses a more significant threat as a disorder due to challenges mentioned previously along with the fact that aggressive forms of the disease have restricted treatment avenues, more tailored approaches toward molecularly identified subtypes of the disease are required to overcome these challenges. This study is designed to include integrated omics methodology to merge the transcriptomics and DNA methylation profiles of different clusters obtained from differentiated molecular subtypes of glioblastoma arising from TCGA datasets. The major aim of the study is to develop and implement an approach to identify the significant molecular pathways and the genes associated with the progression of low grade gliomas along with the potential plausible therapeutic agents. The gene expression patterns arising from transcriptomics data and its compilation with the DNA methylation data showcasing the epigenetic alterations in cancers, researchers can attain a multifaceted comprehension of the disease (Du et al, 2023), which is important in determining the treatment strategies tailored to molecular signatures of the cancer subtype.

The implementation of the sophisticated computational methods can help dissect the gene expression and DNA methylation trends in different genetic subtypes which can further be aligned with the severity of the disease (Binder et al, 2019). With the usage of unsupervised clustering techniques, molecular subcategories for cancers can be discerned, which can mirror the disparities found in the biological responses, signaling pathways and cellular mechanisms (Zhu et al, 2023). The subsequent gene list obtained within each clustered cohort in subtypes can undergo the functional enrichment analysis which can help in determination of the disrupted biological pathways fueling the progression of the disease (Timmons et al, 2015). These bioinformatics pipelines can facilitate the identification of direct pathways involved in specific conditions of the disorder. The integration of omics approaches therefore can lead to certain hits which can be used as potential targets in therapeutic interventions or biomarkers (Lund et al, 2017; Song et al, 2019). Therefore, the use of omics

technologies and their integration has the potential for customized or refined treatment strategies.

As glioma is a complex disorder with complex genetic subtypes, the transcriptomics and DNA methylation data should provide an in-depth view of the molecular signatures of the disease based on genetic profiles of the patients. Then the integration of these two types of omics data can facilitate the potent targets which can be utilized for tailored therapeutic strategies. Therefore, we employed the integration of these molecular datasets and implicated molecular patterns from low grade gliomas to not only recognize the genes that exhibit differential expression but also the significant epigenetic changes as well. We also executed a bioinformatics based search to identify several drugs that can target the genes expressed differentially in one of the genetic subtypes of gliomas.

The significance of this study emanates from the utilization of the TCGA data which is considered the gold standard among researchers in identifying the molecular patterns of various cancers. The specific outcomes of this study has the capacity to inform the targetable pathways associated with genetic subtypes of gliomas through comprehensive and integrated omics approach utilized in this study. The molecular characteristics and a narrow list of genes with the distinct subtypes can help unravel the diverse and complex molecular pathways involved in the progression of LGGs to HGGs, and the information can be used by both researchers and medical practitioners. Moreover, a detailed and integrated approach can not only amplify the effectiveness of the treatment but also can decrease the undesired impacts due to utilization of the molecular targets that contribute to proliferation of the disease (Paananen & Fortino, 2020). In summary, this study can facilitate the identification of driver genes, molecular and biological pathways along with the potential targets for glioma subtypes. We integrated the transcriptomics and DNA methylation of LGG TCGA dataset and external transcriptomics and methylation data of normal samples to arrive at conclusions dictating the molecular drivers of the disease. We also observed the drug-gene interactions focusing on the genes identified as up-regulated or down-regulated in one of the subtypes of lower grade glioma, further highlighting the importance of bioinformatic pipelines and their utilization in targeting pathways associated with different forms of glioma.

2. LITERATURE REVIEW

2.1. Overview of Gliomas

The brain tumors likely originating in the glial cells of the brain are termed as gliomas. These cells are responsible for providing the nourishment and support to the functional units of the nervous system i.e. neurons. Different types of gliomas depending on the types of cells the tumors originate from include ependymomas, oligodendrogliomas, and astrocytomas (Whitfield & Huse, 2022). The grading of the gliomas from Grade I to IV is based on the potential for growth of the tumor and its aggressiveness (Zhuge et al, 2020). The slow growing and least aggressive tumors of gliomas are considered Grade I, with Grade IV representing the most severe form of the disease (Zhuge et al, 2020).

The most common symptoms of the gliomas are numbness in the limbs, speech difficulties, changes in the vision capabilities, seizures, and headaches (Dono et al, 2020). The symptoms vary depending on the location of the tumors. Furthermore, the conventional treatment options available for gliomas include chemotherapy, radiation therapy, and surgery and are considered the first line of defense against the disease and depends on the overall health of the patient, location and grade of the tumor along with other factors (Schaff & Mellinghoff, 2023). The advanced forms of therapies for gliomas also exist and include stem cell therapy, molecular targeted therapy, immunotherapy, oncolytic virus therapy, and electric field therapy (Xiong & Wang, 2019).

2.2. Overview of LGG and GBM

Since LGGs can turn into GBM, we integrated the relevant literature to both LGGs and GBM in our study. As mentioned previously glioma is a term which acts as an umbrella for describing multiple types of brain tumors and usually Grade IV gliomas with the most aggressive presentation are termed Glioblastoma. Such types of tumors are also known as Glioblastoma Multiforme (GBM), due to their extreme heterogeneity. Lower Grade Gliomas (LGGs) correlate with younger age, while GBM typically occurs in adults and accounts for more than 60 % of the brain tumors presented in tumors of the brain in all adults (Grech et al, 2020). The higher incidence rates along with the resistance to treatment modalities and significantly higher

recurrence rates make GBM one of the deadliest cancers. The astrocyte Grade II and Grade III tumors, Grade IV Glioblastomas, and Grade II and Grade III oligodendrogliomas comprise the adult diffuse Gliomas according to the World Health Organization (WHO) (Louis et al, 2016).

Glioblastoma is rarely metastasized and is known to invade the nearby areas of the brain, but remains the most deadly and invasive glioma (Paolillo et al, 2018). The diagnostic reproducibility is limited for glioblastoma and the histologic diagnosis is also prone to changes depending on the clinicians (Kan et al, 2020). The *de novo* or primary GBM is the most common and aggressive form of the disease with secondary GBM less prevalent and initiating at lower-grade as astrocytoma but has been known to become Grade IV tumor as well (Urbańska et al, 2014; Nguyen et al, 2021). The pathogenic features of glioblastoma are diverse. The current failures in treatment are also associated with immunosuppression. Therefore, the long term survival of the patients is hard to achieve (Shergalis et al, 2018).

2.3. Incidence Rate and Survival for LGG and GBM

The neuroepithelial tumors of CNS are gliomas with varying behaviors ranging from LGG as tumors that are more or less resectable to the aggressive phenotypes including diffuse intrinsic pontine gliomas (DIPGs) and GBMs having the worse prognosis of all (Farmanfarma et al, 2019). LGGs likely occur in early childhood and high grade gliomas tend to occur at later stages. Normally LGG are graded as Grade I and Grade II with different types including mixed, diffuse, pleomorphic, giant cell, pilomyxoid, pilocytic forms along with desmoplastic infantile ganglioglioma, and oligodendrogliomas (McKhann & Duffau, 2019). The pediatric LGG annual incidence rate is calculated as 1.3-2.1 per 100,000 in the United States (Miguel Llordes et al, 2023). The adult LGG is considered more common as compared to pediatric LGG with cases rising up to 9.1-12.5 per 100,000 in the United States (Lin et al et al, 2021). The cases of adult LGG occur from 2^{nd} to 4^{th} decade of life as most cases are diagnosed during this period (Diwanji et al, 2017).

Among malignant brain and central nervous system tumor histopathologies, glioblastoma comprise 14.2 percent of all tumors and 50.1 percent of all the malignant forms of the tumors in the United States (Ostrom et al, 2022). It is also more prevalent in males as compared to females with males affected at 1.6 fold more than females

(Chen et al, 2021). The White ethnic patients are at 2-fold higher risk for glioblastoma compared to the Black patients (Ruhban et al, 2019). The latest age-adjusted incidence rate for Glioblastoma was calculated as 3.22 per 100,000 population in the United States (Ostrom et al, 2019). The 5-year survival rate for Glioblastoma is 4-5% (Batash et al, 2017). The largest cumulative study calculating the survival rate for Glioblastoma over a period of 10 years concluded that the ten year survival rate of Glioblastoma is 0.71% (Tykocki & Eltayeb, 2018). This data indicates that glioblastoma is a lethal disorder and treatment strategies need improvement to increase the survival rate of the patients (Janjua et al, 2021).

2.4. Clinical Presentation of LGG and GBM

The histology and the location of the lesion determine the presentation of LGG clinically in children and adults (Byrne et al, 2017). High grade lesions show progressive symptoms rapidly and low grade lesions present with insidious onset spanning over several months. The long history of seizures is also associated with low grade lesions which are known to be refractory to the antiepileptic treatments (Piotrowski & Blakeley, 2015). The malignant gliomas present with more than 30 % cases having seizures with low grade lesions having more seizures (Samudra et al, 2019). Other signs associated with LGG are aphasia or hemisensory deficits and hemiparesis as focal neurologic deficits along with mental status changes, balance changes, and headaches as higher intracranial pressure (Forst et al, 2014).

The course of the disease of glioblastoma is determined upon the factors such as tissue destruction and the extent of edema along with epilepsy and tumor location. These are the reasons why the clinical presentation for glioblastoma is not typical. Standard treatment options can still help in preserving the cognitive functioning and quality of life for the patients despite the fatal prognosis for glioblastoma. However, the decrease in cognition and decline in quality of life occur rapidly once the treatments start failing and lead to severe outcomes (Palmer et al, 2021; McKinnon et al, 2021; Bruhn et al, 2022).

Headache is the the most common symptom with presentation in one-third of the patients which results from the intracranial pressure occuring at night or patient awakening with dullness (Palmieri et al, 2021). At the time of diagnosis, heightened intracranial pressure can further cause slowing of neurocognition, vomiting, nausea,

fatigue, and dizziness. Using steroids such as dexamethasone can lead to amelioration of the symptoms caused by increased intracranial pressure (Grant, 2019).

Glioblastoma can also originate from the brain stem and lead to pediatric glioblastoma with patients typically presenting with occlusive hydrocephalus, dysphagia, and cranial nerve palsies or combination of these symptoms (Alther et al, 2020). Glioblastoma is associated with tropism in the brain, however, cases with liver, bone, lymph nodes, pleura, and lung metastasis are also found (Achi et al, 2023). Therefore, the follow up procedures required to obtain information from other parts of the body are also not required in patients presented with diagnosis of glioblastoma. Moreover, the glioblastoma patients can be donors for organ transplantation as cancer transmission risk is minimal in such scenarios (Warrens et al, 2012; Zhu et al, 2020).

2.5. Intrinsic Risk Factors for Gliomas and GBM

Gliomas can develop without inheritance within a family but there are familial gliomas as well that follow the Mendelian pattern (Louis et al, 2016). Malignant gliomas have been explored using genome wide studies to explore the genetic risks for the development of brain tumors (Ostrom et al, 2019). Such studies indicated that there are 25 susceptibility-asscoiated genomic loci across the incidence variance for glioma in adults encompassing approximately 30 % of the factors, with 70 % remaining unknown (Kinnersley et al, 2015; Ostrom et al, 2019). Somatic mutations plat a major role in the tumorigenesis of glioma (Howell et al, 2018), as explained in the following sections.

Apart from the age, other intrinsic risk factors associated with a small proportion of gliomas include hereditary syndromes of cancer including Turcot syndrome, neurofibromatosis I and II types along with Li-Fraumeni syndrome (Mutiarayani, 2023). Brain tumors including pilocytic astrocytoma and brainstem astrocytoma are mainly associated with *NF1* gene as autosomal dominant syndromes or Neurofibromatosis type 1 (NF1) (D'Angelo et al, 2019). NF2 syndrome is associated with either somatic or germline mutations in *NF2* gene which is a tumor suppressor gene. These mutations give rise to cranial meningiomas and cranial nerve schwannomas which are the tumors of the peripheral and central nervous system (CNS) (Tabor et al, 2023). The mutations in *TP53* tumor suppressor gene lead to Li-Fraumeni syndrome (Rocca et al, 2022) .In Li-Fraumeni syndrome, the germline

mutations in *TP53* have been associated with neuroepithelial tumors of the CNS, which have been found to be more common in females as compared to males (Xiong et al, 2020).

The Turcot syndrome is associated with mutations in *APC* gene which acts in repairing DNA (Khattab & Monga, 2022). The mutations in *NF1* and *NF2* are linked with neurofibromatosis I and II types (Aldape et al, 2015). The cases of more than one patient of glioblastoma in one family are rare, which makes the identification of risk loci difficult in the relatives of the patients (Molinaro et al, 2019).

The risk alleles for gliomagenesis have been observed from population based genome wide association studies including *RTEL1*, *TERT*, *AF*, *EGFR*, and *TP53* (Wrensch et al, 2019). Both *RTEL1* and *TERT* function in the telomere maintenance and risk alleles are specific for histological classification and diagnosis at older age (Walsh et al, 2013). The genome wide linkage investigations also yielded no significant high penetrance variants that can be linked with the risk of glioblastoma (Walsh et al, 2013). However, for gliomas, recently researchers identified *DMBT1*, *ZCH7B3*, and *HP1BP3* as non-coding variants after surveillance of the genomic landscape using CRISPR knockdown approach (Choi et al, 2023).

2.6. Subtypes of LGG or GBM Based on Genetic Mutations

The mutations in oncogenes and tumor suppressors are known to drive the process of carcinogenesis and in the case of gliomas and glioblastoma, there are several genes that have been linked with the disease-status based on their mutation status and can be used as markers for classification of the disease (Whitfield & Huse, 2022).

2.6.1. IDH Wild Type and IDH Mutant Gliomas or GBM

The discovery of mutations in the isocitrate dehydrogenase 1 (*IDH1)* gene marked a significant milestone in the understanding of the molecular pathology underlying glioblastoma and gliomas in general. Initially, a particular point mutation in the *IDH1* gene, R132H, in samples of glioblastoma patients was found (Parsons et al, 2008). Later, the mutations in *IDH1* and *IDH2* genes were associated with higher survival rates in glioblastoma patients as compared to patients with wild-type *IDH1 or*

IDH2 genes (Yu et al, 2010). It was further understood that the features including prognosis and clinical characteristics also differ for patients with *IDH* mutations as compared to patients with wild-type *IDH* genes (Mondesir et al, 2016).

IDH1 enzyme is found in the cytoplasm, while IDH2 and IDH3 enzymes reside in the matrix of mitochondria. These are the critical enzymes responsible for oxidative stress resistance and citric acid cycle (Bergaggio & Piva, 2019). The most commonly found mutation of *IDH1* in gliomas is R132H (Parsons et al, 2008). The current WHO classification of glioblastoma includes IDH wildtype (wt) and IDH mutant (mut), with IDH wt tumors displaying necrosis, microvascular proliferation, mitotic activity, diffuse growth pattern, cellular polymorphism, nuclear atypia, differentiation, and more prominently poor survival (Dono et al, 2020; Motomura et al, 2023).

The development of most low grade glioma (LGG) tumors are associated with the *IDH1* or *IDH2* mutations, as these are considered the earliest events in this transformation (Dono et al, 2021). Like in glioblastoma, IDH mutation leads to overall longer survival in LGG patients, while at the same time being associated with the malignant transformation of the tumors (Leu et al, 2016). IDH wt LGGs are considered heterogeneous with varying outcomes reported in clinical analyses. Furthermore, the inclusion of EGFR amplification, H3F3A mutation, and TERT promoter mutation in this subtype of LGG leads to worse prognosis (Vuong et al, 2019).

The three variants of IDH wt glioblastoma are epithelial-like GBM, gliosarcoma, and giant cell GBM (Louis et al, 2016). The epithelial-like GBM manifests as mesencephalic masses with *BRAF* V600E mutation and is more common in young people and children (Broniscer et al, 2014). Furthermore, this subtype lacks *PTEN* loss and *EGFR* amplification, as well (Alexandrescu et al, 2016). The *CDKN2A* deletion is found in Gliosarcoma (IDH wt) with less common *EGFR* amplification and *TP53* mutations (Lowder et al, 2019). Moreover, giant cell GBMs are also reported to lack the *CDKN2A* deletion and *EGFR* amplification, but entail *TP53* and *PTEN* mutations (Ogawa et al, 2020). IDH mut glioblastoma exhibit susceptibility to treatment with temozolomide as compared to IDH wt form (SongTao et al, 2012).

2.6.2. 1p/19q co-deletion

The concurrent deletion of both the short arm of chromosome 1 (1p) and the long arm of chromosome 19 (19q), known as 1p/19q co-deletion, represents an early

genetic event, along with *IDH1/2* mutations, contributing to the initiation of glioma and influencing overall survival outcomes. (Jenkins et al., 2006). Almost all LGGs that have 1p/19 codeletion also have *IDH1* or *IDH2* mutations (Labussiere et al., 2010). Many studies have shown that the LGGs with both 1p/19 codeletion and IDH mutations, has best prognosis (Kujas et al., 2005; Jenkins et al., 2006; Tews et al., 2006; Iwamoto et al., 2008), but the underlying mechanism is still unclear.

Tews et al. (2006) demonstrated that the genes located on deleted arms (1p/19) were differentially expressed between codel and non-codel samples, giving rise to the thought that the improved prognosis of patients with 1p/19q co-deleted gliomas may be attributed to at least some of these genes.

2.6.3. TERT Promoter Mutations

TERT encodes a telomerase enzyme responsible for inhibiting the shortening of the ends of chromosomes, thereby regulating the cellular aging process. It is also found as dysregulated in different cancers (Colebatch et al, 2019). The presence of either of two mutations (C228T or C250T) in the promoter of *TERT* gene (TERTp) have been identified as a marker for worse prognosis for glioblastoma (Powter et al, 2021). The TERTp mutation status is associated with elderly patients presenting Grade II/III glioma with 40 % frequency, further implicating that the correlation of TERTp mutation with pathology of glioma can be utilized not only as a prognostic feature but can be targeted for therapy as well (Fujimoto et al, 2021; Aquilanti et al, 2023).

In LGG IDH wt tumors with pTERT mutation leads to clustering of tumor samples with GBM based on DNA methylation profiling, implying that pTERT mutation is involved in increasing the invasiveness and severity of the disease (Fujimoto et al, 2021). Therefore, it is suggested that the clinical analysis of pTERT mutation in astrocytic gliomas (LGG) displaying features of glioblastoma are required for the diagnosis of diffuse astrocytic LGG.

2.6.4. EGFR Amplification

The *EGFR* mutation in gliomas usually leads to *EGFR* amplification with EGFRvIII variant being the most common (Garima et al, 2022). The genetic loci of *EGFR* is Chr7 (7p12) and encodes the receptor tyrosine kinase functioning at the

surface of the cells (Romano & Bucci, 2020). The EGFR degradation and internalization is impaired upon EGFRvIII variant activation which further causes the activation of anti-apoptotic signaling mitotic pathways associated with the tumorigenic capacity of GBM (Hoogstrate et al, 2022). Both EGFRvIII and *EGFR* amplification have been implicated in worsening the prognosis for young patients but *EGFR* overexpression is also specifically linked with poor survival in older patients (Chi et al, 2020). In LGG IDH wt (diffuse astrocytoma), EGFR mutations or amplification are also reported as clinically instructional, but the use of EGFR specific or pathway specific inhibitors have been utilized for glioblastoma with EGFR amplification (Yang et al, 2022).

2.6.5. PTEN Mutation

PTEN is a tumor suppressor gene and loss of its function has been associated with carcinogenesis. The PTEN protein is involved in catalyzing the removal of a phosphate group from the inositol ring present in PIP3 phospholipid. This reaction leads to the production of PIP2 and is critical for inhibiting the AKT signaling pathway (Yang et al, 2020). In quiescent and differentiated cells, the PI3K/AKT pathway is dormant, however its activation can lead to positive regulation of cell cycle which is required for carcinogenesis (Jiang et al, 2020). The *PTEN* loss disturbs the lipid phosphatase function via the PI3K/AKT pathway and leads to aggressive phenotypes which can be targeted for therapy as well (Choi et al, 2021). Moreover, PTEN deleted or mutant gliomas including LGG are known to have very aggressive phenotype, worse prognosis and therapeutic resistance (Zhang et al, 2021). Another study indicated that the PTEN mutant LGG showed highest malignancy comparable to GBM, which can be used as an independent prognostic factor for the disease especially in the case of IDH with PTEN mutant cases (Zhang et al, 2021).

2.7. Transcriptomics Profiling for LGG

Cancers are heterogenous disorders with varying degrees of diversity at genomic, transcriptomic, and proteomic levels. Therefore, in order to not only accurately characterize the pathogenesis of the disease, but also to understand the heterogeneity, transcriptome-wide analysis and profiling can be useful (Fan et al,

2020). The changes in gene expression levels monitored by utilizing the next-generation sequencing technologies including RNA sequencing can result in quantifying the high throughput parallels in multiple scenarios. Therefore, phenotypic alterations occurring at molecular level can be monitored in the perspective of systems biology (Cieślik & Chinnaiyan, 2018; Fan et al, 2020). According to transcriptomic profiling, *CRY2, HDAC1, DCLRE1B,* and *KPNA2* genes have been identified as a transcriptomic signature for LGG utilizing the Chinese Glioma Genome Atlas (CGGA) datasets. The same study identified the DNA repair and cell cycle related biological process as dysregulated highly in LGG (Zeng et al, 2019).

The IDH and TP53 mutant cases of LGG were profiled based on integration of genomics and transcriptomics data to identify the biological pathways and it was discovered that the metastases, invasion, immune function, and neuron function related functions were enriched in LGG patients based on selected genes which can be targeted to increase the possible survival outcomes of these patients (Liu et al, 2022). Another study utilized the transcriptomics analysis to separate the LGG patients into two groups of ferroptosis-related clusters (FRC), and concluded that ferroptosis was a major driving force behind forming the tumor microenvironment in LGG (Tu et al, 2022). Therefore, transcriptomics profiling can be utilized in identifying the unknown features associated with the tumor heterogeneity based on different parameters and in the field of LGG biology, has a distinct utilization which should be explored to identify the potential drug targets.

Multiple samples can be included to identify the gene expression patterns using the method of WGCNA, which can cluster the genes in the form of modules displaying the similar patterns of gene expression (Langfelder & Horvath, 2008). The relationship between the modules and specific traits associated with the samples can then be explored. Yang et al in 2018 identified the novel prognostic targets such as *NUSAP1* and *GPR65* that can be utilized as genetic therapeutic strategies against GBM. The study arrived at their conclusion by utilizing the WGCNA based module construction strategy using Differentially Expressed Genes (DEGs) from transcriptomics data from liquid biopsy samples of GBM (Yang et al, 2018).

2.8. DNA Methylation Profiling for LGG

The epigenetic changes have been identified as significant in identifying the cancers, therefore another element of genetic regulation can be used in investigating the common biomarkers (Guo et al, 2019). The genome region rich in CpG dinucleotides are usually linked with the promoters of the genes and DNA methylation of these nucleotides (CpG islands) found in the promoters dramatically silences the expression of the genes (Caiafa & Zampieri, 2005).

The changes in the levels of DNA methylation of promoter regions containing CpG islands can either be hypermethylated leading to gene silencing or hypomethylation causing overexpression of the genes. One of the epigenetic alterations observed in cancers is DNA methylation which functions in the regulation of genomic functions and plays an important role in tumorigenesis (Davalos & Esteller, 2023). It has the potential to be utilized in clinical practice for diagnosis and prognosis of various cancers (Liu et al, 2019). The methylation status of genes are associated with their expression levels as observed for various forms of gliomas (Dabrowski & Wojtas, 2019). For example, for lower grade glioma (LGG) and glioblastoma (higher grade), *MGMT* promoter methylation is considered a prognostic factor and is also associated with the survival rates of the patients suffering these diseases (Brandner et al, 2021; Haque et al, 2022).

Several studies have been conducted to identify the subtypes of gliomas or glioblastoma based on DNA methylation profiles of the patients. In 2013, Brennan et al identified 6 subtypes of glioblastoma based on DNA methylation profiles (Brennan et al, 2013). For diffuse adult LGG, Ferreyra et al (2021) studied 166 cases for genome-wide DNA methylation and associated methylation classes with IDH mutation status and with other genetic factors such as 1p/19q deletion. IDH wt gliomas could also be sub-categorized in the DNA methylation analysis for different molecular signatures (Ferreya et al, 2021). Recently, according to *IDH1* mutation status, three genes were identified based on methylation-driven expression: *STEAP3, CMYA5*, and *ARL9* for LGG (Guo et al, 2022). Therefore, DNA methylation profiling adds another layer of framework to the guidance in prognosing the diseases such as GBM and LGG.

2.9. Multi-Omics Data Integration for Biological Insights

The integration of multiple data forms originating at high throughput and large scale such as genomics, transcriptomics, DNA methylation profiles, and proteomics,

can be integrated in the form of biological pathways or networks which can provide an overview at systems biology level (Subramanian et al, 2020). These approaches have the capacity to combine the individual type of omics data simultaneously leading to comprehension of interplay between different molecules with information flow from one layer of information to another with the goal of bridging the genotype to phenotype gap inherently present in biology (Guo et al, 2021). The field of multi omics analysis is growing fast which has resulted in the development of various platforms, methods and tools for interpretation, visualization, and data analysis of multi omics data (Huang et al, 2017; Rappoport & Shamir, 2018; Wu et al, 2019).

The omics data can either be generated through biological experiments conducted in the laboratory or is readily available from different sources and databases such as The Cancer Genome Atlas (TCGA), International Cancer Genomics Consortium (ICGC), Cancer Cell Line Encyclopedia (CCLE), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC), or Omics Discovery Index (OmicsDI), and archives such as National Center for Biotechnology Information (NCBI) and Gene Expression Omnibus (GEO) (Subramanian et al, 2020). The major biological questions asked in multi omics analysis involve the classification or subtyping of diseases based on multi-omics profiling along with the prediction of driver genes responsible for triggering diseases which can be used as biomarkers as well, and to derive the biological insights behind the disease mechanisms (Li et al, 2021). The methods or tools utilized for multi-omics approaches are subdivided into categories such as multivariate methods, correlation-based analysis, similarity based methods, fusion, bayesian or network based approaches (Subramanian et al, 2020).

The network based integration approaches can provide the gene prioritization for cancers and their integrated framework. The gene aberrations along with alterations in epigenetics and further downstream players such as microRNAs and subsequent protein expression can be integrated in the network interactions (Li et al, 2021). The network-diffusion model utilizes the per-sample method for directing the network deriving the gene ranking at population level based on the aggregates of the individual rankings with the goal of outputting the global ranking for each sample (Subramanian et al, 2020). Another clustering method based on Neighborhood analysis for multi-omics data is dependent on the similarity based analysis and is constructed upon the already established methods. It constructs the interpatient similarity matrix distances for the individual omic datasets and then integrates the data from multiple omics into

a single matrix, which is then separated based on spectral clustering (Demirel et al, 2022). This method can be utilized for partial datasets as well, making it useful in scenarios where complete datasets are missing (Rappoport, N., & Shamir, 2019).

The underlying mechanisms associated with cancer can be deeply investigated with respect to drug targets or biomarkers that can be identified using pathway-based integration of omics data (Kim et al, 2019). The enrichment of pathways and over-representation can be derived from the biological entities such as metabolites, proteins, and genes to link with the already existing biological information in such methods (Tseng et al, 2015). The pipeline of such methods includes the utilization of DEGs or proteins along with their significance in the data sets based on statistical inference and then the over-representation analysis is performed based on certain thresholds to compose the background lists, depending on the number of proteins or transcripts in the reference proteome or transcriptome (Eichner et al, 2014). Any combination of genotypic, transcriptomic, DNA methylation, proteomics, or metabolomics data can be integrated to provide biological insights depending on the research question and appropriate parameters used in the integration or mapping of one dataset to another (Subramanian et al, 2020).

3. MATERIALS AND METHODS

3.1. Transcriptomics Data Collection and Preprocessing

We downloaded the expression data of the lower-grade glioma (LGG) cohort using the TCGABiolinks package (Colaprico et al, 2016). TCGABiolinks enables researchers to query, download, prepare and perform integrative analysis using TCGA data. We first queried NCI's Genomic Data Commons (GDC) portal for the TCGA-LGG project, gene expression quantification data type, Illumina HiSeq platform and legacy data using the "GDCquery" function. After creating our query, both gene expression and clinic data were downloaded using the "GDCdownload" function and prepared with "GDCprepare" function. The GDCprepare function creates a SummarizedExperiment (SE) (Huber et al, 2015) object for downstream analysis.

We divided our samples into two separate groups according to their IDH mutation status: IDH mutant and IDH wildtype (Figure 3.1). These expression matrices were then normalized and filtered with "TCGAanalyze_Normalization" (geneInfo) and "TCGAanalyze_Filtering" (quantile method, qnt.cut 0.25) functions respectively. We constructed sample clustering trees using the average method and removed the outliers. The number of samples in IDH mutant and IDH wild type subtypes of LGG along with the number of genes after normalization and filtering pipelines are shown in Figure 3.1. One outlier was excluded from IDH mutant and 5 outliers were excluded from IDH wild type samples. The selected modules with respective colors as differentiating a parameter are also given in Figure 3.1. The RNA-sequencing data exhibits non-normal distribution characteristics, warranting modeling approaches such as Poisson or negative binomial distributions for an accurate representation (de Torrenté et al, 2020). Prior to constructing weighted gene coexpression networks we normalized RNA-sequencing data with the voom methodology (Law et al, 2014) from the limma library in R (Ritchie et al, 2015) and investigated the gene quality. The voom function calculates the mean-variance of the log-counts and creates each observation's precision weight. This transformation enables us to perform comparative analysis with WGCNA which were developed for

microarray analyses. Employing voom transformation on RNA-seq data is a common practice in computational biology (Wang et al. 2017, Breen et al. 2019, Ota et al. 2021, Michlmayr et al. 2020, Singhania et al. 2018). To ensure robustness in our analysis, we employed the median absolute deviation (MAD) as a reliable metric of variability. This choice was motivated by the need to address situations where two genes exhibit minimal variations in expression levels among patients, leading to strong correlations in WGCNA analysis.

3.2. Construction of the Weighted Gene Co-expression Network (WGCNA)

Scale-free undirected co-expression networks were constructed using the WGCNA library in R (Langfelder & Horvath, 2008; Niu et al, 2023). Soft threshold power beta values were determined using the "pickSoftThreshold" function on the scale-free topology criterion. The suggested β is the lowest value where R2 is bigger than 0.8 (Hou et al, 2019). In accordance with this criterion, we selected beta values "6" and "12" for the IDH mut cohort and the IDH wt cohort, respectively. Setting the module size to 30 and the merge cut height to 0.25, we have constructed the gene modules with the correlation network methodology.

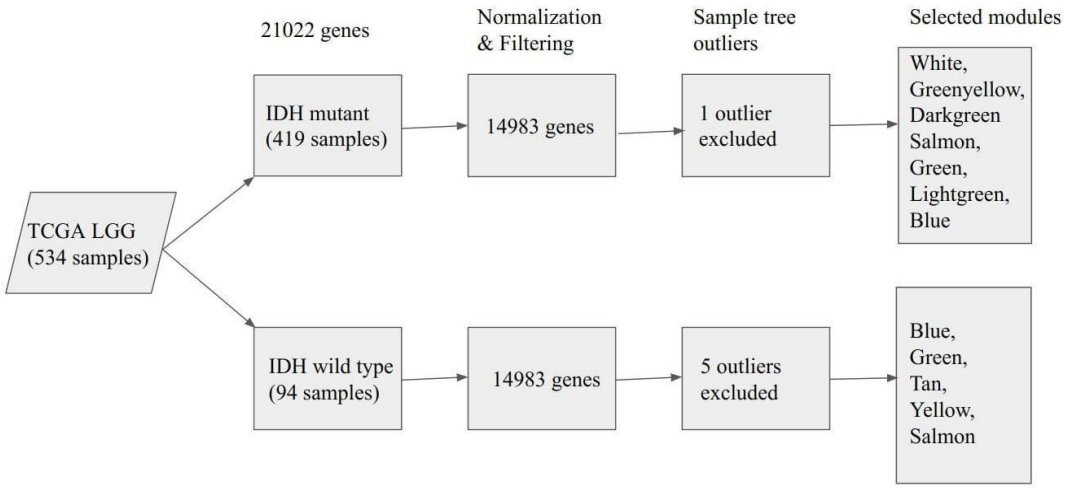

Figure 3.1. Module construction and clinically significant module detection by Weighted Gene Co-expression Network Analysis (WGCNA).

3.3. Identification of Clinically Significant Modules

We related the co-expression modules to clinical traits to identify clinically significant modules. This relationship is based on the eigengene network methodology. The module eigengene (*E*) is the first principal component of a given module and they are defined as the summary of each module (Langfelder & Horvath, 2008). These eigengenes are associated with the external traits (clinical information) to determine most significant associations. TCGA clinical information contains 110 features in the clinical data frame which is a collection of data related to patient diagnosis, demographics, exposures, laboratory tests, and family relationships. In this study we aimed to explore gene expression modules related to significant molecular alterations and we prioritized 8 features (of 110) that have prognostic value in clinical implications to identify clinically significant modules. These 8 features are: Chr7gain.Chr10Loss, Chr19_20.co.gain, TERT.promoter.status, TERTexpression.log2, TERTexpression.status, ATRX.status, Telomere.Maintenance and BRAF.V600E.status. We identified a module as significant if its correlation value with the TERT.promoter.status was above 0.5 or below -0.5.

3.4. Functional Enrichment Analysis

To better understand the biological mechanisms behind the modules of interest, module genes were extracted and converted to Entrez identifiers. Functional enrichment analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) platform to perform enrichment analysis (Sherman et al, 2022).

3.5. Differential Gene Expression Analysis

Differential gene expression analysis (DEA) is a widely used application to elucidate differentially expressed genes (DEGs) across two or more conditions. We analyzed differential expressed genes in LGG IDH wt samples and LGG IDH wt samples compared to solid tissue normal (STN). In addition we have compared two lower grade glioma subtypes with each other to detect differentially expressed genes between them. The TCGA-LGG cohort does not contain normal samples. We used the

Solid Tissue Normal (STN) samples from the TCGA-GBM cohort. Same preprocessing steps were applied as stated in section 3.1.

Differential gene expression analysis was performed using the "TCGAanalyze_DEA" function (TCGABiolinks package) with the "exactTest" method. The false discovery date (FDR) threshold was set at 0.01 and the absolute logarithmic fold change (logFC) was set to 1. We further screened the genes in the selected modules (IDH wt pTERT- blue and IDH mut pTERT- greenyellow) on matching DEGs lists (normal tissue vs LGG IDH wt, normal tissue vs LGG IDH mut and IDH mut vs. IDH wt). In addition, the genes enriched in common pathways between greenyellow and blue modules were inscribed. We focused on blue module genes as these were observed in distinct components of the common pathways,

3.6. Drug-gene Interactions

An association between a medicine and a genetic variation that may have an impact on a patient's response to the treatment. This phenomenon is known as a drug-gene interaction (DGI). We tested drugs that were altering the expression of previously selected genes using the Drug Gene Budger (DGB) tool (Wang et al, 2019). Using the CRowd Extracted Expression of Differential Signatures dataset (Wang et al, 2016), drugs that have inhibitory or activatory effects on up-regulated or down-regulated genes respectively were chosen in our investigation.

3.7. Data Collection and Preprocessing for DNA Methylation

The TCGA-LGG (Brain Lower Grade Glioma) DNA methylation data from Illumina Human Methylation 450 platform (legacy data) were queried, downloaded and prepared with "GDCquery", "GDCdownload" and "GDCprepare" functions of the TCGABiolinks package (Colaprico et al, 2016) respectively. The downloaded data was the beta values (β) of raw methylation data. Beta values represent the estimated levels of methylation, determined by the ratio of intensities between methylated and unmethylated alleles. These values range from 0 to 1, where 0 signifies no methylation (unmethylated) and 1 indicates complete (full) methylation. The TCGA-LGG methylation samples were divided into four groups according to their IDH and TERT

promoter mutation status. We created four different beta matrices for subgroups: IDH mut TERTp mt, IDH mut TERTp wt, IDH wt TERTp mt and IDH wt TERTp wt.

The DNA methylation data of the Glioblastoma Multiforme (TCGA-GBM) cohort were downloaded and 2 Solid Tissue Normal (STN) samples in this dataset were extracted. The probes with beta values missing for both of these normal samples were filtered. To increase our normal sample size we searched the literature for more normal samples of brain methylation data. Capper et al in 2018 reported a study including various central nervous system (CNS) tumors and a variety of control samples (Capper et al, 2018). The methylation data was obtained using Illumina Human Methylation 450 platform and was aligned to hg19, which is compatible with TCGA-LGG and TCGA-GBM methylation data. From the control samples of various CNS normal tissues we decided to use hemispheric cortex samples (CONTR, HEMI). We downloaded the beta matrix from the Gene Expression Omnibus (GEO) (GSE90496) and extracted 6 hemispheric cortex samples (samples 382 - 387).

We further filtered the probe names compatible with the probes coming from TCGA-GBM STN samples and merged them into a beta matrix. A sample clustering tree was constructed with the average method to observe sample similarities as they are coming from different studies. We used the ChAMP package (Morris et al, 2014) to impute missing values with the champ.impute function and normalized with the champ.norm functions using BMIQ method (Teschendorff et al, 2013). On Illumina bead arrays, there are two types of probe designs (type I and type II) with different hybridisation chemistries. This difference affects distributions of probes coming from different types. Type II probes show a reduced dynamic range and this can lead to a bias in the selection of type-I over type-II probes. To avoid this technical effect, normalization of probes against the type-II probe bias is recommended (Dedeurwaerder et al, 2011). After normalization, we visualized normalized data with QC.GUI.

Normal brain expression data is necessary to perform differential expression analysis (DEA). Unfortunately there were no normal samples in the TCGA-LGG cohort and there were only 5 normal samples in the TCGA-GBM cohort. To make our DEA more effective we decided to use the Genotype-Tissue Expression (GTEx) project data. The GTEx portal serves as an extensive public repository for exploring tissue-specific gene expression and regulation. It encompasses data from 54 non-diseased tissue sites, involving the analysis of nearly 1000 individuals. (Carithers &

Moore, 2015). We decided to use brain cortex expression data. The TCGABiolinks package makes users able to use TCGA data and GTex data for downstream analysis. We queried brain tissue data from both TCGA and GTex projects using the "TCGAquery_recount2" function and obtained Summarized Experiment (SE) objects separately.

We only selected the "Brain - Cortex" samples from GTex SE. As GTex data is available as counts and were aligned to hg38, we have queried, downloaded and prepared TCGA - LGG data in "STAR - Counts" workflow for consistency. The counts were scaled and the number of reads were checked. We then created subgroups of LGG data according to their IDH and TERT promoter mutation status. We gathered four subgroups: IDH mut TERTp mt, IDH mut TERTp wt, IDH wt TERTp mt, and IDH wt TERTp wt. Before starting the analysis we replaced Universally Unique Identifiers (UUIDs) with TCGA barcodes and subgroups were merged with normal samples separately using the rownames. All of these count matrices were normalized (by geneLength method) and filtered (by quantile method, qnt.cut=0.25) before performing differential expression analysis.

3.8. Differential Gene Expression Analysis with GTex data

Differential gene expression analysis was performed on previously created subgroups separately to obtain differentially expressed genes between normal samples and LGG samples with different molecular subtypes. To perform DEA we used the "TCGAanalyze_DEA" function with the limma pipeline and the glmLRT method (TCGABiolinks package). "Voom" argument was set to TRUE, false discovery rate cut (fdr.cut) was set to 0.01 and logarithmic fold change cut (logFC.cut) was set to 2. We removed the duplicated Ensenmbl gene identifiers and converted them to the HUGO symbols using the Biomart package (Drost & Paszkowski, 2017) and visualized the results using the "TCGAVisualize_volcano" function.

3.9. Differentially Methylated Regions Analysis

For each subgroup created previously we created Summarized Experiment (SE) objects as the input for the analysis. SE objects consist of features as rows

(rowRanges), samples in columns (colData) and values as matrix (assays). We performed differentially methylated regions analysis using the "TCGAanalyze_DMC" function. We set the differential mean methylation cut (diffmean.cut) to 0.30 and adjusted p-value cut (p.cut) to 0.01. To annotate the differentially methylated CpGs we used Illumina HumanMethylation450 BeadChip [UBC enhanced annotation v1.0] data frame (GPL16304). The first 22 rows of this data frame contain column explanations, so we have duplicated the file but cleaned the informative rows in the second document. In this data frame we were specifically interested in "Distance_closest_TSS" and "Closest_TSS_gene_name" columns. The closest TSS gene names were obtained from the UCSC knownGene table. There is an estimation that 4% of probes have the same distance from more than one TSS. In this case, the annotated gene name of the first gene which appears in the UCSC knownGene table was utilized. We filtered out probes that contained known SNPs and hit the XY chromosomes. For the purpose of this study we filtered probes at 50kB downstream or upstream from the closest gene. In this case for some genes there were multiple differentially methylated probes. To obtain genes based on differential methylation values we aggregated DMC results by median function of the dplyr package in R (Wickham et al, 2019). At the end we obtained differentially methylated genes (DMGs) for each LGG subgroup.

3.10. Integration of DEGs and DMGs

In order to obtain both differentially expressed and differentially methylated genes in each subgroup we integrated our results from previous sections. For a better visualization we filtered DEGs and DMGs by adjusted *p-value* (< 0.01). For DEGs, logFC threshold was set to 2 and for DMGs DNAmethylation difference threshold was set to 0.3.

3.11. Network Analysis

Gene markers identified by differential expression and differential methylation analysis, methylation driven genes (MDGs), were mapped onto STRING v10 [T800] (Szklarczyk et al, 2015) and on the Cytoscape (Kohl et al, 2011) to incorporate our findings with the known interactions. We then selected the first neighbors of the

identified MDGs for better understanding of disease mechanisms. Considering the objectives of our study, we further searched MDGs and their first neighbors on the EpiFactors (Medvedeva et al, 2015; Marakulina et al, 2023) database for the genes which are part of protein complexes playing a role as epigenetic factors. We reduced the network structures to epigenetic factor focused smaller networks as the biological networks are large, complex and contain a number of different molecular system signatures such as signal transduction, and gene regulation. We color coded the network as red (hypermethylated and downregulated MDGs), green (hypomethylated and upregulated MDGs) and blue (genes in the epigenetic factor protein complexes).

3.12. Perturbation Profiling

DepMap is a large-scale functional genomic profiling consortium to map the landscape of cancer vulnerabilities. The genetic perturbation platform specifically aims to screen mammalian cells and identify genetic alterations changing the phenotype, by perturbing genes with different approaches including CRISPR/Cas9 constructs. Cancer dependencies were modeled using analytical methods such as CERES (Meyers et al, 2017) and CHRONOS dependency score (Dempster et al, 2021).

After detecting the large module structures in each subtypes, we screened the gene effect on the DepMap portal using CRISPR (DepMap Public 22Q4+Score, Chronos) data and diffuse glioma models.

4. RESULTS

4.1. Pre-processing of TCGA-LGG RNASeq Dataset and Construction of Weighted Gene Co-expression Networks

The TCGA-LGG cohort expression matrix contained 21,022 genes and 513 samples. 419 of these samples were IDH mut and 94 of them were IDH wt. We preprocessed RNASeq data of TCGA-LGG subtypes separately. After filtering and normalization steps 14,893 genes were left in each group. We detected five outliers in the TCGA IDH mut group (Figure 4.1a) and one outlier in the IDH wt group (Figure 4.1b) and removed them. Sample dendrograms show transcriptionally different samples compared to other samples. The dendrogram signifies the division of complex data into meaningful and differentiated clusters of transcriptionally relevant cohorts.

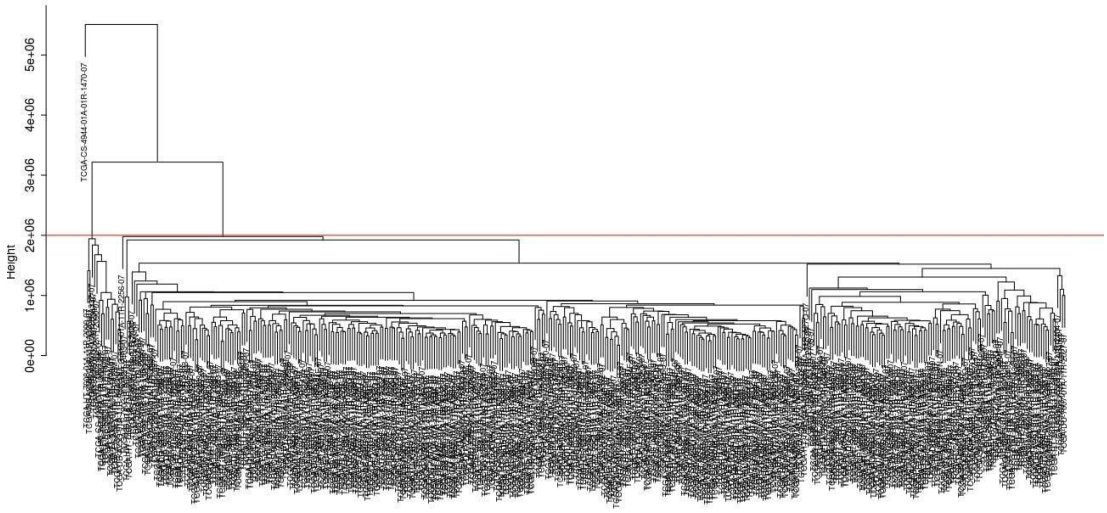

Figure 4.1a. Sample dendrogram of IDH mutant samples to detect outliers.

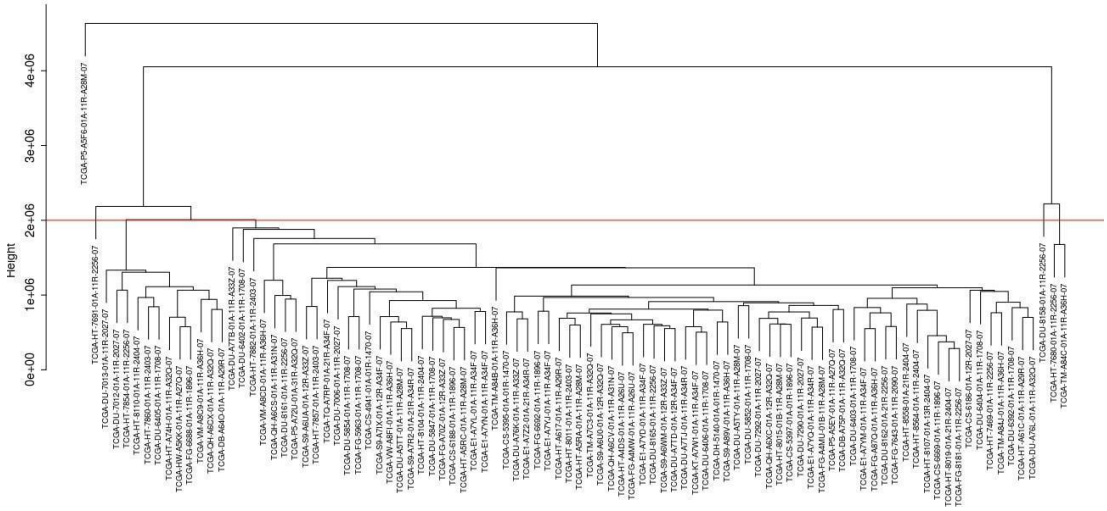

Figure 4.1b. Sample dendrogram of IDH wt samples to detect outliers.

Weighted gene expression networks were constructed with selected soft-thresholding powers (IDH mut subgroup: 6; IDH wt subgroup: 12). The modules were identified with the hierarchical clustering method and each module was color-labeled. Figure 4.2a-b are showing cluster dendograms of IDH mut and IDH wt samples respectively. 28 modules were constructed in the LGG IDH mut group and 14 modules were constructed in the LGG IDH wt group. Genes that have not been clustered in any module were collected in the "Grey" module (LGG IDH mut: 1055 genes, LGG IDH wt: 6314 genes). The clustered dendrograms show the genes as branches and the branch clusters are detected as modules as shown in the figure below. The branch height and module colors are highlighted in the cluster dendrogram. The visualization platform simplified the differentiation of the clusters and the processing of the complex raw data can be observed (Figure 4.2a-b).

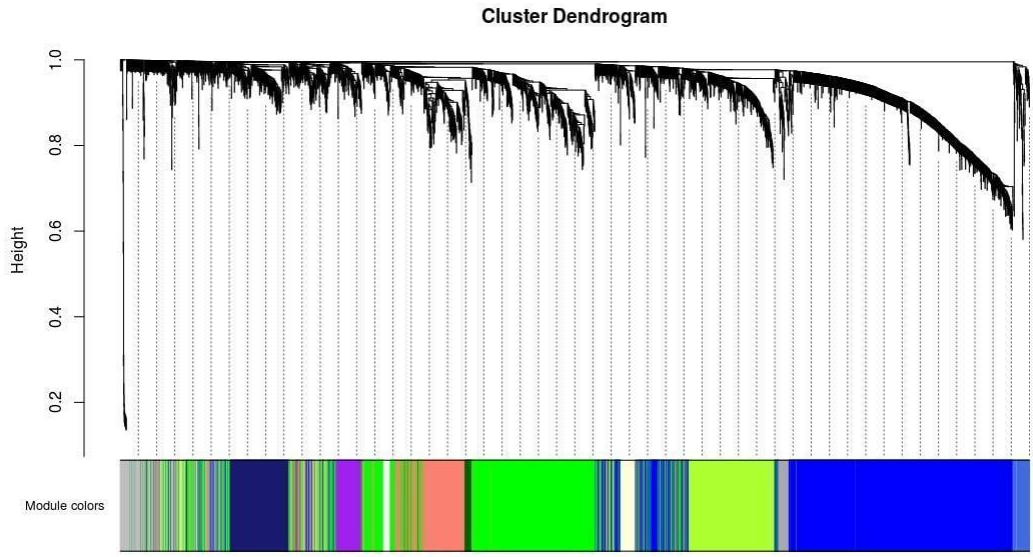

Figure 4.2a. Hierarchical clustering of genes with module colors for IDH mut samples.

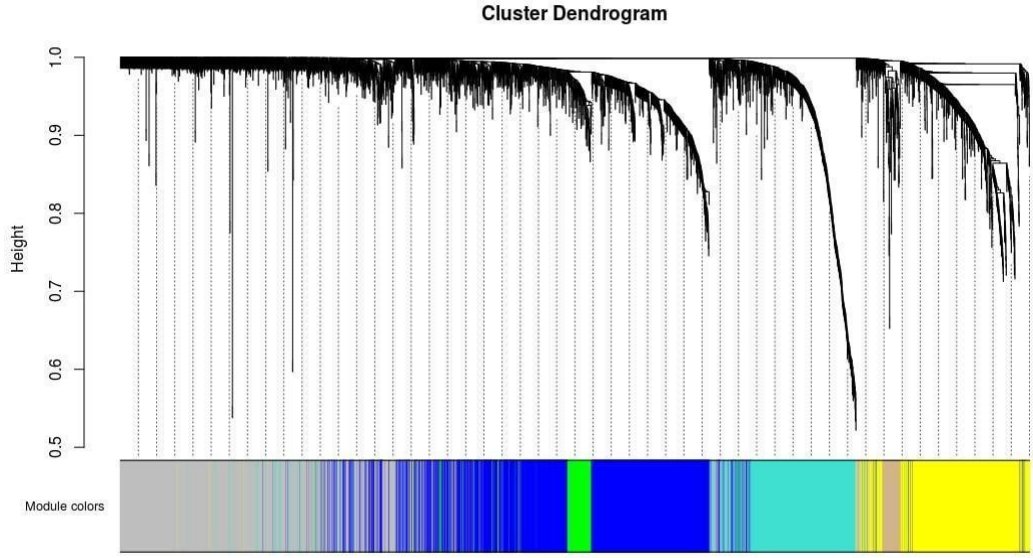

Figure 4.2b. Hierarchical clustering of genes with module colors for IDH wt samples.

4.2. Identification of Clinically Significant Modules

We further the observed module-trait relationships correlation to discover clinically significant modules. The threshold value was set to 0.5 for positive correlations and -0.5 for negative correlations. The IDH mutant group did not show correlation for Chr19_20.co.gain or BRAF.V600E.status features thus we discarded these columns for simplicity. As *TERT* promoter mutations are shared by

glioblastomas and oligodendrogliomas, but with two different prognostic outcomes, we concentrated on modules that correlated with TERT promoter mutation status.

For the LGG IDH wt subgroup, blue module (No. of genes: 1622, cor: 0.61, p-val: 2e-10), green module (No. of genes: 655, cor: 0.60, p-val: 5e-10), and tan module (No. of genes: 102, cor: 0.59, p-val: 9e-10) showed a positive correlation with TERT promoter mutation status. Yellow module (No. of genes: 830, cor: -0.53, p-val: 7e-08) and salmon module (No. of genes: 100, cor: -0.51, p-val: 4e-07) had negative correlation values that were below the threshold (Figure 4.3a).

For the LGG IDH mut subgroup, green module (No. of genes: 1032, cor: 0.86, p-val: 8e-124), salmon module (No. of genes: 391, cor: 0.68, p-val: 2e-58), blue module (No. of genes: 1354, cor: 0.54, p-val: 3e-33), darkgreen module (No. of genes: 64, cor: 0.52, pval: 8e-31), and lightgreen module (No. of genes: 106, cor: 0.52, p-val: 2e-30) showed positive correlation with TERT promoter mutation status. White module (No. of genes: 36, cor: -0.64, p-val: 4e-50) and greenyellow module (No. of genes: 538, cor: -0.57, p-val: 1e-37) showed negative correlation values that were below the threshold (Figure 4.3b).

The clinically relevant modules are showcased in the figures 4.3a-b as the rows indicate the modules detected through WGCNA and the columns indicate the clinical features selected in our study. The Pearson correlation values above 0.5 and below -0.5 were selected for stipulating the clinically significant modules. The darker shades of the color green and red highlight the clinically significant modules with lower and higher Pearson correlation values respectively.

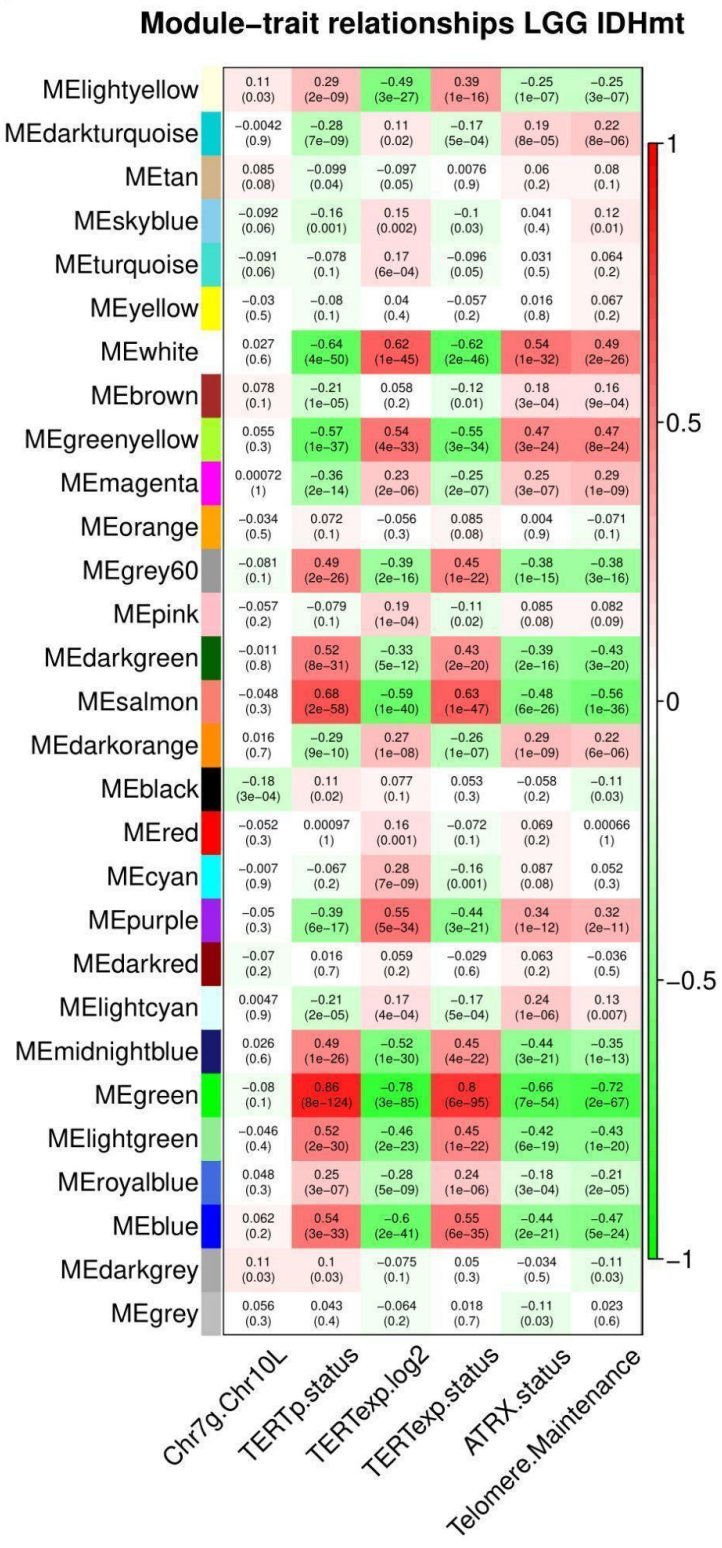

Figure 4.3a. Module-Trait relationship plot for LGG IDH mut samples.

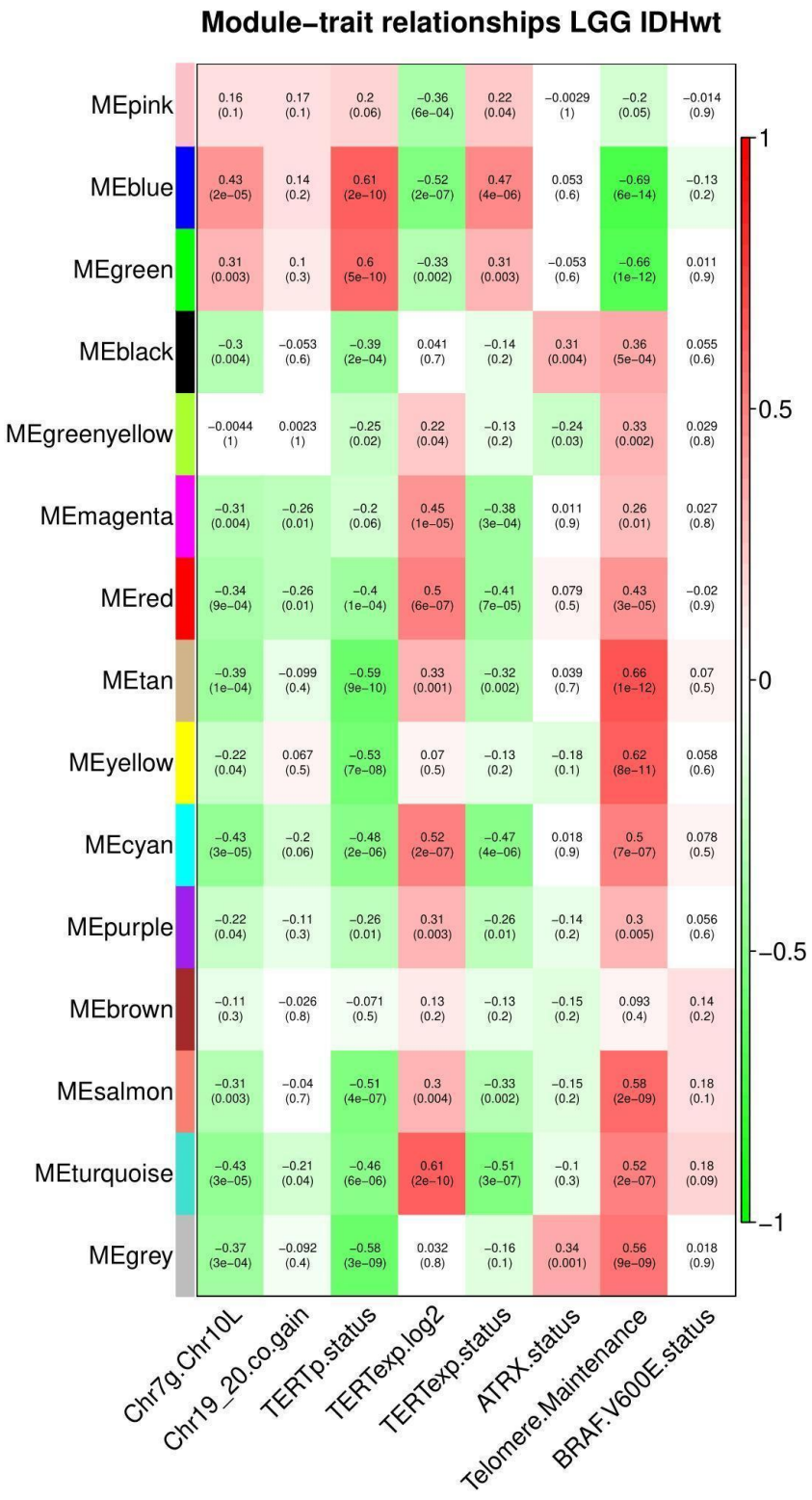

Figure 4.3b. Module-Trait relationship plot for LGG IDH wt samples.

4.3. Enrichment Analysis for Modules of Interest

The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment was conducted on the selected modules and the results are summarized in Table 4.1. For IDH wt group, we observed immune-related pathways enrichment in modules that are negatively correlated with the pTERT status (pTERT-) while synaptic and glutamatergic pathways enrichment in modules that are positively correlated (pTERT+) with the pTERT status. For IDH mut group, we observed cellular differentiation and proliferation, and synaptic pathways enrichment in modules that are negatively correlated with the pTERT status (pTERT-) while metabolomic and immune-related pathways enrichment in modules that are positively correlated (pTERT+) with the pTERT status. Despite being enriched in similar pathways, IDH wt pTERT+ blue module and IDH mut pTERT-greenyellow module have inverse correlations with pTERT status as observed in Table 4.1. The table shows identifiers of the statistically significant modules and their enriched pathways. We identified the biological pathways behind the "clinically significant" modules observed in Figures 4.3a-b.

Table 4.1. KEGG pathway enrichment of the clinically significant modules.

IDH mutation status	pTERT status (correlation)	Module	Term	p-value	Bonferroni
Wildtype	Negative	Yellow	hsa04060:Cytokine-cytokine receptor interaction	2.37E-10	7.29E-08
			hsa04064:NF-kappa B signaling pathway	8.00E-08	2.46E-05
			hsa04620:Toll-like receptor signaling pathway	1.68E-06	0.0005
			hsa04650:Natural killer cell mediated cytotoxicity	2.28E-06	0.0007
			hsa04062:Chemokine signaling pathway	2.62E-06	0.0008
		Salmon	hsa05235:PD-L1 expression and PD-1 checkpoint pathway in cancer	0.0106	0.8196
			hsa04660:T cell receptor signaling pathway	0.0161	0.9265
			hsa04630:JAK-STAT signaling pathway	0.0501	0.9997
			hsa04662:B cell receptor signaling pathway	0.0663	0.9999
			hsa04625:C-type lectin receptor signaling pathway	0.0997	0.9999
	Positive	Tan	hsa04621:NOD-like receptor signaling pathway	1.61E-10	1.34E-08
			hsa04622:RIG-I-like receptor signaling pathway	6.66E-06	0.0006

			hsa04625:C-type lectin receptor signaling pathway	0.0052	0.3486
			hsa04620:Toll-like receptor signaling pathway	0.0052	0.3486
			hsa04062:Chemokine signaling pathway	0.0399	0.9658
		Green	hsa04724:Glutamatergic synapse	1.07E-06	0.0003
			hsa04921:Oxytocin signaling pathway	3.44E-06	0.0009
			hsa04022:cGMP-PKG signaling pathway	4.03E-05	0.0116
			hsa04010:MAPK signaling pathway	4.49E-05	0.0129
			hsa04725:Cholinergic synapse	0.0004	0.1027
			hsa04728:Dopaminergic synapse	0.0005	0.1226
		Blue	hsa04721:Synaptic vesicle cycle	1.19E-18	3.77E-16
			hsa04727:GABAergic synapse	5.60E-13	1.77E-10
			hsa04724:Glutamatergic synapse	3.30E-11	1.04E-08
			hsa04723:Retrograde endocannabinoid signaling	5.64E-11	1.78E-08
			hsa04728:Dopaminergic synapse	5.40E-10	1.71E-07
			hsa04080:Neuroactive ligand-receptor interaction	8.20E-10	2.59E-07
			hsa04024:cAMP signaling pathway	9.60E-08	3.03E-05
Mutant	Negative	White	hsa05202:Transcriptional misregulation in cancer	0.0537	0.9903
			hsa04024:cAMP signaling pathway	0.0682	0.9974
			hsa04014:Ras signaling pathway	0.0765	0.9988
		Greenyellow	hsa04724:Glutamatergic synapse	2.11E-08	4.87E-06
			hsa04080:Neuroactive ligand-receptor interaction	3.42E-06	0.0008
			hsa04721:Synaptic vesicle cycle	6.58E-06	0.0015
			hsa04727:GABAergic synapse	0.0001	0.0294
			hsa04723:Retrograde endocannabinoid signaling	0.0004	0.0825
			hsa04725:Cholinergic synapse	0.0032	0.5218
	Positive	Green	hsa00100:Steroid biosynthesis	0.0031	0.6146
			hsa01100:Metabolic pathways	0.0220	0.9990
			hsa00280:Valine, leucine and isoleucine degradation	0.0374	0.9999
			hsa00250:Alanine, aspartate and glutamate metabolism	0.0416	0.9999
			hsa05231:Choline metabolism in cancer	0.0692	0.9999
		Salmon	hsa04146:Peroxisome	0.0001	0.0312

			hsa05022:Pathways of neurodegeneration - multiple diseases	0.0569	0.9999
			hsa04330:Notch signaling pathway	0.0954	1
			hsa04392:Hippo signaling pathway - multiple species	0.0998	1
		Blue	hsa04060:Cytokine-cytokine receptor interaction	1.81E-09	5.83E-07
			hsa04064:NF-kappa B signaling pathway	6.13E-09	1.98E-06
			hsa04662:B cell receptor signaling pathway	6.13E-08	1.97E-05
			hsa04621:NOD-like receptor signaling pathway	3.70E-07	0.0001
			hsa04620:Toll-like receptor signaling pathway	4.39E-07	0.0001
		Darkgreen	hsa03010:Ribosome	5.58E-11	3.23E-09
		Lightgreen	hsa04015:Rap1 signaling pathway	0.0075	0.7415
			hsa01100:Metabolic pathways	0.0270	0.9927
			hsa04921:Oxytocin signaling pathway	0.0615	0.9999
			hsa04218:Cellular senescence	0.0634	0.9999
			hsa04390:Hippo signaling pathway	0.0644	0.9999

To better understand resemblances and dissimilarities, we further looked into common KEGG pathways. In terms of synaptic pathways, we found 13 connections between the IDH mut green-yellow module and the IDH wt blue module (Table 4.2). Glutamate ionotropic receptor NMDA type subunits were marked in both modules in the "Glutamatergic Synapse" pathway, however glutamate ionotropic receptor AMPA type subunit 1 (GRIA1) was only annotated in the IDH wt blue module. While GRM2 was annotated only in the IDH wt blue module, glutamate metabotropic receptors were annotated in both the blue and greenyellow modules. In addition to GRM2, the IDH wt blue module was enriched for G protein alpha subunits (i1, i3, and o1) and adenylate cyclases (ADCY1, ADCY5). The Greenyellow and Blue modules come from different subtypes (IDH mut and IDH wt respectively) and they are oppositely correlated with pTERT status. The following table 4.2 shows the details of shared pathways between these two modules.

Table 4.2. Common pathways between the IDH mutant (IDH mut) greenyellow module and IDH wild-type (IDH wt) blue module.

IDH mut_greenyellow_pTERT-				**IDH wt_blue_pTERT+**			
Term	**Count**	**%**	**p-value**	**Term**	**Count**	**%**	**p-value**

hsa04724:Glutamatergic synapse	16	3.0246	0.0000	**hsa04721:Synaptic vesicle cycle**	36	2.2402	0
hsa04080:Neuroactive ligand-receptor interaction	24	4.5369	0.0000	**hsa04727:GABAergic synapse**	32	1.9913	0
hsa04721:Synaptic vesicle cycle	11	2.0794	0.0000	**hsa04724:Glutamatergic synapse**	34	2.1157	0
hsa04727:GABAergic synapse	10	1.8904	0.0001	**hsa04723:Retrograde endocannabinoid signaling**	39	2.4269	0
hsa04723:Retrograde endocannabinoid signaling	12	2.2684	0.0004	**hsa04728:Dopaminergic synapse**	35	2.1780	0.0000
hsa04725:Cholinergic synapse	9	1.7013	0.0032	**hsa04080:Neuroactive ligand-receptor interaction**	65	4.0448	0.0000
hsa04726:Serotonergic synapse	9	1.7013	0.0036	**hsa04713:Circadian entrainment**	29	1.8046	0.0000
hsa04360:Axon guidance	11	2.0793	0.0064	**hsa04024:cAMP signaling pathway**	43	2.6757	0.0000
hsa04024:cAMP signaling pathway	12	2.2684	0.0089	**hsa04725:Cholinergic synapse**	27	1.6801	0.0000
hsa04070:Phosphatidylinositol signaling system	6	1.1342	0.0606	**hsa04070:Phosphatidylinositol signaling system**	22	1.3690	0.0000
hsa04713:Circadian entrainment	6	1.1342	0.0606	**hsa04726:Serotonergic synapse**	24	1.4935	0.0000
hsa04728:Dopaminergic synapse	7	1.3233	0.0686	**hsa04360:Axon guidance**	30	1.8668	0.0002
hsa00330:Arginine and proline metabolism	4	0.7561	0.0952	**hsa00330:Arginine and proline metabolism**	9	0.5600	0.0441

The KEGG pathways were further investigated and downloaded from the database as shown in the Figures 4.4a-c. The individual pathways are highlighted in this study for the relevant modules. The different and same components of the pathway with common genes from IDH mut Greenyellow and IDH wt blue modules are shown in the figures. One component of the pathway includes more than one gene. G-protein-coupled receptors activate adenylate cyclase (AC), which in turn activates Protein Kinase PKA. The "Retrograde endocannabinoid signaling" route, where this interaction also takes place, is found in the mitochondria (Figure 4.4c). When adenylate cyclase is activated by G-protein coupled receptors, it turns on PKA. This

interaction also happens within the mitochondria in the “Retrograde endocannabinoid signaling” pathway (Figure 4.4c).

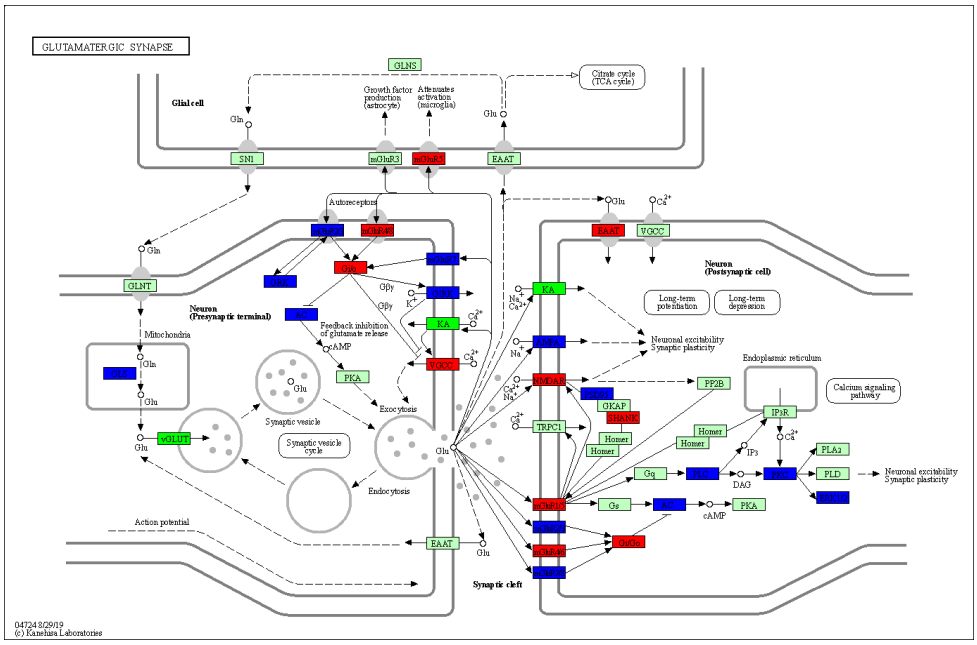

Figure 4.4a. Selected common KEGG pathway schemes of the IDH wt pTERT- blue module and IDH mut pTERT+ greenyellow module ~ Glutamatergic synapse

In figures 4.4a-c the pathways Glutamatergic synapse, GABAergic synapse and retrograde endocannabinoid signalling were selected. The components which were painted red were enriched in both greenyellow and blue module; blue components were only enriched in blue module and darkgreen components were only enriched in greenyellowmodule, while others were left in their default color.

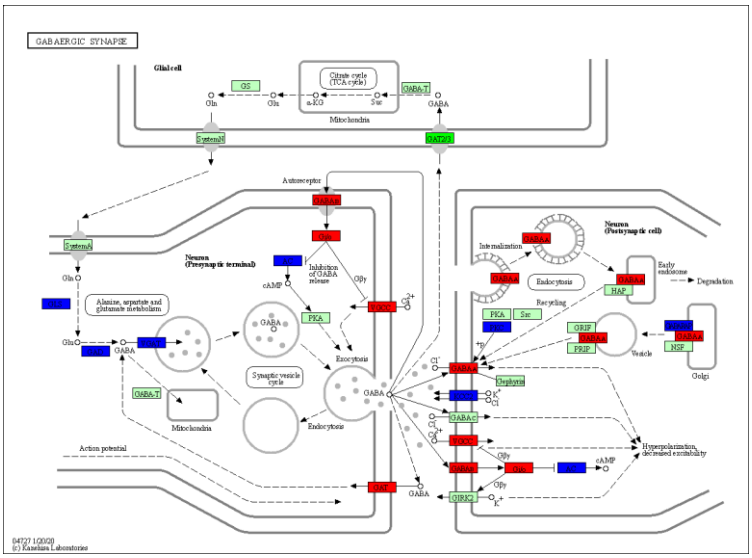

Figure 4.4b. Selected common KEGG pathway schemes of the IDH wt pTERT- blue module and IDH mut pTERT+ greenyellow module ~ GABAergic synapse

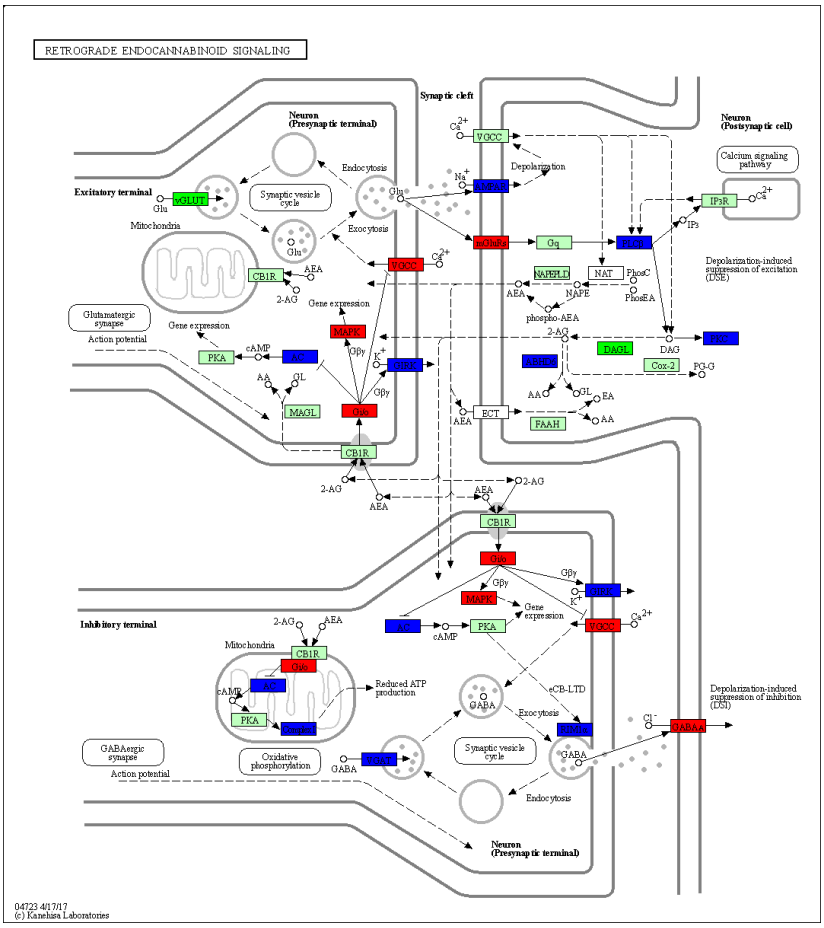

Figure 4.4c. Selected common KEGG pathway schemes of the IDH wt pTERT- blue module and IDH mut pTERT+ greenyellow module ~ Retrograde endocannabinoid signaling

4.4. Differentially Expressed Genes (DEGs)

We identified 3025 DEGs in LGG IDH mut group compared to normal samples and 2804 DEGs in LGG IDH wt group compared to normal samples. Additionally we have identified 1958 genes in LGG IDH wt group compared to LGG IDH mut group. As a result of screening DEGs in the modules of interests, we further identified 206 of co-expressed genes were also differentially expressed in the IDH mut pTERT-greenyellow module (module size 538), and 986 of co-expressed genes were also differentially expressed in the IDH wt pTERT+ blue module, (module size 1622). Moreover, we identified 146 common genes differentially expressed between normal vs. IDH wt and IDH mut vs IDH wt DEGs (upregulated: 53, downregulated: 93). Among DEGs we focused on genes that showed the highest change in expression in IDHwt tumors, while also being altered in IDH mut tumors albeit to a lesser extent in the same direction.

The *GRIN3A* gene had the lowest logFC value (IDH mut vs IDH wt -1.96; normal vs IDH wt -2.95) while the *HPD* gene had the highest logFC value (IDH mut vs IDH wt 2.01; normal vs IDH wt 6.53). Intriguingly, the logFC values for the IDH mut vs IDH wt and normal vs IDH wt groups did not differ significantly for genes that were downregulated, but they did differ significantly for genes that were upregulated. In the IDH wt blue module, we listed 182 genes, and in the IDH mut greenyellow module, 54 genes which were enriched in common pathways. 143 and 27 of them, respectively, were differentially expressed. 22 out of 182 blue module genes were also DEGs in both IDH mut and IDH wt. Additionally, 15 of these 22 genes showed differential expression between normal vs IDH wt samples (Table 4.3). These genes were enriched in the aforementioned pathways and are both differentially expressed between normal vs IDH wt and IDH mut and IDH wt. The genes highlighted (bold) were only enriched in blue module but not in greenyellow module.

Table 4.3. IDH wild-type blue module pathway-enriched DEGs.

	***Normal vs IDH wt**	**Normal (RPKM)**	**IDH wt (RPKM)**	***IDH mut vs IDH wt**	**IDH mut (RPKM)**	**IDH wt (RPKM)**
SLC1A6	-2.2365	989.8	234.4494	-1.1974	514.5239	234.4494
GRIA2	-1.4941	15266	5926.5056	-1.0580	11873.9067	5926.5056

PLCB1	-1.4048	5828.6	2446.4719	-1.0379	4828.8660	2446.4719
GNG12	2.2310	1181.8	6309.8539	1.5414	2087.4952	6309.8539
GRIN3A	-2.9484	1916.6	272.1124	-1.9578	1032.8684	272.1124
SSTR1	-2.4319	2161.4	449.9550	-1.2700	1018.7297	449.9551
SSTR2	-2.5210	2055	399.4607	-1.7434	1284.4665	399.4607
CHRNA4	-1.6524	2051.6	749.1685	-1.0276	1409.4377	749.1685
GABRB3	-2.2102	9864.4	2436.4719	-1.0300	4725.7607	2436.4719
GABRA3	-2.1735	1813.8	444.5843	-1.5058	1251.0909	444.5843
PDYN	-2.1334	5622.8	1506.6517	1.1126	645.9139	1506.6517
VIPR2	2.3059	329.6	1830.6180	-1.5287	4821.8445	1830.6180
CHRM4	-2.3974	334.2	70.5056	-1.022	136.3301	70.5056
ATP6V1G2	-2.0278	13698.6	3727.4045	-1.0769	7554.8421	3727.4045
GNAL	-2.2018	2163.6	530.2809	-1.0779	1077.7105	530.2809

4.5. Drug-Gene Interactions

GRIA2, PLCB1, GNG12, GABRA3, PDYN, and *GNAL* [6 of the 15 genes at the intersection of both DEGs lists (normal vs. IDH mut, IDH mut vs. IDH wt)] were solely enriched in the blue module. *PDYN* was not included since its expression pattern was not complying with our hypothesis of a gradual change of expression from normal to IDH mut to IDH wt. Only *GNG12* was upregulated among the genes above, others were downregulated (Table 4.4). Therefore, we have focused on inhibitory drugs / small molecules for *GNG12* and activatory drugs / small molecules for *GRIA2, PLCB1, GNG12, GABRA3,* and *GNAL*. Since *GNG12* was the most interesting candidate target, we have compared listed drugs for *GNG12* with drugs targeting other candidates (Table 4.4). As cisplatin, imatinib, vanadium pentoxide, and vemurafenib affect the expression of four out of five listed genes (including *GNG12*), these compounds were identified as potential therapeutics candidates. Table 4.4 shows the drugs that inactivate *GNG12* and activate the other four genes (*GRIA2, PLCB1, GABRA3, GNAL*) that are downregulated in our analyses.

Table 4.4. Drug-Gene interaction table for IDH wild-type blue module pathway-enriched DEGs.

Drug Name	***GNG12***	***GRIA2***	***PLCB1***	***GABRA3***	***GNAL***

1,25 dihydroxyvitamin d	x				
4-hydroxynonenal	x		x		x
Adenosine triphosphate	x	x		x	
Alfacalcidol	x				
Aminolevulinic acid	x				x
Androstanolone	x				
Aplidin	x				x
Apratoxin a	x				
Bexarotene	x	x			
Bisphenol a	x		x	x	
Cediranib	x			x	
Cetuximab	x				
Chlorpyrifos	x		x		
Cisplatin	x		x	x	x
Clinafloxacin	x				
Cytarabine	x				
Diclofenac	x	x			x
Doxorubicin	x	x			x
Doxycycline	x				
Estradiol	x		x		
Harman	x				
Imatinib	x		x	x	x
Interferon beta-1a	x				
Interferon gamma-1b	x				x
Mesalazine	x		x		
Metformin	x				x
Nickel	x				
Plx4032	x		x	x	
Puromycin, ec50, 1 d	x			x	
Puromycin, ec50, 5 d	x				
Resveratrol	x				
Sapphyrin pci-2050	x				
Tibolone	x				
Triiodothyronine-[13c6] hydrochloride (t3 thyronine)	x	x			
Trovafloxacin	x				
Vanadium pentoxide	x	x		x	x
Vemurafenib	x	x	x	x	
Vx	x				
Y15	x				

4.6. Data Collection and Preprocessing for DEG analysis and DNA methylation Data

The TCGA LGG and GBM DNA methylation data were carrying 485577 probes at the beginning. We extracted only Solid Tissue Normal samples from the GBM cohort. The LGG cohort is divided into subgroups by IDH mutation status and then each group is divided into further subgroups according to their pTERT mutation

status. Out of 419 LGG IDH mut samples, 93 of them were pTERTmt and 143 of them were pTERTwt. There was no pTERT status information for 183 samples. Out of 94 LGG IDH wt samples, 37 of them were pTERTmt and 19 of them were pTERTwt. There was no pTERT status information for 38 samples. After creating the subgroups according to their IDH and pTERT mutation status we further checked their 1p 19q codeletion status. For the IDH mut pTERTmt subgroup there were 86 1p19q-codeleted and 7 1p19q-intact samples. For the IDH mut pTERTwt subgroup there were 2 1p19q-codeleted and 141 1p19q-intact samples. All of the IDH wt samples were 1p19q-intact samples. We excluded 1p19q-intact samples from the IDH mut and pTERT mutant group and 1p19q-codeleted samples from the IDH mut and pTERTwt group for homogeneity.

The probes that displayed NA values for all the samples were filtered from each subgroup separately. After filtering, 396059 probes were left for solid tissue normal and 396065 probes left for the LGG subgroups. There were 428799 probes for 6 brain hemispheric cortex (CONTR_HEMI) samples extracted from the paper. There were 356774 probes which are the same with solid tissue normal samples. The solid tissue normal samples (TCGA-GBM) and the hemispheric cortex samples (CONTR, HEMI) (Capper et al., 2018) were merged to form a normal brain tissue beta matrix consisting of 8 samples and 356774 probes. The sample tree plot showed that although these normal samples originated from different studies, they were not very different from each other, and therefore, it was concluded that this beta matrix can be used as a normal brain beta matrix (Figure 4.5). The figure highlights the compatibility of normal samples coming from TCGA and the study mentioned above.

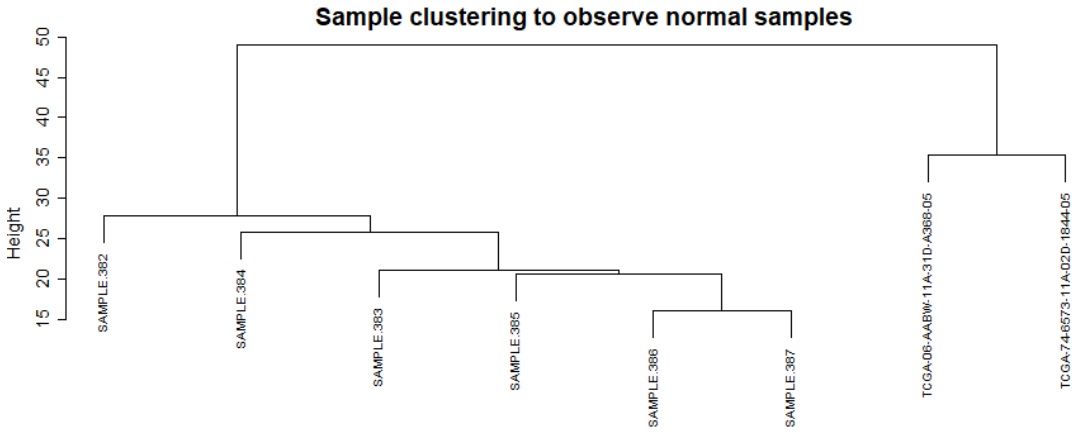

Figure 4.5. Sample clustering of normal samples from paper and TCGA-GBM (Solid Tissue Normal samples only). The branches are samples and the y-axis shows the branch height.

4.7. Differential Gene Expression Analysis with GTex Data to be Integrated with DNA Methylation Data

The probes of four different subgroups of LGG were filtered to be compatible with 356774 probes defined above. GTex expression data for brain tissue was carrying 707 samples, both TCGA and GTex expression data for brain tissue were carrying 58037 genes at the SE objects created at the beginning of differential expression analysis. Out of 707 samples in GTex brain expression data, 132 were expression data from brain cortex. We filtered LGG STAR - Counts data compatible with the subgroups defined above. After normalization and filtering the gene counts at the merged matrices, we ended up with 34003 genes and 218 samples for the IDH mut pTERTmt subgroup, 34148 genes and 271 samples for the IDH mut pTERTwt subgroup, 34188 genes and 169 samples for the IDH wt pTERTmt subgroup and 34390 genes and 151 samples for the IDH wt pTERTmt subgroup (132 samples of these samples are GTex brain cortex data).

Differential gene expression analysis results are shown in Table 4.5. In the IDH mt pTERTmt (1p-19q codeleted) subtype, 54.8% of DEGs were upregulated and 45.2% of DEGs were downregulated. In the IDH mut pTERTwt (1p-19q-intact) subtype, 52.6% of DEGs were upregulated and 47.4% of DEGs were downregulated. In the IDH wt pTERTmt subtype, 60.2% of DEGs were upregulated and 39.8% of DEGs were downregulated. In the IDH wt pTERTwt subtype, 58.1% of DEGs were upregulated and 41.9% of DEGs were downregulated. Interestingly, the IDH wt pTERTmt subtype shows higher upregulation patterns compared to other subtypes. The parameters used for DEG analysis included TCGAanalyze_DEA, fdr.cut = 0.01, logFC.cut = 2, and method = "glmLRT".

Table 4.5. The number of differentially expressed genes in different LGG subtypes.

	Total DEGs	**Upregulated**	**Downregulated**
IDH mut pTERTmt codel	4086	2240	1846

IDH mut pTERTwt noncodel	3873	2039	1834
IDH wt pTERTmt	5342	3216	2126
IDH wt pTERTwt	2937	1706	1231

4.8. Differentially Methylated Regions Analysis (CpG sites)

After imputing the missing values and normalizing the methylation data we observed the samples with QC.GUI function. Although the normal samples from the TCGA GBM and the aforementioned study (Capper et al in 2018) were separated at the dendrograms for all probes, the heatmap and Multidimensional scaling (MDS) plots for the most variable CpGs showed separation of tumor and normal samples for the most variable 1000 CpGs. Three tumor samples from the LGG IDH wt pTERTwt subgroup were clustered with normal samples at the MDS 1000 most variable positions plot. We have observed the adjusted *p-values* in this subgroup were relatively higher compared to other subgroups. Thus, we excluded three samples and reperformed the analysis.

4.9. IDH mut pTERTmt - Quality Control Plots

The dendrogram comparing the normal samples with tumor samples for quality control is shown in Figure 4.6a. A multidimensional scaling plot highlighting the red dots as normal samples and green dots as tumor samples for 1000 most variable positions is shown in Figure 4.6b. A heatmap constructed for 1000 most variable CpGs with rows indicating CpGs and columns displaying the samples, is depicted in Figure 4.6c. These plots summarize the background quality control checks conducted on the samples for DNA methylation data collection and preprocessing.

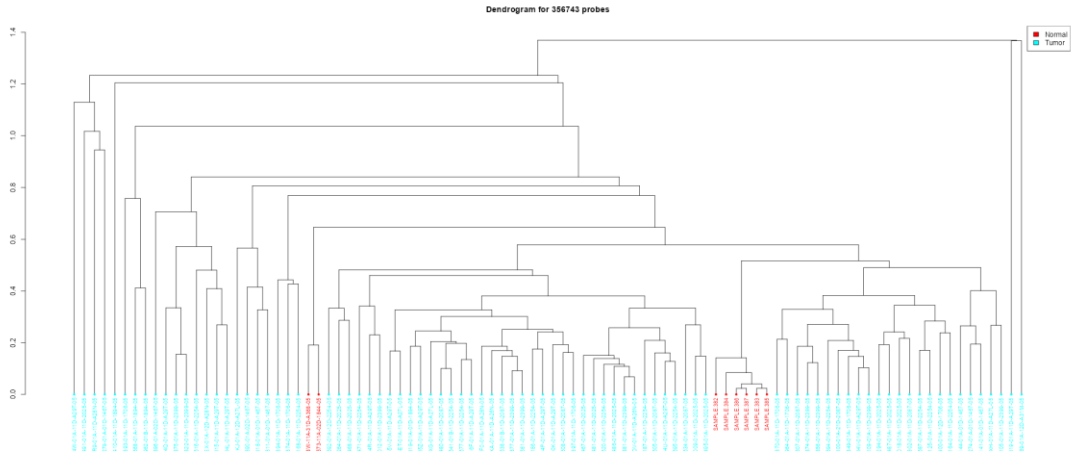

Figure 4.6a. Sample dendrogram for 356743 probes in IDH mut pTERTmt group.

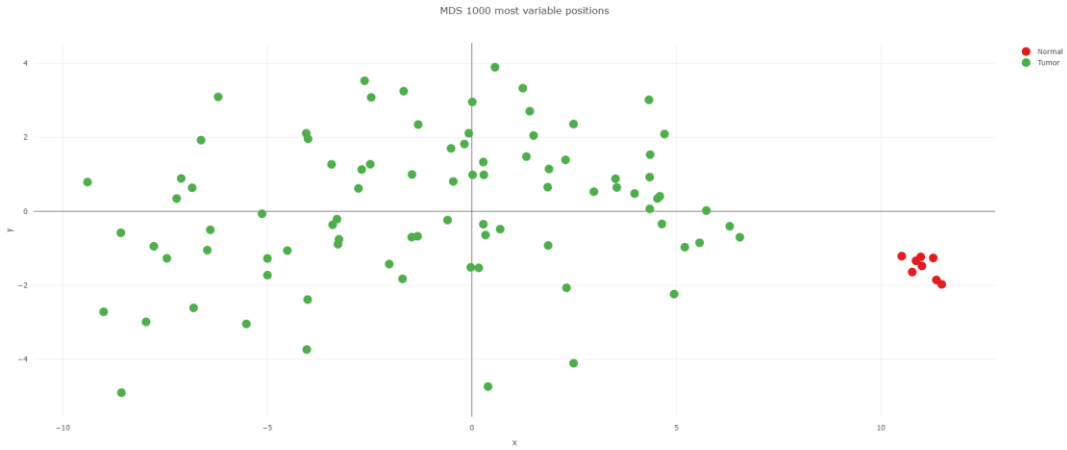

Figure 4.6b. Multidimensional Scaling (MDS) plot for 1000 most variable positions for IDH mut pTERTmt group.

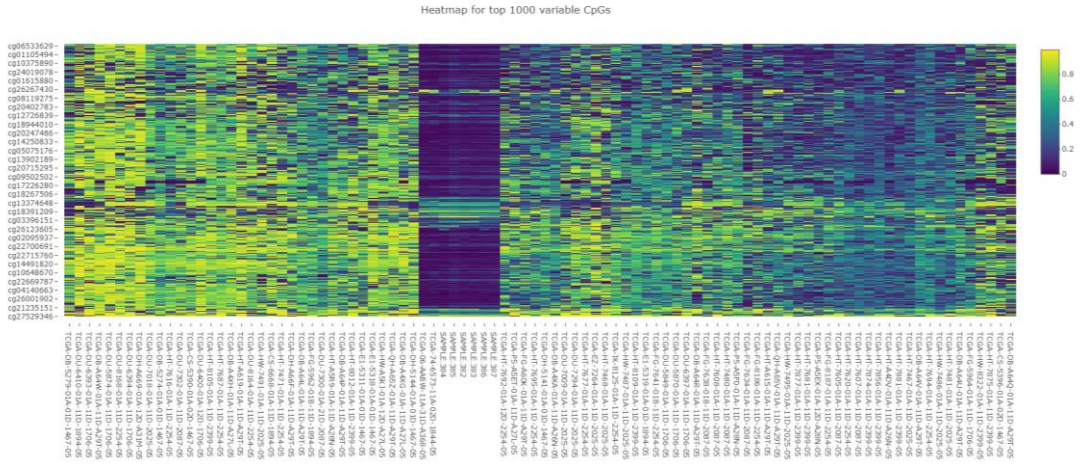

Figure 4.6c. Heatmap for Top 1000 variable CpGs for IDH mut pTERTmt group.

4.10. IDH mut pTERT wt - Quality Control Plots

The dendrogram comparing the normal samples with tumor samples for quality control of IDH mut pTERTwt genetic subtype is indicated in Figure 4.7a. This data contains 356730 probes in total with branch height displayed in y-axis. Another MDS plot for top 1000 variable CpGs for IDH mut pTERTwt is shown in Figure 4.7b, where the red dots show normal samples and green dots indicate the tumor samples. The heat map is also drawn for top 1000 variable CpGs from these samples, where the rows display the CpGs and columns depict the samples (Figure 4.7c).

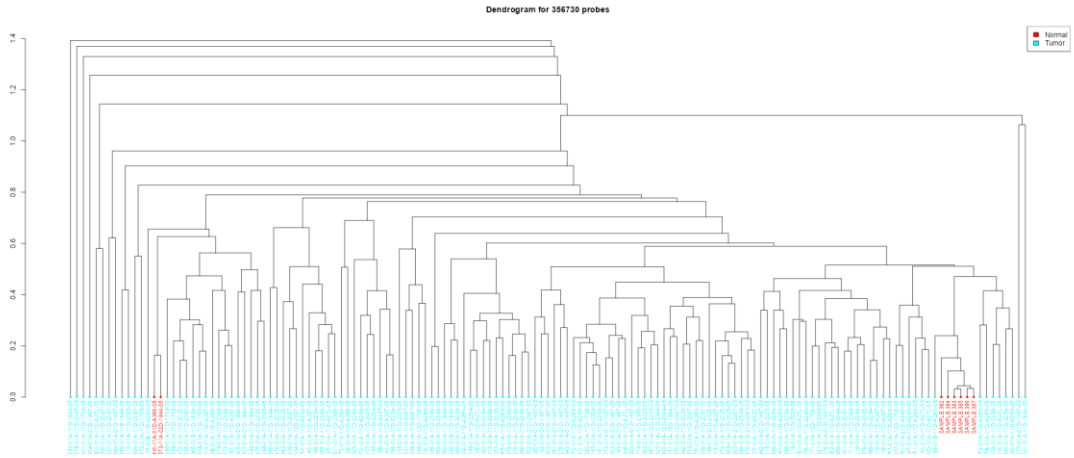

Figure 4.7a. Sample dendrogram for 356730 probes in IDH mut pTERTwt group.

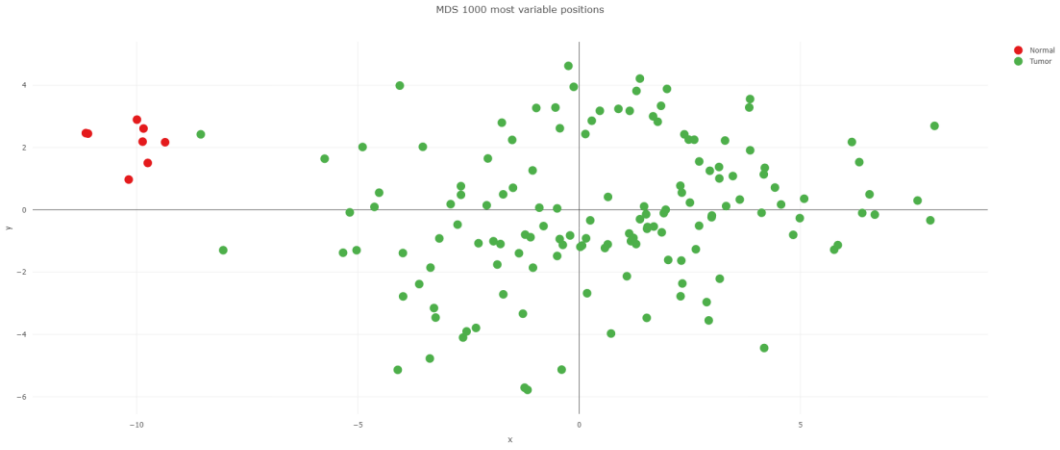

Figure 4.7b. Multidimensional Scaling (MDS) plot for 1000 most variable positions for IDH mut pTERTwt group..

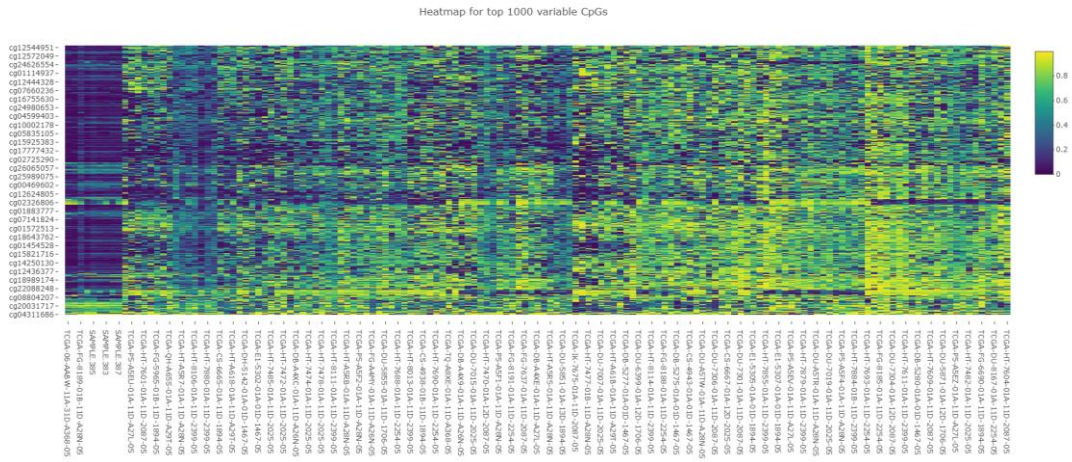

Figure 4.7c. Heatmap for Top 1000 variable CpGs for IDH mut pTERTwt group.

4.11. IDH wt pTERTmt - Quality Control Plots

The dendrogram comparing the normal samples and tumor samples for the genetic subtype of IDH wt pTERTmt- is shown in Figure 4.8a for highlighting the quality control of the samples. Similarly, the MDS plot displaying the normal samples in red and tumor samples in green for the top 1000 variable positions is shown in Figure 4.8b. The heat map for the variable regions is shown in Figure 4.8c.

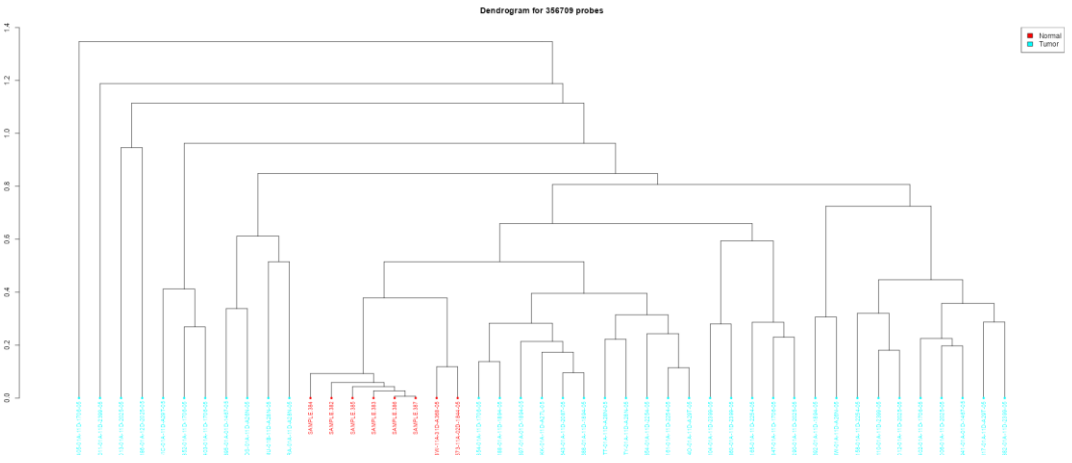

Figure 4.8a. Sample dendrogram for 356709 probes in IDH wt pTERTmt group.

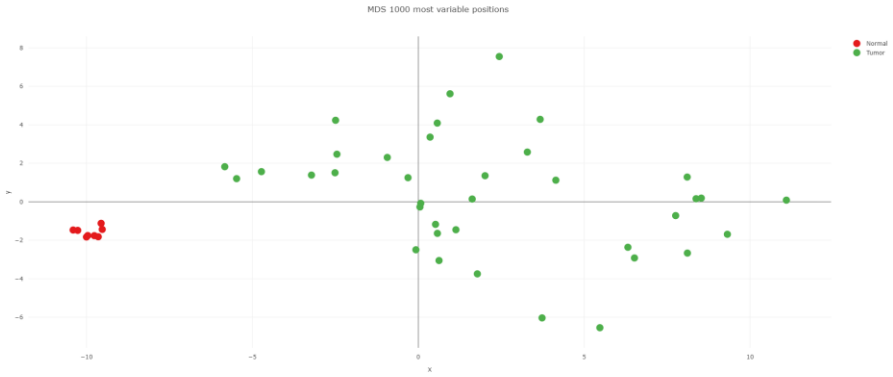

Figure 4.8b. Multidimensional Scaling (MDS) plot for 1000 most variable positions for IDH wt pTERTmt.

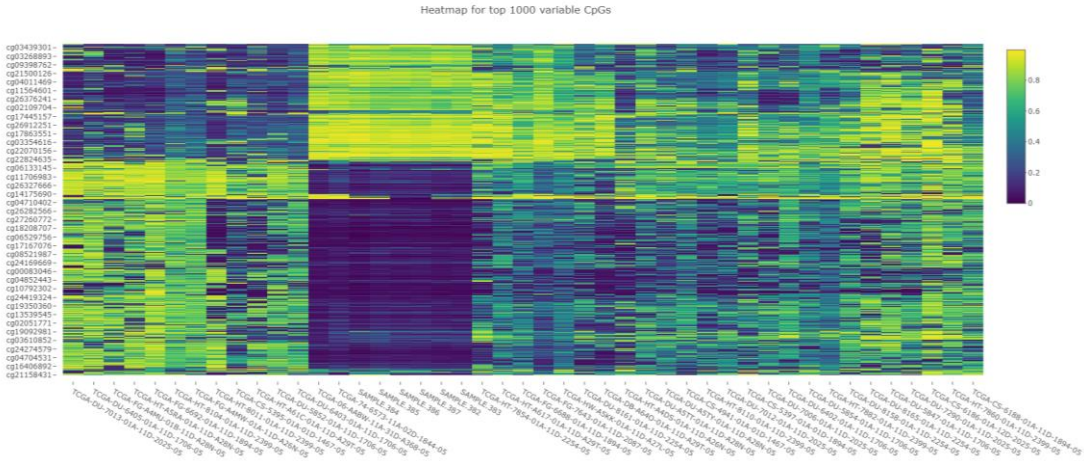

Figure 4.8c. Heatmap for Top 1000 variable CpGs for IDH wt pTERTmt.

4.12. IDH wt pTERTwt - Quality Control Plots

The MDS for 1000 most variable positions in IDH wt pTERTwt is shown in Figure 4.9a, where the red dots display the normal samples and green dots show the tumor samples. Moreover, the dendrogram is shown in Figure 4.9b to show the quality control check and processing of the samples and MDS plot is shown in Figure 4.9c and heat map in Figure 4.9d.

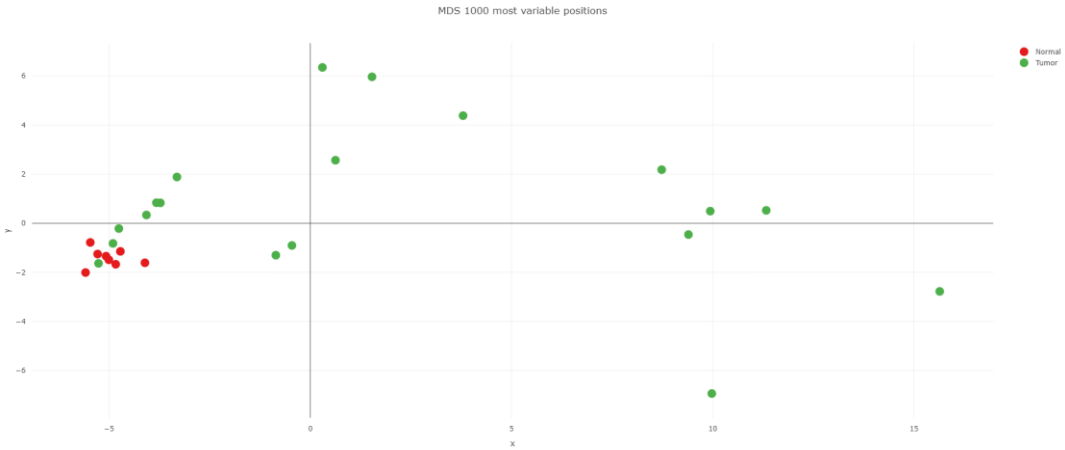

Figure 4.9a. MDS plot of IDH wt pTERTwt samples for IDH wt pTERTwt subtype.

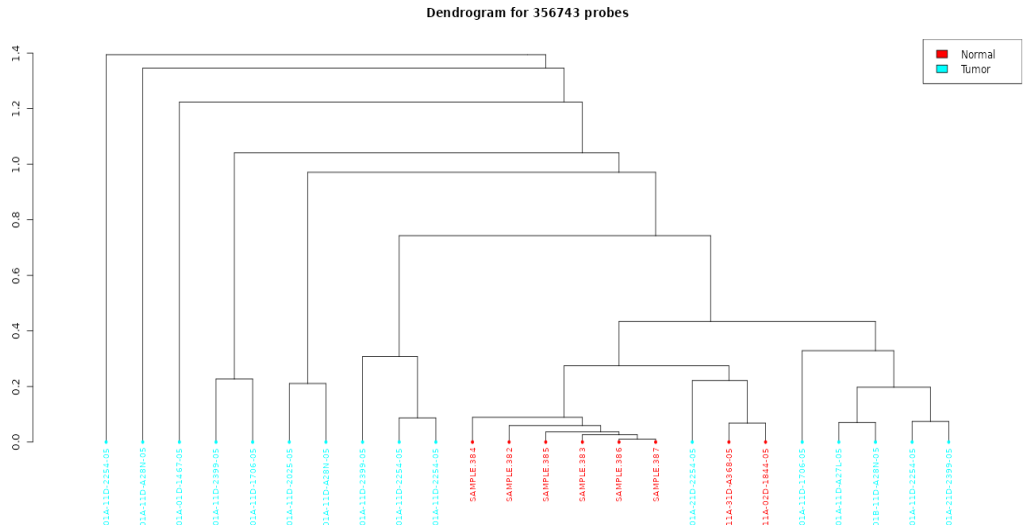

Figure 4.9b. Sample dendrogram for 356743 probes in IDH wt pTERTwt group.

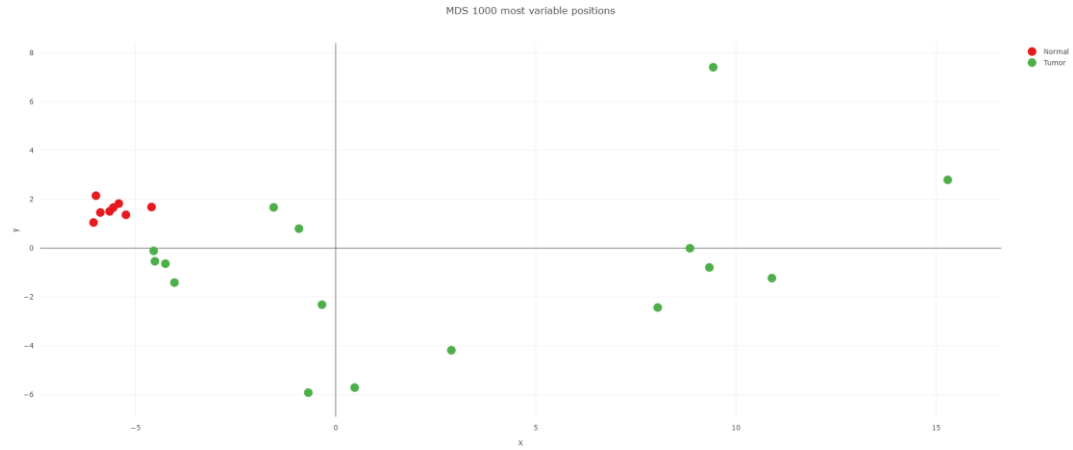

Figure 4.9c. Multidimensional Scaling (MDS) plot for 1000 most variable positions for IDH wt pTERTwt group.

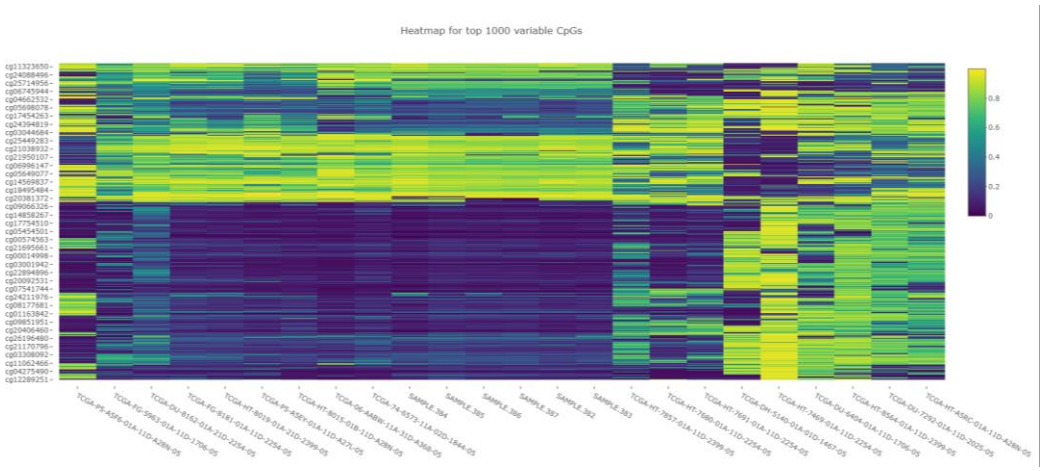

Figure 4.9d. Heatmap for Top 1000 variable CpGs for IDH wt pTERTwt group.

The summary of Differentially Methylated CpGs (DMC) analysis results are shown in Table 4.6. The differentially methylated genes (DMGs) were calculated with median of the probes annotated to the particular gene. These differentially methylated probes and genes were identified in 4 different LGG subtypes. We identified 7802 DMGs in the IDH mut pTERTmt (1p-19q codeleted) subtype; 97.1% of these DMGs were hypermethylated and 2.9% were hypomethylated. In the IDH mut pTERTwt (1p-19q intact) subtype we identified 7137 DMGs; 96% of them were hypermethylated and 4% were hypomethylated.

In the IDH wt pTERTmt subtype we identified 3459 DMGs; 49.2% of them were hypermethylated and 50.8% were hypomethylated. In the IDH wt pTERTwt subtype we identified 626 DMGs; 63.6% of them were hypermethylated and 36.4% were hypomethylated. IDH wt samples showed an increased hypomethylation pattern

compared to IDH mut samples. Interestingly, in the IDH wt pTERTmt subtype higher percentage of genes were hypomethylated than hypermethylated, which is just the opposite of other samples showing higher percentage of hypermethylated genes. The parameters used to arrive at the number of DMCs include TCGAanalyze_DMC, adj.p value < 0.01, and diffmean.cut = 0.3.

Table 4.6. Summary of Differentially Methylated CpG sites (DMC).

	hyper probs	**hypo probs**	**hyper genes**	**hypo genes**
IDH mut pTERTmt	22162	495	7573	229
IDH mut pTERTwt	18129	601	6852	285
IDH wt pTERTmt	3721	2570	1703	1756
IDH wt pTERTwt	508	273	398	228

4.13. Integration of Methylation and Expression Data

The venn diagram was constructed which shows the hypermethylated+downregulated; hypomethylated+upregulated genes' distribution in 4 different subtypes (based on IDH and TERT promoter mutation status) of low grade gliomas. The Figure 4.10 shows that there were 422 genes which were common MDGs between IDH mut pTERTmt and IDH wt pTERTwt subtypes (more than 65% of the MDGs in each group), 337 of them were common only between them (more than 65% of the MDGs in each group). This result was expected as IDH mutations are mainly responsible for epigenetic changes in gliomas and these two subtypes are composed of IDH mutant samples. Meanwhile, only 150 MDGs were common among other groups, out of 422 MDGs (35.5%) compared to 272 out of 422 MDGs from this group. 72 of them were hypermethylated and downregulated while 200 of these were hypomethylated and upregulated.

The summary of integration of differential expression and differential methylation data (hyper methylated and downregulated; hypomethylated and upregulated) in 4 different LGG subtypes is shown in Table 4.7.

Table 4.7. The summary of Methylation Driven Genes (MDGs) in different genetic subtypes of LGG.

MDGs Summary table		
	hyper.down	**hypo.up**
IDH mut pTERTmt	617	31
IDH mut pTERTwt	585	26
IDH wt pTERTmt	200	222
IDH wt pTERTwt	7	7

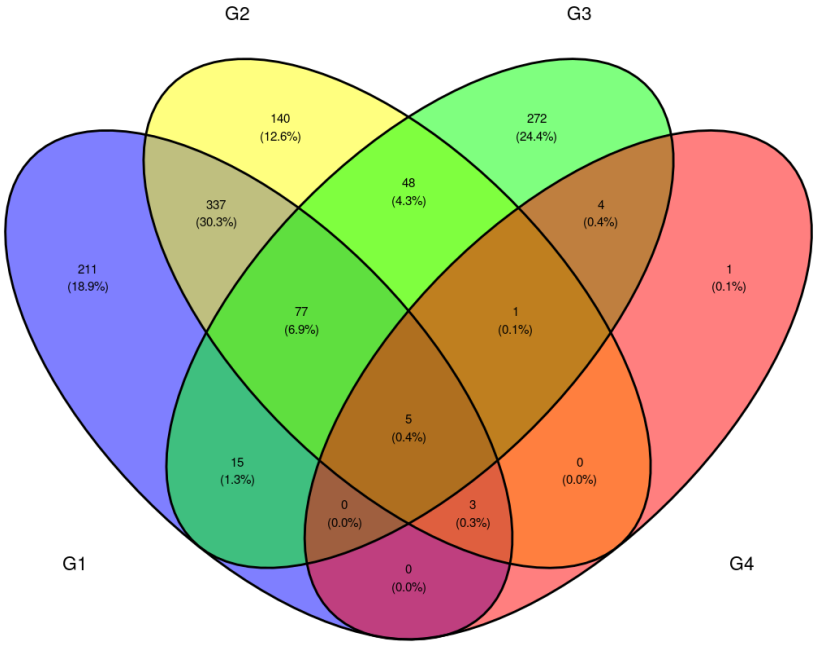

Figure 4.10. Venn diagram of methylation driven genes (G1: IDH mut pTERTmt, G2: IDH mut pTERTwt, G3: IDH wt pTERTmt, G4: IDH wt pTERTwt).

Table 4.8 details the enrichment analysis of the genes in different sections of the venn diagram. Hypermethylated and downregulated genes were enriched in Neuroactive ligand receptors while hypomethylated and upregulated genes were enriched in immune related pathways (Table 4.8).

77 genes at the intersection of G1&G2&G3-only were enriched in ion channel activities, transport activities (GO Molecular Ontology) and in nervous systems (GABAergic synapse, Retrograde endocannabinoid signaling) and environmental information processing (Neuroactive ligand-receptor interaction). 337 genes at the G1&G2-only intersection were enriched with Calmodulin binding and Voltage-gated channel activities (GO Molecular Ontology). 48 genes at the G2&G3-only intersection

were enriched with ion channel activities and transport activities (GO Molecular Ontology). 140 genes at the G2-only area were enriched with Nervous System Pathways (GABAergic synapse, Glutamatergic synapse, Long-term potentiation, Cholinergic synapse, Synaptic vesicle cycle, Dopaminergic synapse, Serotonergic synapse), Signal Transduction (Phosphatidylinositol signaling system, Calcium signaling pathway, cAMP signaling pathway, Rap1 signaling pathway, MAPK signaling pathway) and Environmental Information Processing (Neuroactive ligand-receptor interaction) (Table 4.8).

272 genes at the G3-only area were enriched with Signaling molecules and interaction (Cytokine-cytokine receptor interaction, Neuroactive ligand-receptor interaction) Molecular Ontology enrichment for this subtype was receptor activities (immune receptor activity, signaling receptor activity) and binding (tumor necrosis factor binding, extracellular matrix binding, cytokine binding). Interestingly, we have detected immune system activities for the G3 only area and we further investigated these genes. 72 out of 272 genes were hyper methylated and downregulated, and they were related to the neuroactive ligand-receptor interaction pathway. 200 out of 272 genes were hypomethylated and upregulated, and they were enriched with immune system related pathways. We have further identified 56 of these genes in the immune system at the Reactome pathway [HSA-168256, Immune System (56/1956)] (Table 4.8).

Table 4.8. Enrichment analysis of Venn Diagram genes (*p-value* < 0.05).

Group	Term	Count	%	p-Value	Fold Enrichment	Bonferroni	Benjamini	FDR
G1G2G3 (77 genes)	hsa04727:GABAergic synapse	6	7.79	1.82E-05	17.29	0.0016	0.0008	0.0008
	hsa04080:Neuroactive ligand-receptor interaction	8	10.38	0.0004	5.59	0.0300	0.0077	0.0076
	hsa04723:Retrograde endocannabinoid signaling	5	6.49	0.0022	8.66	0.1736	0.0381	0.0377
	hsa04024:cAMP signaling pathway	4	5.19	0.0498	4.64	0.9883	0.6193	0.6121
G1G2-only (337 genes)	hsa04080:Neuroactive ligand-receptor interaction	16	4.74	6.77E-05	3.34	0.0147	0.0148	0.0147
	hsa04020:Calcium signaling pathway	11	3.26	0.0010	3.51	0.2017	0.1126	0.1120
	hsa04014:Ras signaling pathway	10	2.96	0.0034	3.25	0.5204	0.2445	0.2434
	hsa04723:Retrograde endocannabinoid signaling	7	2.07	0.0121	3.63	0.9300	0.5286	0.5262

	hsa04724:Glutamatergic synapse	6	1.78	0.0158	4.04	0.9691	0.5749	0.5723
	hsa04727:GABAergic synapse	5	1.48	0.0278	4.31	0.9979	0.7585	0.7550
	hsa04728:Dopaminergic synapse	6	1.78	0.0278	3.49	0.9979	0.7585	0.7550
	hsa04713:Circadian entrainment	5	1.48	0.0365	3.95	0.9997	0.7965	0.7929
	hsa04721:Synaptic vesicle cycle	3	6.25	0.0198	13.15	0.7281	0.5734	0.5734
G2-only (48 genes)	hsa04020:Calcium signaling pathway	12	8.57	7.21E-07	6.95	0.0001	0.0001	0.0001
	hsa04724:Glutamatergic synapse	7	5	0.0001	8.54	0.0226	0.0059	0.0052
	hsa04024:cAMP signaling pathway	9	6.42	0.0001	5.66	0.0233	0.0059	0.0052
	hsa04727:GABAergic synapse	6	4.28	0.0004	9.38	0.0610	0.0126	0.0111
	hsa04080:Neuroactive ligand-receptor interaction	10	7.14	0.0010	3.79	0.1473	0.0234	0.0206
	hsa04929:GnRH secretion	5	3.57	0.0010	10.86	0.1545	0.0234	0.0206
	hsa04725:Cholinergic synapse	6	4.28	0.0012	7.38	0.1707	0.0234	0.0206
	hsa04721:Synaptic vesicle cycle	5	3.57	0.0022	8.91	0.2955	0.0338	0.0298
	hsa04728:Dopaminergic synapse	6	4.28	0.0023	6.32	0.3111	0.0338	0.0298
	hsa04015:Rap1 signaling pathway	7	5	0.0035	4.64	0.4309	0.0433	0.0382
	hsa04010:MAPK signaling pathway	8	5.71	0.0044	3.78	0.5108	0.0452	0.0399
	hsa04070:Phosphatidylinositol signaling system	5	3.57	0.0048	7.17	0.5377	0.0453	0.0399
	hsa04726:Serotonergic synapse	5	3.57	0.0087	6.05	0.7545	0.0777	0.0685
	hsa04720:Long-term potentiation	4	2.85	0.0117	8.30	0.8487	0.0988	0.0872
	hsa04270:Vascular smooth muscle contraction	5	3.57	0.0146	5.19	0.9066	0.1177	0.1038
	hsa05214:Glioma	4	2.85	0.0158	7.42	0.9230	0.1211	0.1068
	hsa05207:Chemical carcinogenesis - receptor activation	6	4.28	0.0166	3.94	0.9325	0.1215	0.1071
	hsa04723:Retrograde endocannabinoid signaling	5	3.57	0.0203	4.70	0.9633	0.1422	0.1254
	hsa04014:Ras signaling pathway	6	4.28	0.0251	3.54	0.9833	0.1684	0.1485
	hsa04022:cGMP-PKG signaling pathway	5	3.57	0.0300	4.16	0.9926	0.1846	0.1628
	hsa04713:Circadian entrainment	4	2.85	0.0309	5.73	0.9937	0.1846	0.1628
G3-Only (272 genes)	hsa04060:Cytokine-cytokine receptor interaction	16	5.88	4.68E-05	3.48	0.0102	0.0102	0.0100
	hsa04080:Neuroactive ligand-receptor interaction	17	6.25	0.0002	2.97	0.0345	0.0176	0.0172
	hsa04610:Complement and coagulation cascades	8	2.941	0.0004	5.96	0.0741	0.0257	0.0251
	hsa04640:Hematopoietic cell lineage	8	2.94	0.0008	5.18	0.1654	0.0452	0.0442

	hsa04350:TGF-beta signaling pathway	7	2.57	0.0035	4.72	0.5311	0.1138	0.1112
	hsa05150:Staphylococcus aureus infection	7	2.57	0.0036	4.67	0.5497	0.1138	0.1112
	hsa04061:Viral protein interaction with cytokine and cytokine receptor	7	2.57	0.0044	4.49	0.6233	0.1218	0.1190
	hsa05145:Toxoplasmosis	7	2.57	0.0077	4.01	0.8155	0.1683	0.1645
	hsa04621:NOD-like receptor signaling pathway	8	2.94	0.0239	2.79	0.9950	0.4024	0.3932
	hsa04072:Phospholipase D signaling pathway	7	2.57	0.0271	3.03	0.9976	0.4238	0.4141
	hsa05200:Pathways in cancer	15	5.51	0.0347	1.81	0.9996	0.5063	0.4947

4.14. Network Analysis and EpiFactors

The methylation driven genes alone are not informative enough on a biological level as these are individual genes. We further included the first neighbors of MDGs and constructed a more biologically interpretable network. The MDGs + first neighbor were still very crowded. Therefore, we focused on Epigenetic factor protein complexes and attempted to understand which epigenetic protein complexes might be playing a role in methylation differences.

Table 4.9. Network statistics on epigenetic data with MDG nodes, edges, neighbors and protein complexes.

	MDGs nodes	**MDGs edges**	**MDGs + first neighbors nodes**	**MDGs + first neighbors edges**	**Protein complex focused nodes**	**Protein complex focused edges**
IDH mut pTERTmt	427	390	3907	93668	435	415
IDH mut pTERTwt	411	477	3620	88792	419	505
IDH wt pTERTmt	309	313	4131	155613	326	363
IDH wt pTERTwt	9	1	125	596	11	5

Table 4.9 summarizes the statistics from different networks constructed. Epigenetic factor protein complex focused subnetworks show a clear difference between IDH mut pTERTwt (G3) group and IDHmut groups (G1, G2). For simplicity smaller networks and single nodes were excluded during visualization. Interestingly GABA complexes created a separate network in G1, G2 and G3. Potassium voltage-

gated channel subfamily members were separate in G1 but they were wired to the large network in G2 and G3. As shown in Table 4.9, G2 large network structures have higher numbers of calcium signaling pathways (Figure 12a-d).

The G1 protein complex focused subnetwork was highlighted in Figure 4.12a for IDH mut pTERTmt subtype, where red genes were hypermethylated and downregulated, green genes were hypomethylated and upregulated, and blue genes showed protein complex genes. Furthermore, G2 protein complex focused subnetwork for IDH mut pTERTwt data contained red genes, which are hypermethylated and downregulated, green genes are hypomethylated and upregulated, whereas blue genes are protein complex genes (Figure 4.12b). The G3 protein complex focused subnetwork for IDH wt pTERTmt subtype is highlighted in Figure 4.12c, where red genes are hypermethylated and downregulated, green genes are hypomethylated and upregulated, and blue genes protein complex genes. Subsequently, the G4 protein complex focused subnetwork for IDH wt pTERTwt subtype is displayed in Figure 4.12d, where red genes are hypermethylated and downregulated, green genes are hypomethylated and upregulated, and blue genes showcase the protein complex genes.

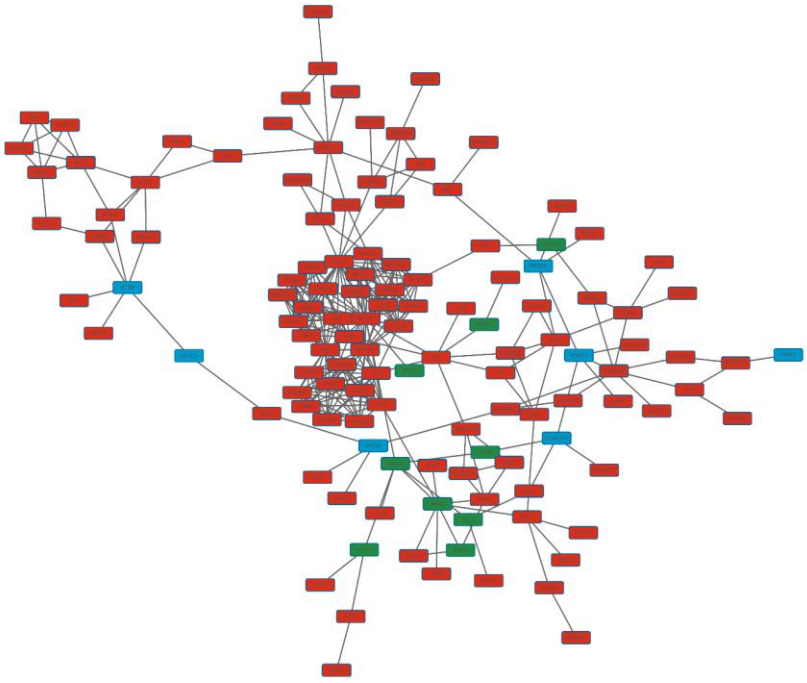

Figure 4.11a. IDH mut pTERTmt - Protein complex focused subnetwork

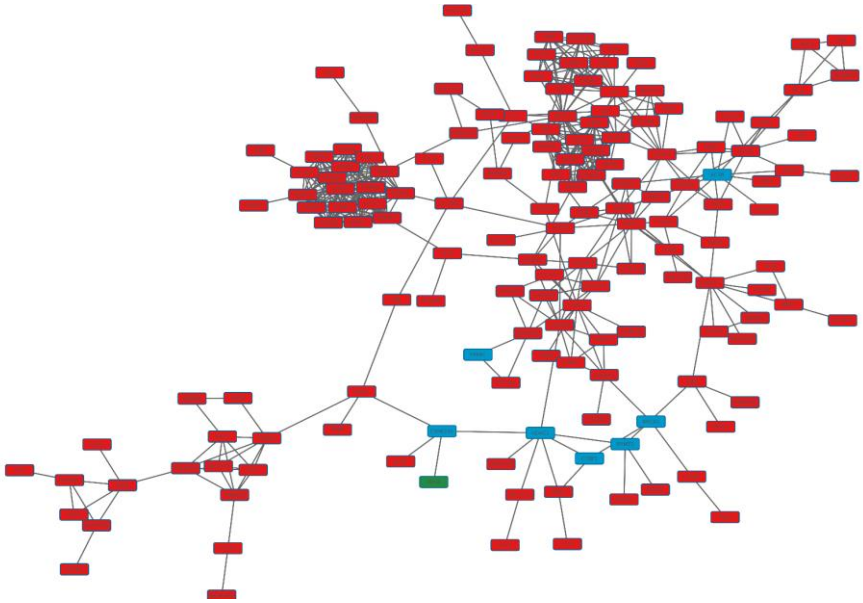

Figure 4.11b. IDH mut pTERTwt - Protein complex focused subnetwork

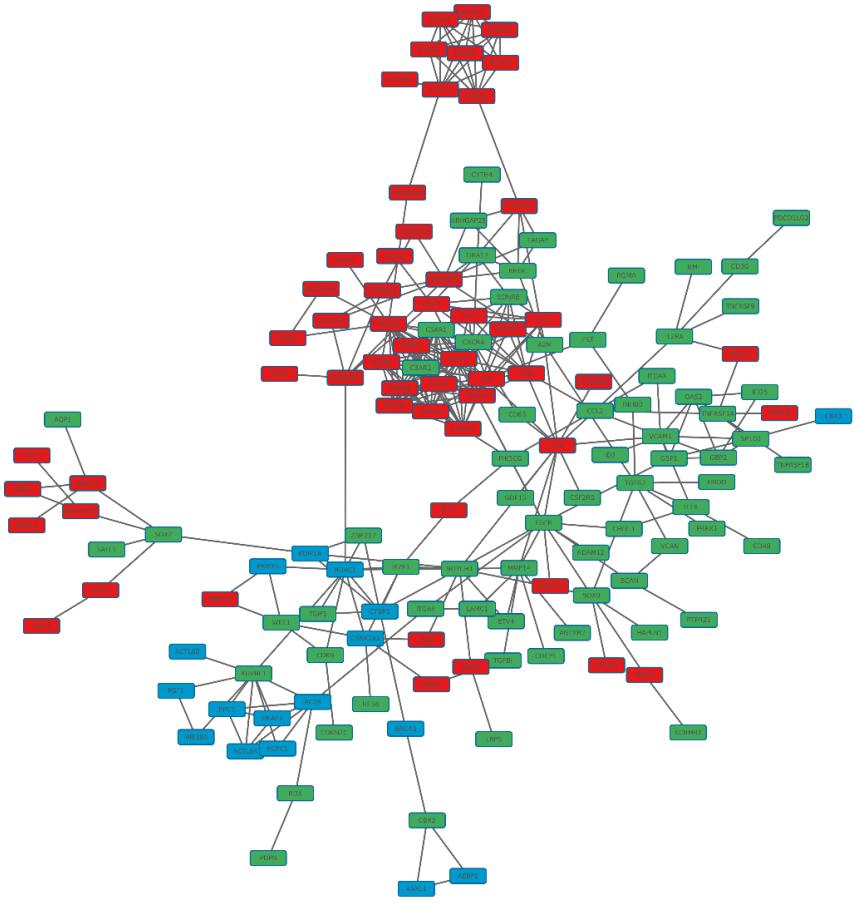

Figure 4.11c. IDH wt pTERTmt - Protein complex focused subnetwork

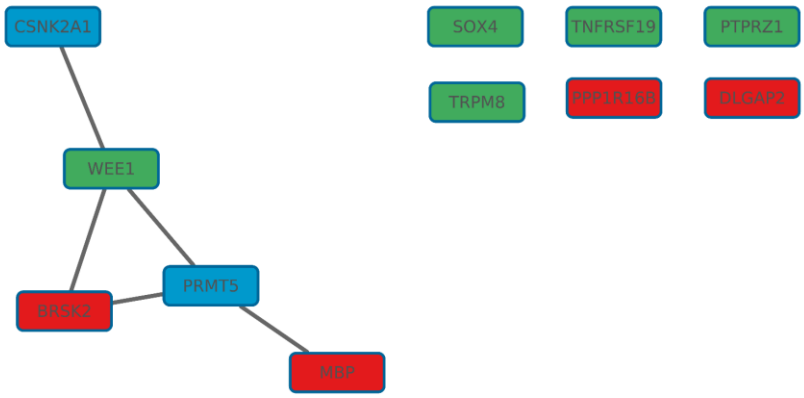

Figure 4.11d. IDH wt pTERTwt - Protein complex focused subnetwork

4.15. Perturbation Profiling

DepMap screening of the protein complex focused subnetwork genes returned perturbation effect of 6 genes (*ACTB, BRCA1, HCFC1, MET, PPP4C, PRMT5*) from IDH mut pTERTmt subtype, 4 genes (*ACTB, BRCA1, PPP4C, PRMT5*) from IDH mut pTERTwt subtype, 11 genes (*ACTB, ACTL6A, BRCA1, CDK6, HCFC1, MIS18A, PRMT5, RUVBL1, SOX2, WEE1, ZNF217*) from IDH wt pTERTmt subtype and 2 genes (*PRMT5, WEE*1) from IDH wt pTERTwt subtype at least in one diffuse glioma model analyzed by the DepMap (Table 4.10). The Chronos dependency score is derived from a cell depletion assay, where a lower score suggests a greater likelihood of the gene being essential in a specific cell. A score of 0 implies non-essentiality, while -1 is akin to the median among all pan-essential genes. The Chronos dependency scores were available for 66 diffuse glioma models which included 2 for Anaplastic Astrocytoma (AASTR), 13 for Astrocytoma (ASTR), 49 for Glioblastoma (GB), and 2 for Oligodendroglioma (ODG). The data is compiled to show the genes which were selected from network modules processed in our study for various subtypes and they display the perturbation effect on different aforementioned models.

Table 4.10. CHRONOS dependency scores of 66 diffuse glioma models

Models	*ACTB*	***ACTL6A***	*BRCA1*	*CDK6*	***HCFC1***	*MET*	*MIS18A*	*PPP4C*	***PRMT5***	***RUVBL1***	*SOX2*	***WEE1***	*ZNF217*
AASTR_ACH-001126	NA	-1.75	NA	-1.17	-2.04	NA	-1.10	-1.67	-1.05	-2.40	NA	-2.60	NA
AASTR_ACH-002257	-1.05	-1.99	NA	-1.08	-1.96	-1.28	-1.03	NA	-1.48	-2.50	NA	-1.94	NA
ASTR_ACH-000040	NA	-1.69	NA	NA	-1.80	NA	NA	NA	-1.05	-2.27	NA	-2.53	NA
ASTR_ACH-000128	NA	-2.29	NA	NA	-2.14	NA	-1.29	-1.50	-1.06	-1.82	-1.50	-1.35	-1.33
ASTR_ACH-000232	NA	-1.74	NA	NA	-1.65	NA	-1.18	NA	NA	-2.08	-1.06	-1.96	NA
ASTR_ACH-000329	-1.00	-1.84	NA	NA	-2.29	NA	NA	NA	NA	-2.45	NA	-3.03	NA
ASTR_ACH-000389	NA	-1.46	NA	NA	-1.93	NA	NA	-1.13	NA	-2.43	NA	-2.79	NA
ASTR_ACH-000437	NA	-1.46	NA	NA	-1.75	NA	-1.36	NA	-1.99	-2.34	NA	-2.31	NA
ASTR_ACH-000591	NA	-1.29	NA	NA	-1.59	NA	-1.53	NA	-1.10	-2.30	NA	-2.50	NA
ASTR_ACH-000592	NA	-1.91	NA	NA	-2.18	NA	-1.34	NA	-1.74	-2.10	NA	-2.89	NA
ASTR_ACH-000655	NA	-1.85	NA	NA	-2.07	NA	NA	NA	-1.58	-2.40	NA	-2.87	NA
ASTR_ACH-001016	-1.30	-1.78	NA	NA	-2.14	NA	-1.21	-1.08	-1.64	-1.98	NA	-2.15	NA
ASTR_ACH-001172	NA	-1.31	NA	NA	-1.71	NA	-1.62	-1.18	-1.29	-2.38	NA	-3.00	NA
ASTR_ACH-002269	NA	-1.45	NA	NA	-1.91	NA	-1.57	-1.53	-1.47	-2.04	NA	-2.42	NA
ASTR_ACH-002304	-1.30	-1.57	NA	NA	-1.35	NA	-1.27	-1.31	-1.12	-2.21	NA	-2.83	NA
ODG_ACH-000067	NA	-1.59	NA	NA	-1.81	NA	NA	NA	-1.35	-1.92	NA	-2.41	NA
ODG_ACH-000807	-1.30	-1.73	NA	NA	-1.56	NA	-1.65	-1.19	-1.72	-2.14	NA	-1.99	NA
GB_ACH-000036	NA	-1.86	NA	-1.11	-1.68	NA	-1.23	NA	-1.53	-1.35	NA	-2.78	NA
GB_ACH-000075	NA	-1.74	NA	NA	-2.05	NA	NA	-1.03	-1.46	-2.11	NA	-2.48	NA
GB_ACH-000098	NA	-1.52	NA	NA	-1.97	NA	-1.14	NA	-1.26	-2.00	NA	-2.55	NA
GB_ACH-000137	NA	-1.83	NA	NA	-2.04	NA	-1.33	-1.05	-1.35	-1.96	NA	-2.57	NA
GB_ACH-000152	NA	-1.65	-1.11	NA	-2.18	NA	NA	-1.07	-1.08	-2.32	NA	-2.41	NA
GB_ACH-000200	NA	-1.69	NA	NA	-2.23	NA	-1.24	-1.14	-1.17	-1.87	NA	-2.59	NA
GB_ACH-000215	NA	-2.29	NA	NA	-2.16	NA	-1.49	-1.60	-1.38	-2.25	NA	-2.79	NA
GB_ACH-000231	NA	-2.11	NA	NA	-1.61	NA	-1.61	-1.37	-1.70	-1.92	NA	-2.91	NA
GB_ACH-000244	NA	-1.77	NA	NA	-2.03	NA	-1.01	-1.52	-1.28	-2.02	NA	-1.76	NA
GB_ACH-000269	NA	-1.79	NA	-1.15	-1.92	NA	NA	NA	-1.90	-1.79	NA	-2.17	NA
GB_ACH-000323	NA	-1.54	NA	NA	-2.02	NA	NA	NA	-1.58	-2.19	NA	-2.45	NA
GB_ACH-000368	NA	-1.58	NA	-1.38	-1.97	NA	-1.04	-1.04	-1.23	-2.28	NA	-2.33	NA
GB_ACH-000370	NA	-1.52	NA	NA	-1.86	NA	NA	-1.07	NA	-2.01	NA	-2.58	NA
GB_ACH-000376	NA	-1.55	NA	NA	-1.74	NA	-1.27	-1.32	-1.75	-2.55	NA	-2.22	NA
GB_ACH-000445	NA	-1.45	NA	NA	-1.17	NA	-1.18	NA	-1.54	-2.27	NA	-2.84	NA
GB_ACH-000455	NA	-2.07	NA	NA	-1.71	NA	-1.00	-1.51	-1.84	-1.70	NA	-2.93	NA

GB_ACH-000464	-1.09	-2.13	NA	NA	-2.28	NA	-1.17	NA	-1.71	-2.33	NA	-2.72	NA
GB_ACH-000469	NA	-1.82	NA	NA	-1.63	NA	-1.31	NA	-1.54	-2.15	NA	-2.55	NA
GB_ACH-000479	NA	-1.61	NA	-1.01	-1.69	NA	-1.25	-1.53	-1.34	-2.03	-1.24	-2.36	NA
GB_ACH-000504	-1.13	-1.42	NA	NA	-1.73	NA	NA	NA	NA	-2.16	NA	-2.65	NA
GB_ACH-000558	NA	-1.67	NA	NA	-2.07	NA	NA	-1.27	-2.25	-2.06	NA	-2.61	NA
GB_ACH-000570	NA	-1.55	NA	NA	-1.97	NA	-1.50	-1.14	-1.60	-2.47	NA	-2.65	NA
GB_ACH-000571	NA	-1.84	NA	NA	-2.03	NA	NA	NA	-1.30	-2.18	NA	-2.82	NA
GB_ACH-000595	NA	-2.20	NA	NA	-1.78	NA	-1.41	-1.35	-1.67	-2.31	NA	-2.62	NA
GB_ACH-000609	NA	-1.54	NA	NA	-1.64	NA	-1.11	-1.15	NA	-1.94	NA	-2.65	NA
GB_ACH-000622	NA	-1.27	NA	-1.19	-2.00	NA	NA	NA	NA	-1.92	-1.76	-2.74	NA
GB_ACH-000623	-1.04	-1.10	NA	NA	-1.68	NA	NA	-1.08	-1.84	-2.13	NA	-2.60	NA
GB_ACH-000634	-1.16	-1.58	NA	NA	-1.88	NA	-1.42	-1.59	-1.55	-2.12	NA	-2.80	NA
GB_ACH-000673	NA	-2.24	NA	NA	-2.07	NA	-1.39	-1.10	-1.84	-2.19	NA	-2.54	NA
GB_ACH-000676	NA	-1.59	NA	NA	-1.76	NA	-1.30	-1.39	-1.55	-2.38	NA	-2.93	NA
GB_ACH-000738	NA	-1.73	NA	NA	-2.06	NA	-1.32	-1.17	-1.54	-2.27	NA	-1.64	NA
GB_ACH-000760	NA	-1.56	NA	NA	-1.87	NA	NA	-1.45	-1.69	-1.94	NA	-2.96	NA
GB_ACH-000819	NA	-1.64	NA	NA	-1.79	NA	-1.02	NA	NA	-1.56	NA	-2.21	NA
GB_ACH-000863	-1.16	-1.17	NA	NA	-1.59	NA	-1.45	-1.20	-1.23	-1.77	NA	-2.80	NA
GB_ACH-000887	NA	-1.54	NA	NA	-2.20	NA	-1.65	-1.62	-1.52	-1.64	NA	-2.93	NA
GB_ACH-001329	NA	-2.16	NA	NA	-2.49	NA	NA	NA	NA	-2.19	NA	-2.49	NA
GB_ACH-001605	NA	-1.31	NA	NA	-2.08	NA	NA	NA	NA	-2.19	NA	-1.11	NA
GB_ACH-001606	NA	-1.74	NA	NA	-1.46	NA	NA	NA	NA	-1.91	NA	-2.49	NA
GB_ACH-001608	-1.23	-1.54	NA	NA	-2.25	NA	-1.77	-1.56	-1.83	-2.36	NA	-1.97	NA
GB_ACH-001609	NA	-2.15	NA	NA	-2.09	NA	-1.22	-1.20	-1.76	-2.24	NA	-2.95	NA
GB_ACH-001610	NA	-1.72	NA	NA	-2.00	NA	-1.69	-1.23	-1.86	-1.91	NA	-1.49	NA
GB_ACH-001611	NA	-1.52	NA	NA	-1.56	NA	-1.17	-1.19	-1.53	-2.25	NA	-2.19	NA
GB_ACH-001622	NA	-1.91	NA	NA	-2.06	NA	-1.07	-1.14	-1.04	-2.34	NA	-3.00	NA
GB_ACH-001623	NA	-1.91	-1.30	NA	-2.02	NA	-1.36	-1.14	-1.18	-2.21	NA	-2.96	NA
GB_ACH-001624	NA	-2.07	NA	-1.01	-1.59	NA	-1.39	NA	NA	-1.40	NA	-2.67	NA
GB_ACH-002228	NA	-1.40	NA	NA	-1.57	NA	-1.47	-1.72	-1.35	-2.41	NA	-2.44	NA
GB_ACH-002229	NA	-1.75	NA	NA	-1.81	NA	-1.17	-1.26	-1.45	-1.80	NA	-2.60	NA
GB_ACH-002230	-1.08	-2.28	NA	NA	-2.08	NA	NA	NA	NA	-1.93	NA	-2.60	NA
GB_ACH-002680	NA	-1.98	NA	-1.40	-2.37	NA	-1.38	-1.13	-1.41	-2.43	NA	-2.74	NA

5. DISCUSSION

The gliomas present massive challenges in terms of innovating new treatment strategies despite the efforts put forth by researchers in recent years (Sanders & Debinski, 2020). The co-expression patterns of genes in cancers have become a new trend in identifying the targetable entities (Yi et al, 2020). Therefore, initially we investigated the co-expression profiles and their differences between different subtypes of LGG based on *IDH* mutation status and to explore the common pathways between these subtypes.

Even though these subtypes belong to LGG, the survival rate and the progression of the disease are significantly dissimilar for LGG with *IDH* mutation status (Sharma & Graber, 2020). The LGG with IDH1 mut, pTERT wt and 1p19q-co-deleted genotype has the best chance of survival (Aoki et al, 2018). The overall survival shortenst with only incorporation of pTERT mutations on the aforementioned genotype (Eckel-Passow et al, 2015). Therefore, these genotypic traits are important to consider while exploring either genotypic or transcriptomic profiles of gliomas.

In this study, we explored the discrepancies between transcriptomic profiles of LGG subtypes and incorporated the perspective of co-expression data to identify the shared biological pathways that significantly differ between the normal samples and LGG subtypes, and further looking at the differentiating features from one subtype to another (clinical parameters). Furthermore, the tumors belonging to CNS were previously characterized on the basis of histological parameters; however, with the advances and popularization of NGS and high throughput methodologies, the molecular profiling or characterization of complex disorders has become possible (Varghese et al, 2017).

The genetic subtypes especially in the cases of gliomas such as IDH mutation parameter being a biomarker of the disease, can further assist in understanding the molecular heterogeneity of the cancers by utilizing sequencing data (Carter et al, 2017). It can further be utilized in understanding the oncogenic drivers which can be targeted at the earliest stages to combat the worse prognosis associated with eventual progression to GBM (Lim-Fat et al, 2022). Therefore, we utilized datasets detailing the transcriptomics and DNA methylation events at genomic scale to arrive at specific conclusions in our study regarding the biological pathways, targetable genes with

specific drugs, and how the integration of transcriptomics and DNA methylation can guide in understanding the diverse profiles of LGG subtypes.

5.1. Subtype Specific Co-expression Modules and Their Enrichment Analysis

The clinical features of interest were correlated with different modules from the weighted gene co-expression analysis performed. Our analysis showed that Chr7 gain, which is associated with *EGFR* amplification and Chr10 loss, which is linked with PTEN deletion (Stichel et al, 2018), are not associated with either IDH wt or IDH mut subtypes of LGG. Although these clinical features have been extensively studied and may be clinically relevant, Sienkiewicz et al. found that these features are not enough to be used in diagnosis as well (Sienkiewicz et al, 2022).

The module-trait relationship plots are shown in Figures 4.3a-b. Moreover, for IDH mut subtype, pTERT mutation status as a clinical feature was positively correlated with five co-expression modules in our study and negatively correlated with only two modules (white and GreenYellow module). This feature has been linked with the prognosis of LGG for IDH wt subtype as well (Aoki et al, 2018); therefore, the co-expression modules negatively correlating with IDH mut indicates that this set of genes are inactive in IDH mut, which is less aggressive form of LGG. IDH wt subtype had four negatively correlated and two positively correlated co-expression modules for pTERT mutation status. Here, the modules of interest are positively correlated ones (blue, green). In IDH mut samples, TERTexp.status feature was positively correlated with three modules (Blue, Green and Salmon) and negatively correlated with two modules (White and GreenYellow). Moreover, in IDH wt samples, TERTexp.status was not positively correlated with any module based on our threshold values, but found as negatively correlated with only one module (Turquoise).

Dono et al. identified patients of gliomas with IDH wt subtype and *BRAF* V600E mutation have the potential to be given BRAF-targeted therapies (Dono et al, 2020). Therefore, *BRAF* V600E mutation status was explored in IDH wt subtype for co-expression based analysis and according to our correlation criteria (less than -0.5 or more than 0.5); the co-expression modules were not correlated with IDH wt LGG samples. Braf protein functions in cellular growth and V600E mutation in *BRAF* gene is associated with tumor metastasis for different brain tumors including GBM, ganglioma, and astrocytomas (Sithanandam et al, 1990; Davies et al, 2002; Kaley et

al, 2018). Based on our analysis of *BRAF* mutation status, there are no or few co-expressed targets that can be identified and targeted together if BRAF inhibitors are utilized as intervention therapy for IDH wt LGG patient.

The GBM and oligodendrogliomas represent opposite prognosis, but both types possess pTERT mutations (Aoki et al, 2018); therefore, IDH wt and IDH mut LGGs were explored in our study further with regard to pTERT mutation status for selecting modules of interest (based on cut-off of correlation values). Subsequently, enrichment analysis was performed on such modules for identifying the enriched pathways associated with individual modules based on weighted co-expression, which are shown in Table 4.1.

Immune-related pathways were enriched in IDH wt Salmon and Yellow modules having negative correlation with pTERT status, meanwhile IDH mut Blue module showing positive correlation with pTERT was enriched for immune system related pathways. The tumor associated responses from the immune system for differing oncogenic profiles (Berghoff et al, 2017) along with the heterogeneous tumor microenvironment (Anderson et al, 2021) explains the correlation variance for immune related pathways being regulated differently, which could possibly explain our findings regarding the pTERT mutation statuss.

The synaptic pathways were shared among Green and Blue modules which are positively correlated with pTERT mutation status in IDH wt subtype. Moreover, the GreenYellow module in IDH mut subtype was also enriched for synaptic pathways but was negatively correlated with the status of pTERT mutation. A study exploring DNA methylation status and molecular profiling discovered that IDH wt LGG exhibit the activation of synaptic pathway (Ceccarelli et al, 2016), however, our analysis is the first of its kind which investigated co-expression modules to identify that synaptic pathways are linked with pTERT mutation status in IDH wt LGG.

The IDH wt Blue pTERT positively correlated and IDH mut GreenYellow pTERT negatively correlated modules contained 13 shared pathways as shown in the Table 4.2. The Glutamatergic synapse pathway of both modules contained Glutamate ionotropic receptor NMDA type subunits. NMDA receptor mediated signaling has been associated with invasive nature and proliferation of glioma and glioblastoma cells (Ramaswamy et al, 2014; Müller-Längle et al, 2019). However, Glutamate ionotropic receptor AMPA type 1 (GRA1) was a component of IDH wt blue models and was absent in GreenYellow module, which is known to facilitate the growth and

cytoskeletal regulation in glioma cells (Wirsching & Weller et al, 2020). This data indicates that in IDH wt pTERT mutant-correlated LGG, the activity of NMDA type receptors play a role in oncogenic transformation. However, in IDH mut pTERT mutant negatively correlated LGG, *GRA1* mediated signaling is activated. This data can assist in differentiating the molecular subtypes of LGG further at the transcriptomics level.

5.2. DEGS from Module of Interest

After exploring the clinically relevant modules and their pathways along with the parameters of IDH mutation and pTERT mutation, we further investigated the differentially expressed genes (DEGs) between normal samples and LGG IDH wt or between normal and IDH mut samples; then mapped these genes on the relevant final module of interest (IDH wt Blue). Since the IDH wt Blue module was correlated with pTERT mut status in both IDH wt and IDH mut samples, this type of analysis assists in understanding the transcriptionally active or inactive genes correlating with the clinical features.

After mapping the DEGs, we observed that *GNG12* was upregulated and *PLCB1*, *GRIA2*, *GABRA3*, and *GNAL* were downregulated in the IDH wt Blue module, furthermore these were absent in the IDH mut GreenYellow module. Interestingly, the downregulated genes were also downregulated in our Normal vs IDH wt and Normal vs IDH mut LGG analyses. Therefore, we postulate that these are the hub genes based on co-expression analysis and relevant with IDH and pTERT mutation status in LGG samples.

GNG12 has been associated with metastases, differentiation and cell division along with other tumor promoting pathways and belongs to the G-family of proteins (Morishita et al, 1995; Luo et al, 2018). Liu et al. has recently identified *GNG12* as not only a novel biomarker for glioma but *GNG12* can also be utilized in targeted therapy as well (Liu et al, 2022). Our analysis also revealed that *GNG12* expression was correlated with severity of LGG indicated with lowest expression in normal samples and highest in IDH wt LGG. This finding is significant from the perspective of this independent study as our data can be utilized for clinical relevancy as it is validated from other investigations.

A G-protein alpha subunit stimulatory protein *GNAL* was found downregulated among our DEGs, which has already been identified as one of the 24 genes linked with prognosis of glioma with an inverse correlation of its expression with glioma grades (Zhang et al, 2020). In our analysis, *GNAL* expression was highest in normal samples and downregulated in IDH mut with lowest expression in LGG IDH wt samples. This data further validates our findings. Another downregulated gene *GABRA3* (Gamma-amino butyric acid receptor alpha subunit 3) has also been identified in an independent study as downregulated in glioma, where its loss has been associated with invasion and metastasis of glioma (Patil et al, 2020). Patil et al utilized the Cancer Cell Line Encyclopedia glioma data and TCGA data in their analyses, further implying that our study produced validated results.

Glutamate ionotropic receptor AMPA type subunit 2 (*GRIA2*) has been studied extensively in glioma and glioblastoma cells with invasive behavior of cells linked with higher expression levels of *GRIA2* (Oakes et al, 2017; Zhang et al, 2018; Hu et al, 2020). However, AMPA receptor expression (GRIA1-4) was negatively correlated with glioma grade in one study, with RNA editing of GRIA2 depending on the glutamate environment of the cells (van Vuurden et al, 2009). In our study, we found *GRIA2* to be among the downregulated cohort of genes, which is interesting. Moreover, *GRIA2* being enriched in Blue module only and not in GreenYellow, indicates that there is a difference in how *GRIA2* is regulated among different subtypes of LGG. AMPA receptors are involved in regulation of calcium homeostasis via activation of the Akt pathway (Ishiuchi et al, 2007).

The inhibition of the Akt pathway in combination with ERK has been identified as a synergistic and potent inhibitor of glioma (Wu et al, 2017). The glutamatergic synapse pathways are involved in ERK regulation via PKC and Phospholipase C beta 1 (*PLCB1*). We identified *PLCB1* as downregulated in IDH wt LGG as compared to IDH mut LGG and normal samples. Moreover, *PLCB1* was also discovered as downregulated in glioblastoma as compared to normal samples recently (Marvi et al, 2022). The low expression of *PLCB1* has been significantly associated with poor survival of GBM (Cai et al, 2022). Therefore, the gene signature of downregulated *GNAL, GABRA3, GRIA2, PLCB1* and upregulated *GNG12* presents a unique opportunity as a result of this study for researchers to explore in the biomarker, diagnosis, prognosis, and therapeutic intervention of LGG based on its molecular subtypes.

5.3. Interaction of Identified Gene Signature with Known Drugs

After identifying the gene signature associated with the progression of LGG, we investigated the drugs that can be used to target this signature for best utilization with the aim of inactivating the upregulated *GNG12* and activating the downregulated *PLCB1, GRIA1, GABRA3,* and *GNAL*. The comprehensive analysis is shown in Table 4.4. We identified four drugs namely Vemurafenib, Vanadium Pentoxide, Imatinib, and Cisplatin, which targeted four out of five genes in our gene signature.

Temozolomide and Cisplatin have been investigated for combined treatment against recurrent GBM with acceptable toxicity and active performance (Wang et al, 2017). Temozolomide is used against glioma, however the treatment leads to resistance and eventual recurrence (Daniel et al, 2019). Cisplatin has been used with other drugs for incurring synergistic effects (Macieja et al, 2019; Zhai et al, 2021) but not with other drugs targeting our gene signature such as Imatinib and Vanadium Pentoxide. Moreover, Imatinib is known to elicit autophagy and subsequent cytotoxicity in glioma cells (Shingu et al, 2009) and has been used for investigating its synergistic potential with other drugs (Lu et al, 2020) and also against GBM (Rashidi et al, 2020). However, the combination of Imatinib with other drugs targeting our gene signature has not been investigated and provides an opportunity to researchers to explore this new targeted approach.

Vanadium Pentoxide has not been investigated to date directly in treating glioma, but its effect on cancer cell lines and animal models has been studied (García-Rodríguez et al, 2016; Zwolak, 2014). Furthermore, Vanadium-based nanoparticles have been explored in cytotoxicity and drug encapsulation studies with modest results (Guo et al, 2020), however, the direct use of Vanadium Penoxide is yet to be investigated. Vemurafenib is known to target the BRAFV600E in pediatric glioma (Bautista et al, 2014) and LGG (Del Bufalo et al, 2018) in a Phase I clinical study (Nicolaides et al, 2020) with high grade glioma leading to resistance (Lehmann et al, 2022). Vemurafenib was combined with Cobimetinib in a study against GBM with positive outcome in a recent study (Rajan et al, 2023). It was also combined with Cobimetinib, and Dabrafenib and Trametinib in patient derived pediatric LGG cells (Usta et al, 2020). However, the combination of Vemurafernib in similar models or

patients with aforementioned identified drugs targeting our gene signature is yet to be performed.

Although imatinib shows poor penetration through the blood brain barrier (BBB) (Takayama et. al., 2002), Bihorel et al. (2007) demonstrated that combining imatinib with P-gp (and Bcrp1) transporter inhibitors like elacridar may increase imatinib's brain delivery and increase its ability to treat malignant gliomas. In addition, Durmus et al. (2012) showed that brain penetration of Vemurafenib can be enhanced with elacridar coadministration. Vanadium (V) crosses the BBB (Avila-Costa et al. 2015). Dorado-Martínez et al. showed that Vanadium Pentoxide (V_2O_5) induced death in Alzheimer-like cells in rats through cytoskeletal and synaptic alterations. According to pharmacokinetic analyses in nonhuman primates, cisplatin and carboplatin only partially penetrate the CNS (3.7% and 2.6%, respectively) (Jacobs et. al. 2005). Zhang et. al. (2017) has shown that cisplatin-loaded nanoparticles (70 nm in diameter) penetrate deeper into the brain with local administration by either manual injection or convection-enhanced delivery (CED). The drugs prioritized in this study might have poor or limited BBB penetration, which is one of the grand challenges in glioma treatment, but drug delivery systems such as nanoparticles, convection-enhanced delivery, focused ultrasound, and disruption of the BBB using microbubbles might be helpful to combat this issue in the future applications (Arvanitis et. al. 2020, Wen et. al., 2020, Drappatz et. al., 2013, Idbaih et. al., 2019).

5.4. Methylation Data Integration with Transcriptomics Data

After identifying the gene signature based on co-expression analysis and exploring DEGs among different subtypes of LGG and drugs that can target the gene signature comprehensively, we aimed to investigate the alterations occurring at DNA methylation levels among different subtypes of LGG. Since DNA methylation is directly involved in regulating the gene expression and co-expressed genes can be co-regulated with common epigenetic factors functioning in DNA methylation profiles (Yang et al, 2017), this analysis can be integrated into our processed transcriptomics data as well.

Before processing the DNA methylation data, we explored more normal samples to get enriched data to be compared with cancer samples in our study. In the previous analysis exploring the co-expression data, we had 5 samples belonging to the

normal group for analyzing the differentially expressed genes and then mapped them onto the genes contained in specific modules. Therefore, we repeated the DEGs with normalized samples from the GTEx database and included them in our analysis compared with TCGA LGG samples (Table 4.5). In short, we analyzed the 4 subtypes 1) IDH mut pTERTmt Chr1p19q codel 2) IDH mut pTERTwt Chr1p19q noncodel, 3) IDH wt pTERTmt Chr1p19q noncodel, 4) IDH wt pTERTwt Chr1p19q noncodel. After conducting the DEG analysis, we explored the differentially methylated regions, which were later explored to arrive at the hypo- or hyper-methylated genes.

Each subtype explored in our study showed a subset of DEGs, but the most aggressive subtype of LGG with IDH wt pTERT mut showed the most downregulated genes. This data is consistent with our previous DEG analysis performed with normal samples from TCGA GBM data. In our differentially methylated genes analysis, we observed that IDH mut subtypes showed a significantly higher proportion of hypermethylated genes, which has been studied in literature as well (Unruh et al, 2019; Braun et al, 2021). The CpG island hypermethylation is associated with epigenetic alterations occurring at genomic level due to IDH mutation has already been reported (Noushmehr et al, 2010).

The number of hypomethylated genes observed in IDH wt pTERT mut subtype as compared to normal samples was significantly higher as compared to other subtypes when compared with normal samples (Table 4.6). Genome instability has been linked with DNA hypomethylation (Ehrlich & Lacey, 2012), which can result in more aggressive phenotype in cancer cells (Yao & Dai, 2014). The role of DNA hypomethylation has been reviewed recently in the context of glioma where the emphasis on chromatin regulation was made, in order to identify the heterogeneity associated with different subtypes of gliomas (Dabrowski & Wojtas, 2019). Furthermore, the Akt signaling pathway has also been associated with DNA hypomethylation in glioma (Briand et al, 2019), which we discussed in our analysis of gene signature being involved in the intersection with Akt signaling. The detailed analysis of differentially methylated genes are shown in Figure 4.10 showcasing the subtype specific genes driven through differential methylation status.

These subtype specific MDGs were enriched for biological pathways (Table 4.8). The subtype specific or shared MDGs for the explored subtypes showed enrichment for CNS specific pathways or major signaling pathways we identified in our previous analysis for co-expression modules. However, interestingly, the

hypomethylated and overexpressed genes for IDH wt pTERT mut subtype which is the most aggressive subtype, were enriched for immune-related pathways. The pTERT alteration and its associated effects has been associated with the tumor microenvironment and changed signaling referring to immune evasion for gliomas and GBM (Olympios et al, 2021).

Moreover, the hypermethylated and downregulated genes for IDH wt pTERT mut were enriched for neuroactive ligand-receptor interaction pathways, implying that the aggressive oncogenic transformation in glioma is associated with over activation of ligand-receptor mediated signaling. Recently, MD2 was observed as a biomarker for immune infiltration for gliomas, with MD2 mediated neuroactive ligand-receptor interactions for T cell co-stimulation as a differentiating factor (Zhao et al, 2022). Therefore, our analyses revealed that the most aggressive subtype of LGG based on IDH and pTERT mutation status (IDH wt pTERT mut) has differential response to immune related pathway and neuroactive ligand receptor interaction based pathways, as an underlying feature at both transcriptome and DNA methylation level, leading to tumor heterogeneity and consequent aggressiveness and worse prognosis.

5.5. Network Construction based on MDGs + Modules of Interest and Essential Epifactor Analysis

We further explored our WGCNA-based co-expression analysis in the context of DMGs to investigate the co-expressed and co-methylated cohorts.The MDGs were mapped onto the co-expressed cohorts (modules) as shown in Figure 4.10a-b. The mapped IDH wt pTERT+ Blue module genes were mostly co-hypomethylated and IDH mut pTERT- GreenYellow module genes were co-hypomethylated. This data supports the hypothesis that the DNA methylation profiles can serve as a differentiating factor for further characterizing the LGG subtypes that are already based on genetic alterations. The subnetworks obtained as a result of our analyses inform us that the DNA methylation has a role in regulation of the genes that are co-expressed and can be targeted for specific subtypes.

Since, the aforementioned subnetworks associated with modules of interest are specific with MDGs incorporated onto co-expressed genes, we further explored their neighbors to construct a more biologically interpretable network containing nodes that can be targeted. Therefore, we included epigenetic factor protein complexes (also

known as epifactors) in our analysis to illustrate the comprehensible subtype module or subtype specific network. The number of complexes with nodes and edges for subtype specific networks and their statistics are displayed in Table 4.9. This analysis resulted in targetable entities which are involved in regulation of both DNA methylation and also the co-expressed genes extracted from transcriptomics data which are associated with specific subtypes of LGG based on IDH mutation status and its correlation with pTERT mutation status.

The IDH mut pTERT- subtype (least aggressive subtype among the ones investigated) co-expression module of interest with incorporated MDGs showed targetable Epifactors such as *HDAC1, BRCA1, CTBP1, PRMT5, PPP4C, and CSNK2A1* (Figure 4.12b). Furthermore, the IDH wt pTERT+ subtype (most aggressive subtype among the ones investigated) co-expression module of interest with incorporated MDGs showed more targetable Epifactors such as *CBX3, PRMT5, KDM1A, HDAC1, CTBP1, CSNK2A1, ACTL60, EPC1, BRCA1, ACTB, HCFC1, ACTL6A*, and *MIS18A* among others. Recently, in high grade gliomas, the Epifactors were screened and identified using CRISPR screening method, and targeted to illustrate the importance of these factors in targeting the brain tumors (Wenger et al, 2023), however, for LGG and adult GBM, such an approach has not been utilized. As a consequence of our study, we propose that the Epifactors can be identified as potential targets for LGG as these factors regulate the genes expressed together in a subtype specific manner.

Lastly, we performed perturbation analysis on the Epifactors that are common for all the subtypes we explored in our analysis for LGG. Thirteen Epifactors were screened for perturbation effects in 66 models of diffuse glioma. We identified a 5-Epifactor signature (*ACTL6A*, *HCFC1*, *RUVBL1*, *PRMT5*, and *WEE1*), which can be considered essential for cell survival (Table 4.10). Interestingly, *PRMT5* was shared across all subtypes of LGG used in all analyses across our study with *WEE1* common for only IDH wt subtypes regardless of pTERT mutation status. PRMT5 is a protein arginine methyltransferase which acts in the methylation of both histones and non-histone proteins for regulating the gene expression (Pollack et al, 2009; Branscombe et al, 2001). It also affects the stability, recruitment, and activity of different transcription factors (Koh et al, 2015). *PRMT5* inhibition has been investigated as a potential therapeutic intervention in glioblastoma with positive outcomes in targeting cancer stemness (Yan et al, 2014; Banasavadi-Siddegowda et al, 2018; Sachamitr et

al, 2021). In glioma, the expression of PRMT5 has been associated with malignancy of the tumor (Han et al, 2014) and its interaction with HOXC10 leading to upregulation of *VEGFA* as a mechanism of inducing malignant behavior (Tan et al, 2018). Moreover, in glioblastoma, the inhibition of PRMT5 can lead to increased sensitivity to mTOR inhibition for targeting cancer cells (Holmes et al, 2019). However, there is a dearth of literature concerning the mode of action of *PRMT5* in gliomas, therefore our study revealing *PRMT5* as a common essential Epifactor observed in all clinically relevant co-expression modules of LGG subtypes observed, makes it a prime target that can be utilized in therapeutic strategies.

Wee1 protein, encoded by the *WEE1* gene, functions in the regulation of G2/M checkpoint of the cell cycle, and is overexpressed in many cancers and considered a potential target for therapy with many clinical trials underway (Vakili-Samiani et al, 2022). In glioblastoma, both genetic and pharmacological inhibition of *WEE1* has been associated with sensitivity of cancer cells to ionizing radiation leading to mitotic catastrophe, further implying that *WEE1* is a potent target for brain tumors (Mir et al, 2010). WEE1 is overexpressed in many cancers including glioblastoma (De Witt Hamer et al, 2011). Caretti et al in 2013 showed that WEE1 inhibition using MK-1775 leads to enhanced response to radiation in the xenograft model of diffuse intrinsic pontine glioma (Caretti et al, 2013). In high Grade gliomas, a specific inhibitor of *WEE1* MK-1775 in combination with radiation treatment was proposed as a targeted therapy option with promising results (Mueller et al, 2014). MK-1775 has been used in combination with Temozolomide to study the heterogeneous distribution across the blood brain barrier, which indicated the limited distribution to brain tumors (Pokorny et al, 2015; Lescarbeau et al, 2016). Therefore, other ways of delivery may be explored to utilize this compound in clinical settings against glioblastoma. In glioblastoma, *WEE1* is considered a prognostic marker and in glioma, its expression was correlated with *MGMT* status (Music et al, 2016). Adavosertib, a WEE1 inhibitor in synergy with cranial radiation therapy (CRT) has been utilized in a Phase I consortium study against diffuse intrinsic pontine glioma (Mueller et al, 2022). An integrated analysis using only DNA Damage and Repair (DDR) related genes in DEG analysis for LGG showed that *WEE1* is associated with IDH mutation status (Pang et al, 2020). Our independent study initiating from clinically relevant co-expression modules based on IDH and pTERT mutation status and integration with DNA methylation profiles with

subsequent Epifactor analysis resulted in identifying *WEE1* as a IDH wt specific Epifactor complex, which can be targeted for LGG therapy.

6. CONCLUSION

Gliomas present challenges as they subsequently progress from lower grade to higher grade forms of the disease with worse prognosis for patients. With high-throughput data and analysis techniques available to researchers, the molecular signatures of the disease can be deciphered with the goal of finding pathways or targets that can be readily targeted for incorporation in treatment strategies. Therefore, we designed a bioinformatics pipeline to investigate the clinical data available from the relevant databases in order to explore the common pathways shared between different molecular subtypes of LGG. We further identified the enriched pathways associated with the co-expression modules and identified the drugs that can target a cohort of targets based on our investigation. We further implemented DNA methylation analysis to discover the altered methylation profiles associated with different subtypes of LGG and integrated our analysis with transcriptomics data. This deep analysis exercise was further enhanced by incorporating the epigenetics-based interactome analysis to find the potential Epifactors enriched in a subtype specific manner and can be used as direct targets in intervention strategies. Our initial findings based on co-expression modules of LGG subtypes were already established in the relevant literature and suggested that co-expression based modules associated with clinical subtypes of IDH mutation status and pTERT mutation status can differentiate the subtypes based on co-expressed genes, which are enriched in synapse-related or immune-related pathways. We further arrived at a gene signature of upregulated *GNG12* and downregulated *PLCB1*, *GRIA2*, *GABRA3,* and *GNAL* after mapping DEGs on our co-expression modules constructed using WGCNA. Based on our analysis, we identified four drugs Vemurafenib, Vanadium Pentoxide, Imatinib, and Cisplatin, that can target four out of five genes in our gene signature simultaneously. These findings have direct implications in developing clinical and pre-clinical intervention strategies for aggressive LGG IDH wt pTERT mut patients having the worse prognosis.

We further explored the LGG samples and compared the DNA methylation profiles of normal samples incorporated from different datasets with appropriate normalization techniques. We integrated the hyper and hypo-methylated genes with differentially expressed genes from different subtypes of LGG and observed a higher number of hypomethylated genes in IDH wt pTERT mut subtype as compared to normal samples. To arrive at biologically interpretable insights from DNA methylation

integration with co-expression modules, we found the co-expressed and co-methylated cohorts and explored their neighbors to construct the networks associated with different genetic subtypes of LGG. The epigenetic factors of the individual networks were then investigated to find the targetable entities for each subtype. This deep analysis converged on two targets which can be utilized in further studies in combination with aforementioned drugs to aggressively treat the LGG tumors. *PRMT5* was found to be enriched in all subtypes as a methylator factor, and therefore can lead to a broad target for brain tumors and has already been explored in clinical studies. Interestingly, WEE1 was discovered as a LGG IDH wt specific target in our analysis, and its inhibitors has been explored against high grade gliomas, but LGGs with IDH wt having worse prognosis can be targeted using *WEE1* inhibitors as well. Therefore, this multi-omics approach of integrating transcriptomics and DNA methylation high throughput data provided us with a narrow gene signature that can be targeted with already established drugs and also unique targets which can be targeted in combination with established drugs against different types of LGG or GBM.