Fall 12-15-2017

Comparative Molecular Characterization of Typical and Exceptional Responders in Glioblastoma

Kristin Wipfler

COMPARATIVE MOLECULAR CHARACTERIZATION OF TYPICAL AND EXCEPTIONAL RESPONDERS IN GLIOBLASTOMA

by

Kristin Wipfler

A DISSERTATION

Presented to the Faculty of
the University of Nebraska Graduate College
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Genetics, Cell Biology & Anatomy
Graduate Program

Under the Supervision of Professor Chittibabu Guda

University of Nebraska Medical Center
Omaha, Nebraska

September 2017

Supervisory Committee:
Michele R. Aizenberg, M.D.
James Eudy, Ph.D.
Howard S. Fox, M.D., Ph.D.
Fang Yu, Ph.D.

DEDICATION

To Donnie: Thank you for your unwavering love and support and for being my partner in everything. We accomplished this together.

And to our daughter Rowan: This is for you, Noodle. I love you more than I ever thought one person could possibly love another.

ACKNOWLEDGEMENTS

First, I would like to thank Dr. Guda for his support, encouragement, and guidance throughout my graduate career. I would also like to thank the other members of my supervisory committee, Dr. Aizenberg, Dr. Eudy, Dr. Fox, and Dr. Yu, for their suggestions and insightful questions.

I would like to express my gratitude to Dr. Andrew Dudley and Dr. Karen Gould for their contributions to graduate education at UNMC and for being tireless advocates for the students here. I would also like to thank all the other educators, from elementary school through graduate school, who have inspired me and fostered my dedication to science and to learning.

I would also like to thank Dr. Dele Davies and everyone in the Graduate Studies Office for supporting me in my endeavors as a student and as GSA President. Your support has helped me achieve some of my proudest accomplishments during my time at UNMC.

Thank you to all of the current and former members of the Guda lab, especially Adam, whose support, contributions to this project, and friendship have been invaluable to me.

Finally, I would like to thank my family and friends. I have been extremely fortunate and privileged to have you all as my support system and village. I cannot take all the credit for this accomplishment because I could not have reached this point without all of you. You each own a piece of this work. Thank you.

COMPARATIVE MOLECULAR CHARACTERIZATION OF TYPICAL AND EXCEPTIONAL RESPONDERS IN GLIOBLASTOMA

Kristin Wipfler, Ph.D.

University of Nebraska, 2017

Supervisor: Chittibabu Guda, Ph.D.

Glioblastoma (GBM) is the most common and the deadliest type of primary brain tumor, with a median survival time of only 15 months despite aggressive treatment. Although most patients have an extremely poor prognosis, a small number of patients survive far beyond the median survival time. Investigation of these "exceptional responders" has sparked a great deal of interest and is becoming an important focus in the field of cancer research. To investigate the molecular differences between typical and exceptional responders in GBM, comparative analyses of copy number, methylation, gene expression, miRNA expression, and protein expression data sets from The Cancer Genome Atlas were performed, and the results of these analyses were integrated via correlation studies and pathway analyses to assess the functional significance of the differential aberrations. Typical responders are characterized by upregulation of NF-κB signaling and of pro-inflammatory cytokines and their associated pathways, while exceptional responders are characterized by upregulation of Alzheimer's and Parkinson's disease pathways, as well as of genes involved in synaptic transmission and plasticity. The upregulated pathways and processes in typical responders are consistently associated with more aggressive tumor phenotypes, while those in the exceptional responders suggest a retained ability in tumor cells to undergo cell death.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GBM	Glioblastoma
TCGA	The Cancer Genome Atlas
NCI	National Cancer Institute
GDC	Genomic Data Commons
ERI	Exceptional Responders Initiative
CGH	Comparative Genomic Hybridization
SNP	Single Nucleotide Polymorphism
FDR	False Discovery Rate
KS	Kolmogorov-Smirnov
GCRMA	Guanine Cytosine Robust Multi-Array Analysis
RMA	Robust Multi-Array Analysis
Log_2FC	Log_2 Fold Change
GSEA	Gene Set Enrichment Analysis
EIEE	Early Infantile Epileptic Encephalopathy
miRNA	microRNA
RPPA	Reverse Phase Protein Array
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes

INTRODUCTION

Glioblastoma

Glioblastoma (GBM) is the most common and the deadliest type of primary brain tumor. GBM is highly malignant and nearly uniformly fatal, with a median survival time of only approximately 15 months despite aggressive treatment, including surgical resection followed by concurrent radiation and chemotherapy with temozolomide.[1–3] GBM tumors are particularly aggressive due to their high degree of heterogeneity and tentacle-like projections that infiltrate surrounding brain tissue, making them extremely difficult to fully excise.[4,5]

The central nervous system is comprised of neurons and glia (including astrocytes, oligodendrocytes, and microglia).[6] GBM arises from astrocytes, star-shaped glial cells that play a variety of diverse roles in the central nervous system, including maintenance of homeostasis, regulation of blood flow, and synaptic transmission.[5,7] GBM usually arises in the cerebral hemispheres, but can be found anywhere in the brain or spinal cord. Most cases occur sporadically, without genetic predisposition. The only known risk factors are some specific genetic diseases (neurofibromatosis, tuberous sclerosis, Li-Fraumeni syndrome, retinoblastoma, and Turcot syndrome) and some environmental exposures (ionizing radiation, vinyl chloride, pesticides, smoking, petroleum refining, and synthetic rubber manufacturing). The most common symptom is a progressive neurological deficit resulting in personality changes or memory loss, but headaches and seizures may occur

as well. The incidence of GBM is higher in men than in women and presents at a median age of 64 years.[2]

There are three pathways that are consistently dysregulated in GBM: the p53 pathway, the receptor tyrosine kinase/Ras/phosphoinositide 3-kinase signaling pathway, and the retinoblastoma pathway. Other common alterations include overexpression of EGFR (epidermal growth factor receptor), mutations in PTEN (phosphatase and tensin homolog), and loss of chromosome 10.[2] Treatment options targeting these genes and pathways have been explored, primarily anti-EGFR agents, but their efficacy is limited by drug resistance.[8]

The Cancer Genome Atlas

The Cancer Genome Atlas (TCGA) was a project led by the National Cancer Institute (NCI) and the National Human Genome Research Institute that began in 2005. It was a database that contained genomic data obtained from a variety of high-throughput genome analysis techniques for 33 different cancer types. Data types investigated in the TCGA project included gene expression profiling, copy number variation, SNP genotyping, DNA methylation profiling, and many more. The primary goal of the project was to demonstrate that genomic data from a variety of sources could be integrated and utilized to identify statistically and biologically significant alterations in cancer.[9,10]

The TCGA project has now concluded, but the more than two petabytes of genomic data generated in the project have been made publicly available through the Genomic Data Commons (GDC). This massive amount of data provides a unique

opportunity to analyze a variety of data types for a large number of cancer patients. By the end of the TCGA project, the GBM dataset included data for 528 patients. The different data types included in the GBM dataset were clinical information, gene expression, exon expression, miRNA expression, copy number arrays, methylation arrays, SNP arrays, trace files, somatic mutations, protein expression, and RNAseq (Table 1).[10–12]

Exceptional Responders Initiative

With the end of the TCGA project, NCI is now developing multiple new genomics databases, one of which is the Exceptional Responders Initiative (ERI). The goal of this project is to identify molecular features that predict whether or not a particular drug or class of drugs will help patients live longer. In many cases, a treatment is deemed unsuccessful after a clinical trial, but 10% or fewer of the patients still have a favorable response. The ERI project intends to identify markers that predict positive responses in such cases. The database will include patients that receive standard treatments as well, not just patients in clinical trials.[13,14]

The idea for the ERI came about based on the concept of exceptional responders, patients who have a unique response to treatments that are not effective for most other patients. The exact definition of "exceptional" varies by specific disease, stage, and treatment. In general, exceptional responders achieve a complete or partial response that only up to 10% of patients experience, and they sustain that response for a much longer duration than the median response.[13–15]

Data Type	*N* (full dataset)	*N* (subset)
Clinical	499	75
Copy Number (SNP array)	493	72
Methylation (27k)	287	32
Gene Expression (DNA microarray)	440	67
miRNA Expression	474	72
Protein Expression	210	20

Table 1: Number of patients in each data type in the TCGA GBM dataset. Of the data types available for GBM, seven were analyzed, six of which are discussed in this dissertation. The number of patients for each of these data types is listed for the full dataset (499 Total) as well as for the dataset analyzed in this work (75 total). The selection of this subset of patients is discussed in Chapter 1. The exon expression and copy number array data types were excluded due to redundancy with gene expression and SNP arrays, respectively. RNAseq was excluded because the available data files were highly processed with a methodology that is not preferred, while the gene expression data type was available in a raw format. Trace files were excluded as they are outdated, having a been replaced by GAM files. When data were available from multiple platforms for a single data type, the most comprehensive option was chosen for analysis whenever possible.

Investigating these exceptional responders has become an important focus for the future of cancer research. There is a great deal of interest in studying these rare patients to learn how to improve therapies for patients who have a more typical response. Several studies of exceptional responders have already been published, and they have helped uncover molecular alterations and mechanisms of resistance. With the huge amount of interest and funding being directed at this topic, it is expected that studies of exceptional responders will be a major focus of cancer research in the near future.[16–19]

Hypothesis

Analyzing and integrating the information from the variety of next generation sequencing and array-based data available in TCGA (now the GDC) for typical and exceptional responders will reveal aberrations that produce more aggressive tumors in typical responders as well as protective effects in exceptional responders. This will provide a clearer picture of the molecular basis of GBM and also reveal possible therapeutic targets and markers for a positive or negative response to standard therapy.

CHAPTER 1

SURVIVAL ANALYSIS AND DEFINING THE RESPONSE GROUPS

Introduction

With current therapies, the median survival time for GBM patients is approximately 15 months. Although most patients have an extremely poor outcome, a small number of patients survive far beyond the median survival time.[1] Survival analyses in the current literature often have a very small sample size and look specifically at a very small set of genes, such as IDH1 and MGMT or EGFR and TP53.[20–23] These studies often have arbitrarily chosen survival time cutoffs, typically >36 months for long survival and <36 months for short survival. These cutoffs are not appropriate for the study of survival outcomes in GBM, as the resulting short survival groups would include many patients who survive well beyond the median survival time. Defining survival groups to compare based on specific characteristics of GBM and the survival curve of TCGA GBM patients would be a vast improvement over the commonly used arbitrary methods described above.

In addition to an improved method of defining survival groups, a main focus of this study is an investigation of so-called "exceptional responders" in GBM. NCI researchers conducting the ERI study define exceptional responders as "patients who have dramatic and long-lasting responses to treatments for cancer that were not effective for most similar patients." The precise definition of exceptional is specific to the disease, stage, and treatment.[13–15] For GBM, defining what constitutes an exceptional response may

be more straightforward than for most cancers, since all primary GBM tumors are classified as grade IV and most patients receive the same standard treatment of tumor resection followed by radiation and chemotherapy with temozolomide. The response to this treatment is also very consistent, with most patients surviving very close to the median survival time and a relatively small number of patients surviving a substantially longer time.[1–3]

Due to the shortcomings in the methodology of defining survival groups in current GBM studies, there is a need for a GBM survival analysis utilizing cutoff parameters specific to characteristics of this disease. With the recent shift of focus to exceptional responders in cancer research, it is also important to incorporate this concept into survival studies in cancer, particularly in cancers like GBM, where most patients respond poorly to treatment but an exceptional few respond very positively. This study addresses both of these needs, utilizing an improved method of defining survival groups guided by the concept of exceptional responders.

Methods

Inclusion criteria were applied utilizing clinical information contained within the TCGA Biotab files for GBM. Only untreated primary GBM samples from patients with known survival times were included in the survival analysis. A Kaplan-Meier survival curve was generated based on the survival times of the patients remaining after the application of the inclusion criteria. The top 10% of patients with the longest survival times were designated as “exceptional responders.” The 10% cutoff was chosen based on

the survival curve and loosely based on the ERI definitions of exceptional responders.[15] The median 10% of patients were classified as "typical responders" in order to have comparable sample sizes between the two groups. Linear regression models were generated with XLSTAT to investigate possible confounding variables that may influence survival, including sex, race, ethnicity, diagnosis method, age, and Karnofsky score. The term with the highest non-significant p-value was removed and the model was regenerated until the overall model and each term were significant ($p < 0.05$). An age cutoff was applied and linear regression models were generated again using the reduced number of samples in order to identify any remaining confounding variables.

Results

After the application of the inclusion criteria, 408 patients remained in the dataset. The Kaplan-Meier survival curve (Figure 1) for those 408 patients shows a steep drop in the first two years, with the survival time for the vast majority of patients within one year of the median 345 days. The curve levels off between two and three years, and a relatively small number of patients survive beyond that time. The patients within that range are in roughly the top 10% for survival time, which was the defining factor for the exceptional responders group.

Age and sex were determined to be confounding variables in the linear regression models. An age cutoff of ≥30 years was applied, which reduced the exceptional responders group by five patients and corrected for the confounding variable of age. Ethnicity was the same for all patients in this group (not Hispanic or Latino) and Karnofsky score, age,

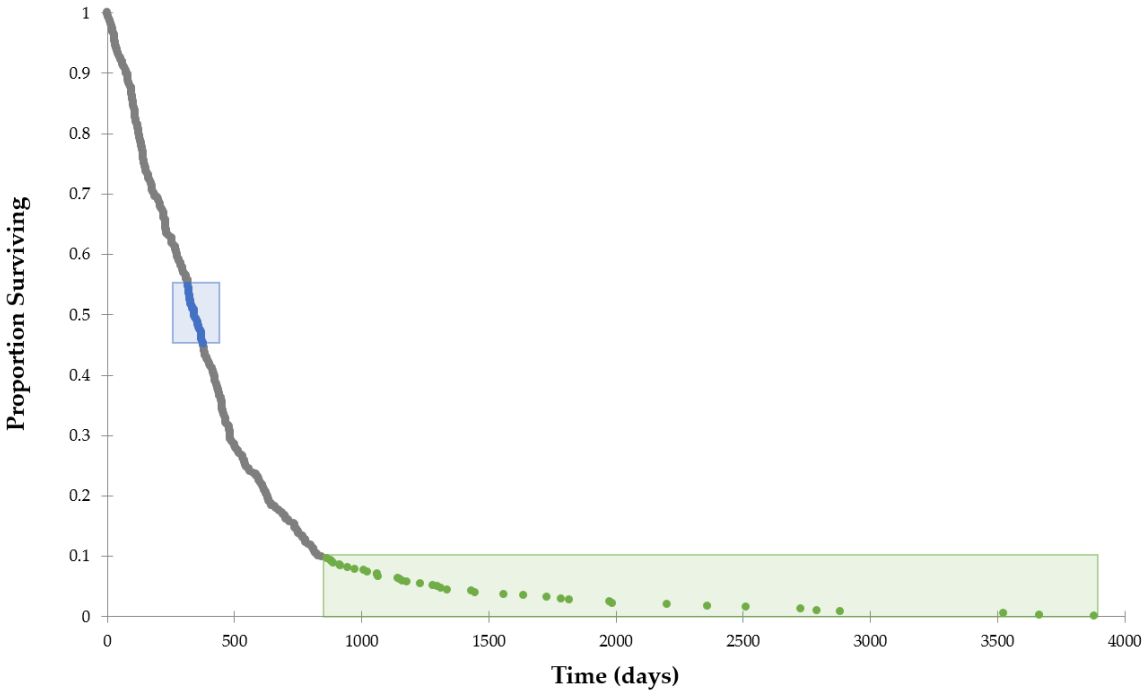

Figure 1: Survival curve for TCGA GBM dataset. This curve includes the 408 TCGA GBM patients that met the inclusion criteria. The curve is characterized by a steep drop off centered around the median of 345 days, with a small number of patients surviving beyond approximately 2.5 years. Typical responders are labeled in blue and exceptional responders are labeled in green.

race, and diagnosis method were not significant predictors of survival time. However, sex was predictive of outcome, with female patients enriched in the exceptional responders group (regression model p=0.021, chi-squared test p=0.034). Sex is only associated with survival in the typical and exceptional response groups, not in the full dataset of 408 patients.

The final dataset included 40 typical responders and 35 exceptional responders (Table 2). Males are more highly represented in the typical response group, and the exceptional responders tend to be a bit younger with a mean age of 49.8 years compared to the typical responders' mean age of 58.7 years. However, this age difference is not statistically significant. The median survival for the typical group is the same as the full dataset (345 days) with a range of 320-378 days. Median survival for the exceptional group is 1282 days (approximately 3.5 years) with a range of 864-3881 days (approximately 2.4-10.6 years).

Discussion

Although the current median survival time for GBM is approximately 15 months, the 12 month median in the full dataset and the typical responders group is consistent with the time period in which most of these samples were obtained.[24] The characteristics of the survival curve are as expected based on previous GBM survival studies.

While the confounding variable of age was addressed with an age cutoff, the only way to fully address the confounding variable of sex is to completely exclude either males or females from the study. Rather than sacrificing such a large number of samples and

	All	Typical Responders	Exceptional Responders
Patients (n)	408	40	35
Male	253	29	17
Female	155	11	18
Age (years)			
Mean	58.7	58.0	49.8
Range	10 - 89	33 - 73	30 - 74
Survival (days)			
Mean	459	347	1600
Median	345	345	1282
Range	3 - 3881	320 - 378	864 - 3881

Table 2: Descriptive statistics for the full dataset and response groups. Statistics on sample size, age, and survival time are included for the full group of patients that met the inclusion criteria as well as for the typical and exceptional response groups. Typical responders closely resemble the norm for GBM in general, while exceptional responders tend to be younger (though this is not statistically significant) and have an equal representation of males and females as opposed to the usual higher proportion of males (this is statistically significant).

restricting the relevance of this study to only one sex, this issue was addressed in each individual analysis for all data types, as described in subsequent chapters.

In order to ensure that no samples in the study had been exposed to radiation or other treatments that may corrupt results, only untreated tumor samples were included in the dataset. The dataset also includes only primary GBM samples in order to avoid statistical noise from secondary GBM samples, which develop through progression of low-grade astrocytomas and should be approached as a different disease.[25]

The typical and exceptional response groups defined in this chapter were utilized throughout the entirety of this study.

CHAPTER 2

COPY NUMBER

Introduction

A variety of copy number variations have been identified in GBM, the most prevalent of which is amplification of chromosome 7, particularly of EGFR.[26–28] Other frequently occurring copy number changes include losses of chromosomes 9p (particularly of CDKN2A, cyclin-dependent kinase inhibitor 2A)[26,28,29] and 10[26,27,29] as well as gains in chromosomes 19 and 20[26,30]. Some studies have associated these copy number alterations with prognosis, while others have determined that they are not significantly associated with outcome.[27,30–32]

Copy number alterations across the genome can be assessed with comparative genomic hybridization (CGH) arrays as well as single nucleotide polymorphism (SNP) genotyping arrays. The Affymetrix Genome-Wide Human SNP Array 6.0, which was the type of array utilized in this study, assesses more genetic variation than any other array. It includes over 1.8 million markers, including over 946,000 probes that detect copy number variation.[33]

Methods

Affymetrix Genome-Wide Human SNP Array 6.0 CEL files were obtained from the GDC Legacy Archive on April 15, 2017 for 38 typical and 34 exceptional responders (Table 3). The files were divided into four groups based on response group and sex and

	Typical	Exceptional
Female	11	17
Male	27	17

Table 3: Sample number by sex and response group in copy number analysis. Nearly all of the patients in the two response groups had copy number data available. As expected, there is a higher proportion of males in the typical response group and an equal number of males and female in the exceptional response group.

processed with the R tool *Rawcopy*.[34] Log_2 ratio values (relative to normal) obtained from the genelist files generated by *Rawcopy* were compared between typical and exceptional responders to identify any differential gains or losses. A log_2 ratio cutoff of +/- 0.25 was used to define a copy number gain/loss, and only probes where the mean log_2 ratio indicated a gain or loss for at least one of the response groups were included in the analysis. An additional cutoff was applied in which the difference in the mean log_2 ratio between typical and exceptional responders must be > 0.2. Redundant probes (probes for the same gene with the same log_2 ratio value) were removed. Welch's unequal variances t-tests were performed for each remaining probe and a multiple testing correction was performed using the Benjamini-Hochberg false discovery rate (FDR) method ($q < 0.1$). This process was repeated for sex-specific analyses (typical male versus exceptional male and typical female versus exceptional female) with sex chromosomes excluded. A subset of results from this analysis were investigated with respect to the distribution of gains/losses and amplifications/deletions (log_2 ratio cutoff +/- 0.8) between typical and exceptional responders. Significance was determined with chi-squared tests ($p < 0.05$) when the distributions of gains/amplifications and losses/deletions were compared.

Results

Copy number heatmaps for each patient (Figure 2) show consistent alterations regardless of response group or sex. These include gains in chromosomes 7, 19, and 20 as well as losses in chromosomes 9p, 10, 13, and 14, all of which have been described previously in the literature.[26–32] A plot of the mean log_2 ratios across the genome for each

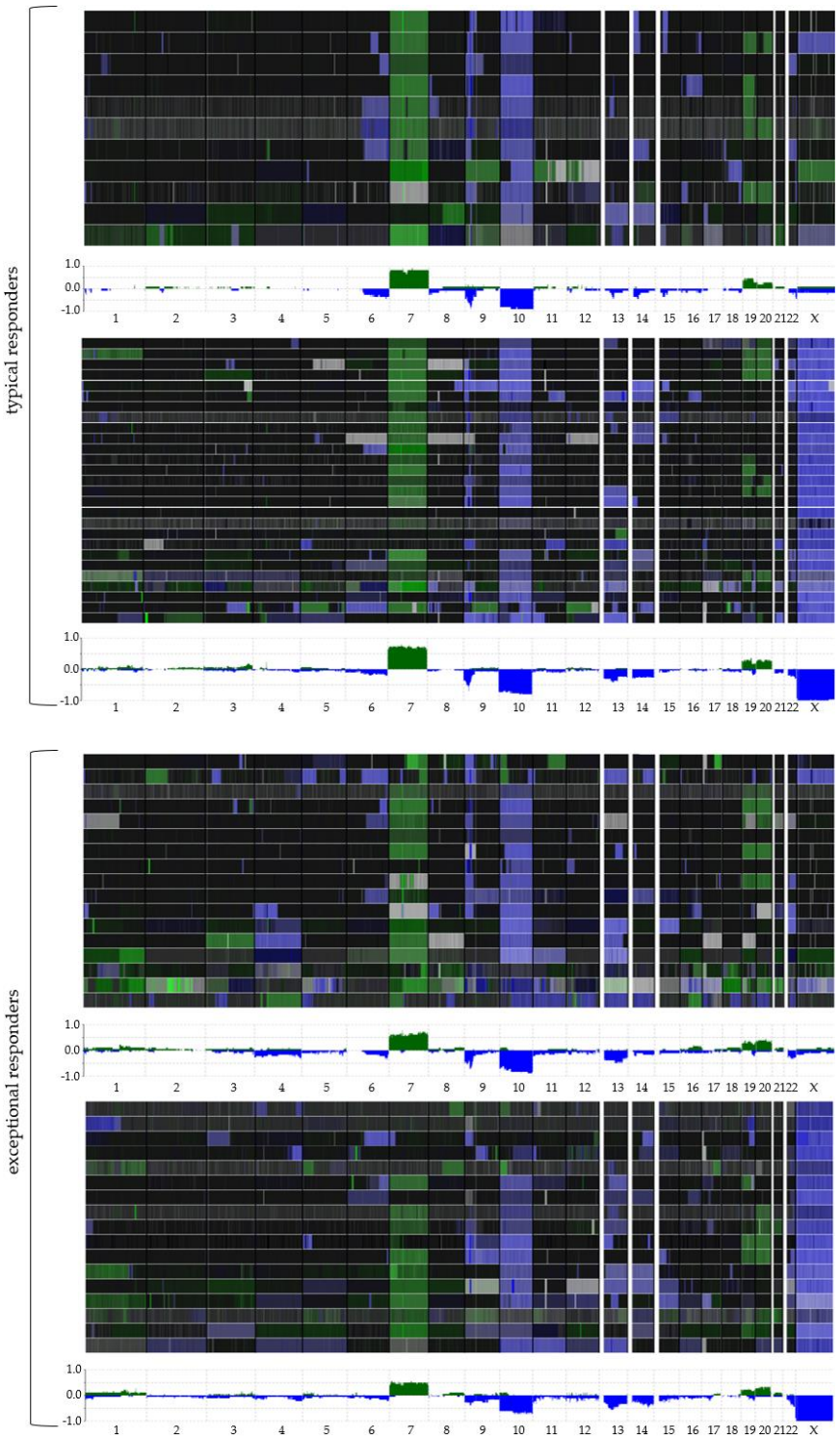

Figure 2: Copy number heat maps by sex and response group. Heat maps generated by Rawcopy for typical and exceptional responders, separated by sex. For each response group, female patients are grouped above and male patients are grouped below. All groups are characterized by gains in chromosomes 7 (particularly around EGFR in 7p), 19, and 20 as well as losses in chromosomes 9p and 10.

response group (Figure 3) also indicates that those gains and losses occur consistently in both response groups, although the magnitude of the gain or loss typically appears to be greater in typical responders.

The *Rawcopy* analysis generates $\log_2$ ratios relative to normal for each probe in each array. After applying the cutoffs described above to the mean $\log_2$ ratios for each response group, 10 probes associated with alterations at 10 genes remained. Following the t-tests and multiple testing correction, 5 of these were determined to be differentially altered between the response groups (Table 4). The 5 genes identified are the olfactory receptors OR4M2 ($p = 0.018$) and OR4N4 ($p = 0.022$), as well as LOC285878 ($p = 0.026$), VSTM2A ($p = 0.025$), and CDKN2A-AS1 ($p = 0.048$). VSTM2A and LOC285878 were both characterized by gains, while OR4M2, OR4N4, and CDKN2A-AS1 were characterized by losses. The sex-specific analyses did not yield any significant results.

Utilizing a $\log_2$ ratio cutoff of +/-0.25 to define copy number gains/losses, several regions as well as specific genes were identified as altered relative to normal (Table 4). Most of these changes have already been implicated in GBM, and most of them were not significantly differentially altered between typical and exceptional responders. These included gains in chromosome 7p (Figure 4), losses in chromosome 9p (Figure 5), losses in chromosome 13q, and losses across the entirety of chromosome 10. Smaller regions of altered copy number included losses at LCE3C, ADAM3A, OR52N5, OR4M2, and OR4N4 as well as gains at FKBP9, PRSS3P2, AND PRSS2.

Based on three of the genes identified as significantly differentially altered between the two response groups, in combination with regions of gain/loss that were

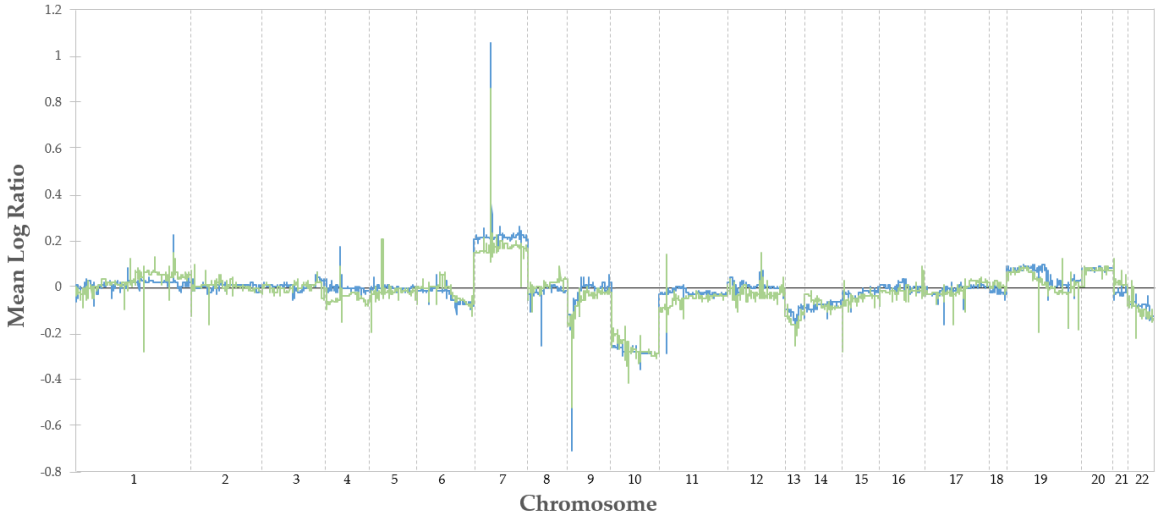

Figure 3: Genome-wide mean copy number for each response group. Mean $\log_2$ ratios assessed at approximately 40,000 probes are shown in blue for typical responders and green for exceptional responders across the genome, excluding sex chromosomes. The most prominent alterations are gains in chromosome 7 and losses in chromosomes 9p and 10. Peaks tend to be of a greater magnitude in the typical response group.

Gains	
Region and Gene(s)	**Response Group Affected**
7p11.2 HPVC1, **VSTM2A***, **LOC285878***, SEC61G, EGFR, EFGR-AS1, LANCL2, VOPP1, FKBP9L, SEPT14, MRPS17, GBAS, PSPH, CCT6A, SNORA15, SUMF2, PHKG1, CHCHD2, and NUPR1L	typical; exceptional to a lesser extent
7q21.2 AKAP9, CYP51A1, LRRD1, KRIT1, ANKIB1, GATAD1, PEX1, RBM48, MGC16142, FAM133B, and CDK6	typical
7q34 PRSS3P2 and PRSS2	typical
Losses	
Region and Gene(s)	**Response Group Affected**
1q21.3 LCE3C	exceptional
8p11.22 ADAM3A	typical
9p21.3, 9p21.2 FOCAD, MIR491, PTPLAD, IFNB1, IFNW1, IFNA21, IFNA4, IFNA7, IFNA10, IFNA16, IFNA17, IFNA14, IFNA22P, IFNA5, KLHL9, IFNA6, IFNA13, IFNA8, IFNA1, MIR31HG, IFNE, MIR31, MTAP, **CDKN2A-AS1***, CDKN2A, CDKN2B-AS1, CDKN2B, DMRTA1, FLJ35282, ELAVL2, IZUMO3, TUSC1, LOC100506422	typical; exceptional to a lesser extent
entirety of chromosome 10	typical and exceptional
11p15.4 OR52N5	typical
13q14.2 DLEU2, MIR16-1, MIR15A, DLEU1, and ST13P4	exceptional
15q11.2 **OR4M2*** and **OR4N4***	exceptional

Table 4: Regions of copy number gain and loss. Regions of copy number gain (mean $\log_2$ ratio > 0.25) and loss (mean $\log_2$ ratio < -0.25) are shown with lists of specific genes affected in each region. The affected response group is described for each region. Genes in bold and labeled with a * reached statistical significance based on t-tests comparing typical and exceptional responders.

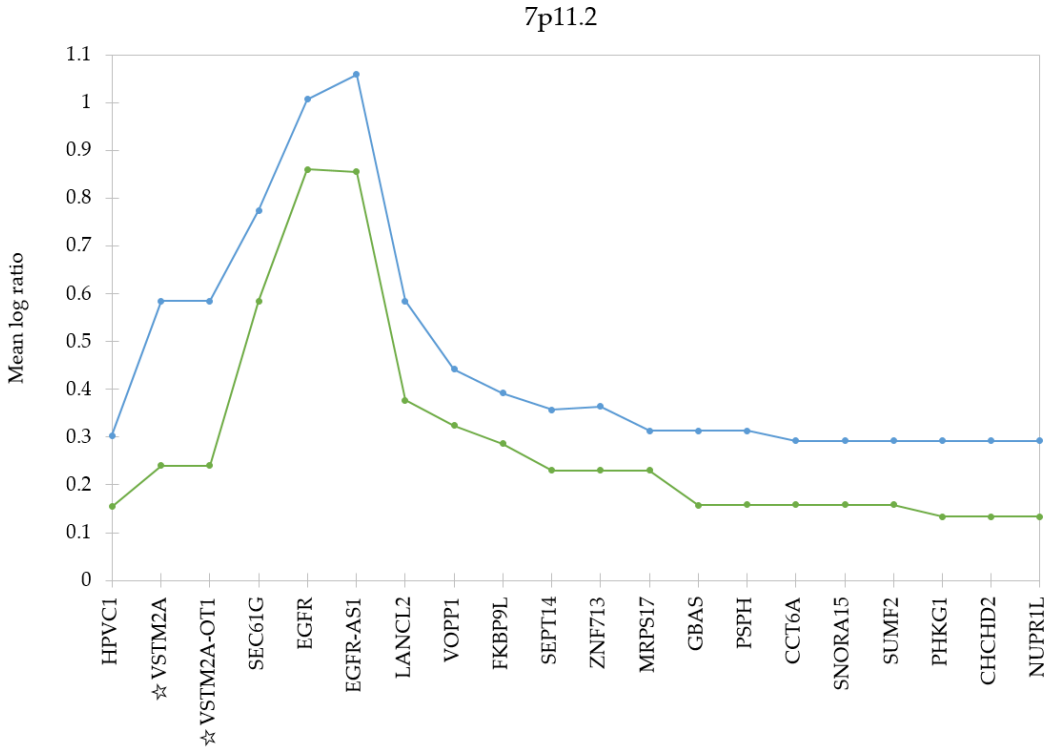

Figure 4: Copy number gain in 7p11.2. Mean $\log_2$ ratios for typical (blue) and exceptional (green) responders in the region of chromosome 7p11.2 described in Table 3. Both groups are characterized by gains ($\log_2$ ratio > 0.25) and amplifications ($\log_2$ ratio > 0.8) in this region, but the magnitude is greater in typical responders. VSTM2A and VSTM2A-OT1 (marked with stars) have differential copy numbers between the two groups that reach statistical significance.

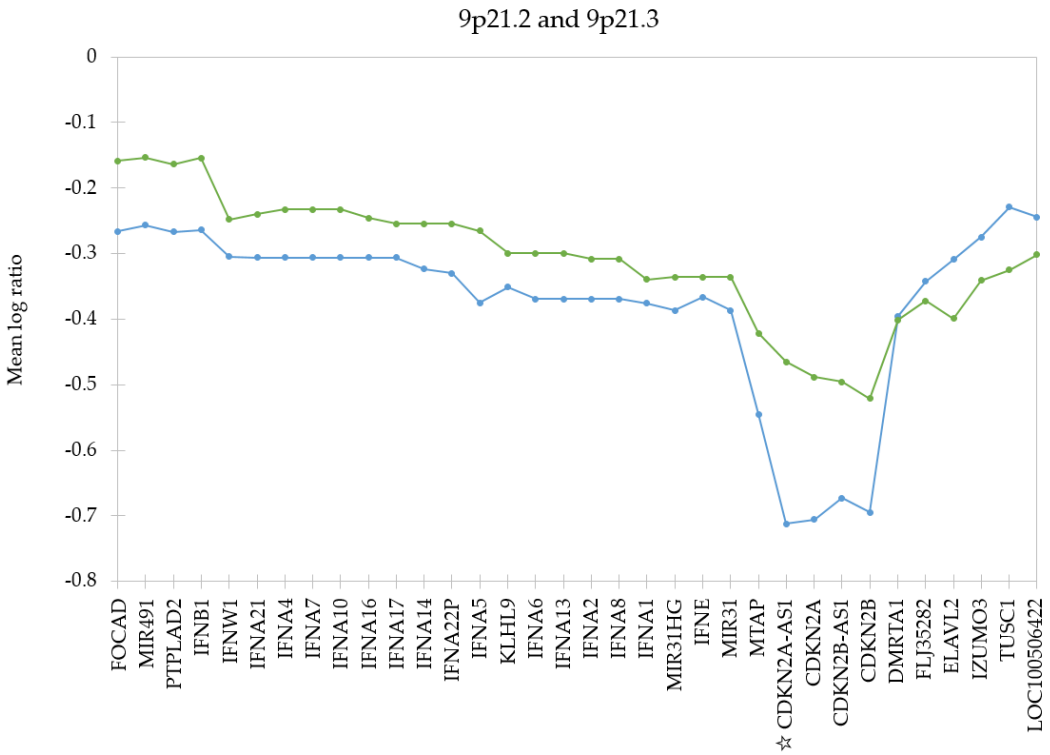

Figure 5: Copy number loss in 9p21.2 and 9p21.3. Mean $\log_2$ ratios for typical (blue) and exceptional (green) responders in the region of chromosome 9p21 described in Table 3. Both groups are characterized by losses ($\log_2$ ratio < -0.25) in this region, but the magnitude is generally greater in typical responders. CDKN2A-AS1 (marked with a star) has differential copy numbers between the two groups that reach statistical significance.

found, EGFR and CDKN2A/B were investigated more closely due to their proximity to VSTM2A, LOC285878, and CDKN2A-AS1 as well as their previously established relevance to GBM.[35] Chi-squared tests examining the distribution of gains/amplifications and losses/deletions among all patients in each response group indicate that typical responders are more likely than exceptional responders to experience loss or deletion of CDKN2A. The distributions of EGFR gain/amplification and CDKN2B loss/deletion were not significantly different between the response groups.

Of the 1812 probes that meet the definition of copy number gain/loss, 1752 of them are losses, and 1201 of those have a greater magnitude in typical responders. Only 60 probes indicate copy number gains, and the magnitude is greater in typical responders for all 60 of them. Overall, 69.6% of the alterations have a larger magnitude in typical responders.

Discussion

Defining Cutoff Values

There is no standardized $\log_2$ ratio cutoff to define copy gain and loss or amplification and deletion. However, a cutoff of ±0.25 for gain/loss and ±0.8 for amplification/deletion is commonly utilized for copy number studies in cancer, which is why those definitions were applied in this study.[36–38]

The vast majority of $\log_2$ ratios in this study are between 0 and 1, making a fold change cutoff at worst misleading and at best uninformative. Rather than apply a fold change cutoff to identify regions of differential copy number alterations between the

response groups, a cutoff was applied in which the absolute value of the difference between the mean $\log_2$ ratios for typical and exceptional responders must be greater than 0.2. This is the same value that is typically chosen for a $\Delta\beta$ cutoff in methylation studies, which also have values between 0 and 1.[39–41]

Differentially Altered Genes Between the Response Groups

Two of the significantly differentially altered genes identified were olfactory receptors. This is likely an artifact and not actually associated with survival in GBM, because there is huge variation in copy number in the general population for approximately 50% of olfactory receptors.[42]

The other significant results include LOC285878 and VSTM2A, both of which are immediately upstream of EGFR, and CDKN2A-AS1, which overlaps slightly with CDKN2A and precedes CDKN2B. It is likely that these genes are significantly differentially altered between the response groups due to their very close proximity to EGFR and CDKN2A/B. Copy number alterations in GBM in both of those regions are very well characterized. It is for this reason that those genes were investigated further. Chi-squared tests indicate that loss or deletion of CDKN2A is more likely to occur in typical responders than in exceptional responders, suggesting that copy number alteration of CDKN2A could serve as a prognostic factor in GBM, associated with a poorer outcome.

Other Gains and Losses

Several regions of copy number gain or loss relative to normal were identified in this analysis. Most of these met the definition of gain or loss for one response group and not the other, but did not reach statistical significance. Nevertheless, these may still be of clinical interest and are worth further exploration in larger studies in the future. Some of these alterations include loss of LCE3C, MIR16-1, MIR15A, and ST13P4 in exceptional responders, loss of ADAM3A in typical responders, and gain of PRSS2 in typical responders. Upregulation of MIR16 and MIR15a is associated with adverse prognosis and poor overall survival in multiple myeloma.[43] Loss of these miRNAs in exceptional responders suggests reduced expression and lower tumorigenic potential in that response group. PRSS2 (protease, serine 2) is thought to play a role in tumor invasion in multiple cancers[44–46] and may be contributing to more aggressive tumors in typical responders.

Rates of Gains and Losses

Losses occurred much more frequently than gains in both response groups, and these alterations were consistently of a higher magnitude in typical responders (Figures 3, 4, and 5; Table 4). This is consistent with previous studies showing that losses occur more frequently than gains in GBM, as well as cancers in general. This suggests that typical responders have more frequent or more severe copy number alterations, possibly contributing to more aggressive tumors and a poorer prognosis.

CHAPTER 3

METHYLATION

Introduction

Methylation and β Values

DNA methylation is an epigenetic mechanism in which a methyl group is added to a CpG site in the DNA. Methylation is typically associated with gene silencing, particularly when the CpG site is located within a promoter.[47] The level of methylation at any given site is reported as a β value, which is the ratio of intensities between methylated and unmethylated alleles. This value ranges from 0 (unmethylated) and 1 (fully methylated). A β value greater than 0.7 is indicative of hypermethylation, while a β value under 0.3 is defined as hypomethylation.[48–50]

In glioblastoma, a recurrent methylation aberration occurs in the promoter for MGMT (O-6-methylguanine-DNA methyltransferase). Methylation in this region silences expression of MGMT, which leaves tumor cells susceptible to alkylating agents. Methylation of MGMT is therefore a marker of a positive response to chemotherapy treatment with temozolomide in GBM.[51]

HumanMethylation27 Array

The HumanMethylation27 array utilizes Infinium genotyping technology to assess the methylation level at 27, 578 CpG sites covering 14,495 genes. This method begins with bisulfite conversion, which converts unmethylated cytosine into uracil and leaves

methylated cytosine alone. The DNA is then amplified, and uracil is matched with adenine, which then pairs with thymine for subsequent replications. The DNA is then hybridized to a chip, which has two probes for each locus (one for the methylated version and one for the unmethylated version). The different probes are then stained with different fluorescent agents and the level of methylation is determined with the calculation of β values.[48,52]

Methods

Illumina HumanMethylation27 idat files were acquired from the GDC Legacy Archive on September 27, 2016 for 16 typical responders and 16 exceptional responders (Table 5). The analysis was performed with *RnBeads*, an R package designed to perform an analysis of DNA methylation at single nucleotide resolution in a more comprehensive manner than other methylation tools.[53] The arrays were normalized with the beta-mixture quantile normalization method and the *Greedycut* algorithm was utilized for filtering. *RnBeads* includes a module that addresses batch effects. There was not an adequate number of samples to complete separate analyses to address the confounding variable of sex. Sex chromosomes were not included in the analysis to address this issue. The resulting lists of CpG sites and promoters were narrowed further with a Δβ (the absolute value of the difference between the mean β value for each response group) cutoff of 0.2, which is a commonly used cutoff for studies of differential methylation.[39–41] The Benjamini-Hochberg multiple testing correction[54] was performed ($q < 0.1$) to identify differentially methylated sites and promoters. The "normal" dataset was obtained from a

	Typical	Exceptional
Female	3	13
Male	10	6

Table 5: Sample number by sex and response group in methylation analysis. Nearly half of patients had methylation data available. The distribution of sex in this group does not follow the proportions of the full response groups, with nearly all of the typical responders being male and nearly all the exceptional responders female, rather than a slight male majority and an even split, respectively. Because of this, there was an insufficient sample size for sex-specific analyses.

2013 study on methylation in neuronal and glial cells in which Illumina HumanMethylation450 experiments were performed for non-neuronal cells of 6 different subjects, with 2 experiments for each subject.[55] Mean β values were calculated from the signal intensities for all 12 sets. Kolmogorov-Smirnov (KS) tests were performed to compare the distributions of beta values for typical, exceptional, and normal groups.

Results

The *RnBeads* report includes a differential methylation file comparing the mean degree of methylation between the typical and exceptional response groups at each of the approximately 27,000 CpG sites assessed by the Illumina assay. After applying the cutoffs and a multiple testing correction, 41 differentially methylated CpG sites corresponding to 37 unique genes were identified (Table 6), 39 of which had a higher degree of methylation in the exceptional response group. A modified volcano plot (Figure 6) indicates which of these sites are outliers with the highest Δβ values and the lowest p values, with PCDHB12 (protocadherin beta 12), LY6K (lymphocyte antigen 6 family member K), and NKX2-5 (NK2 homeobox 5) among the top results.

The *RnBeads* report also includes a differential methylation file comparing the mean degree of methylation across promoter sites between the two response groups. In this case, data from multiple CpG sites in the same promoter (1.5 kb upstream and 0.5 kb downstream of the transcription start site) are combined to reflect the overall methylation level across the region. Utilizing the same cutoffs and multiple testing correction from the site analysis, 5 differentially methylated promoters, all with a higher degree of

cgid	gene name	mean β exceptional	mean β typical	Δβ	p value
cg12343638	PCDHB12	0.5078026	0.262902748	0.244899925	0.000792036
cg08569678	LY6K	0.441856355	0.178005278	0.263851077	0.000901718
cg12052765	CHAT	0.351379153	0.140280129	0.211099024	0.001210066
cg03294619	NKX2-5	0.363597909	0.088957022	0.274640887	0.001724144
cg21480743	PTEN	0.381009425	0.18046069	0.200548735	0.00232438
cg04369341	C20orf100	0.63529373	0.426659616	0.208634114	0.002811554
cg11532513	LRTM1	0.406058029	0.630219234	0.224161205	0.002811802
cg17651821	HIST1H4L	0.513744205	0.256118974	0.257625231	0.003395508
cg25957124	DNAH3	0.705059137	0.476860941	0.228198195	0.003418798
cg01888566	MEST	0.74056681	0.521977296	0.218589514	0.003421299
cg23519022	CAPZB	0.534783267	0.74339535	0.208612083	0.003806896
cg24101578	CDH22	0.618165843	0.406560371	0.211605472	0.003809712
cg25946389	MGMT	0.394523362	0.179592629	0.214930733	0.003829408
cg25509184	CFTR	0.600842388	0.382027484	0.218814904	0.004066663
cg09522147	KRT7	0.588742794	0.36717856	0.221564233	0.004232271
cg10303487	DPYS	0.431869687	0.181799459	0.250070228	0.006518306
cg09595479	PRPH	0.732006613	0.52721146	0.204795153	0.007883096
cg18676237	SERPINB9	0.474358981	0.201989707	0.272369273	0.009013391
cg12558519	KLHL26	0.34932085	0.099339316	0.249981535	0.009626013
cg26980692	SLC15A3	0.401463369	0.190110148	0.211353221	0.010153909
cg00949442	ABCA3	0.424217955	0.221278093	0.202939862	0.012334811
cg24264506	TTC12	0.455964348	0.192059252	0.263905096	0.013005415
cg16363586	BST2	0.679897957	0.474140835	0.205757122	0.01303555
cg27090216	TNFRSF10C	0.323192125	0.115625389	0.207566736	0.013564396
cg13067215	CGI-38	0.317780555	0.106700898	0.211079656	0.01374847
cg20050826	K6IRS2	0.412443387	0.211270568	0.201172819	0.014985767
cg21215336	LRRC8E	0.595433055	0.390559013	0.204874042	0.016306681
cg12981137	MGMT	0.376274797	0.150035143	0.226239654	0.017000628
cg01009664	TRH	0.460950967	0.249015311	0.211935657	0.018029927
cg09160477	SUSD3	0.274926433	0.067484552	0.207441881	0.01810372
cg07753583	LRRC61	0.62589652	0.417196705	0.208699815	0.01842675
cg07952391	FLJ10916	0.409459146	0.184952884	0.224506262	0.018626547
cg23244913	HCG9	0.543443874	0.332649804	0.21079407	0.018700098
cg12177743	TTC12	0.331577699	0.116007362	0.215570337	0.019122814
cg06274159	ZFP42	0.630768539	0.422974537	0.207794002	0.019292771
cg25057743	PTHR2	0.385272981	0.183412711	0.201860269	0.022032263
cg07260592	LPA	0.653635527	0.435082623	0.218552904	0.022092278
cg17965019	HIST1H3J	0.352617133	0.142842183	0.20977495	0.024151396
cg17860158	CNTN2	0.394115837	0.183061088	0.21105475	0.025666055
cg12768605	LYPD5	0.498629758	0.298008361	0.200621398	0.033938924
cg00630164	KCNQ4	0.489497043	0.277459749	0.212037294	0.042287698

Table 6: Significantly differentially methylated CpG sites. These 41 CpG sites were determined to be differentially methylated between typical and exceptional responders following a Δβ cutoff of 0.2 and multiple testing correction ($q < 0.1$). The raw p values are shown.

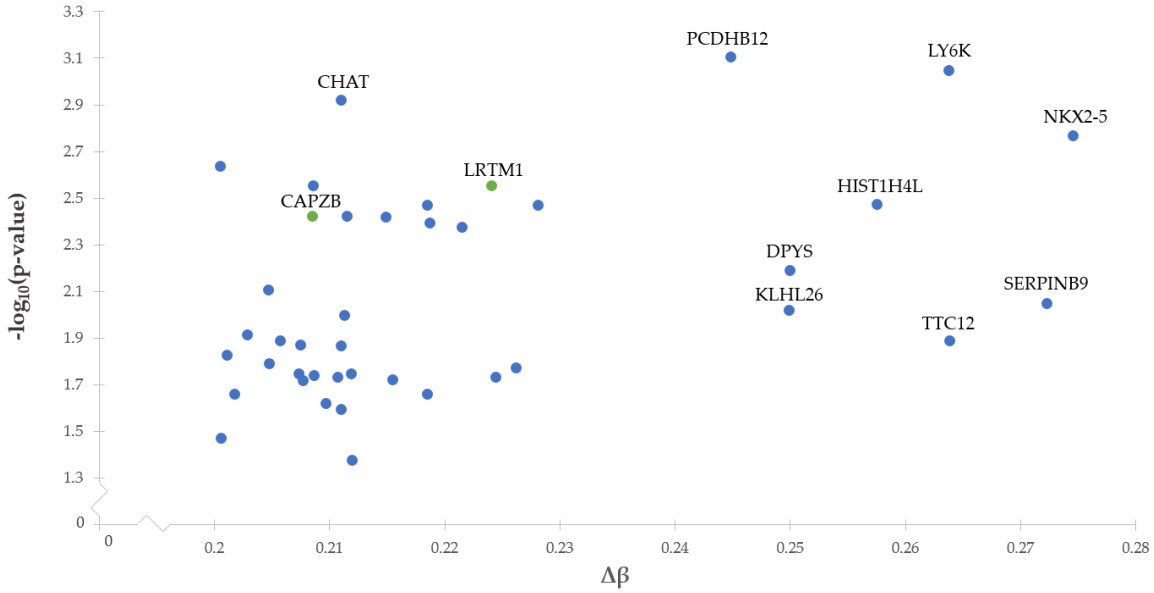

Figure 6: Modified volcano plot of significantly differentially methylated CpG sites. This includes the 41 CpG sites listed in Table 5. Each axis is skewed to reflect the cutoffs made to assess significance ($p < 0.05$ and $\Delta\beta > 0.2$). Sites with a lower degree of methylation in typical responders are shown in blue and sites with a lower degree of methylation in exceptional responders are shown in green. Sites in the upper right hand side have the largest $\Delta\beta$ values and the smallest p values. Outliers and both green sites are labeled with their associated gene name.

methylation in exceptional responders, were identified (Table 7). These include SLC15A3 (solute carrier family 15 member 3), TTC12 (tetratricopeptide repeat domain 12), LRRC8E (leucine rich repeat containing 8 family member E), SUSD3 (sushi domain containing 3), and LRRC61 (leucine rich repeat containing 61).

There are 45 CpG sites with a $\Delta\beta$ value greater than 0.2 between the typical and exceptional response groups, 41 of which are also present in the normal dataset. Histograms for each of these groups (Figure 7) show larger proportions of hypomethylated ($\beta < 0.3$) sites in the typical and normal groups and more moderate β values in the exceptional group. This observation was investigated further with KS tests and cumulative distributions plots for each group (Figure 8). There is no difference in the distribution of β values between typical responders and normal glial cells ($p=0.127$), but the exceptional response group β value distribution is significantly different from both of the other groups ($p<0.0001$ in both cases). The D statistic (a measure of the magnitude of the difference between two datasets) is 0.622 for the typical versus exceptional comparison and 0.734 for the normal versus exceptional comparison.

Discussion

Addressing the Confounding Variable

The distribution of sex between the typical and exceptional response groups is quite skewed in the methylation dataset. For the other data types in this study, analyses were typically performed three times: once for typical versus exceptional responders overall, and then one analysis for each sex in order to address the confounding variable.

gene name	mean β exceptional	mean β typical	Δβ	combined p value
SLC15A3	0.401463369	0.190110148	0.211353221	0.010153909
TTC12	0.393771023	0.154033307	0.239737716	0.01224649
LRRC8E	0.595433055	0.390559013	0.204874042	0.016306681
SUSD3	0.274926433	0.067484552	0.207441881	0.01810372
LRRC61	0.62589652	0.417196705	0.208699815	0.01842675

Table 7: Significantly differentially methylated promoters. The promoters of these five genes were determined to be differentially methylated between typical and exceptional responders following a Δβ cutoff of 0.2 and multiple testing correction ($q < 0.1$), based on the degree of methylation of all the CpG sites that fall within the promoter range (1.5 kb upstream and 0.5 kb downstream of the transcription start site) for each gene. The raw combined p values are shown.

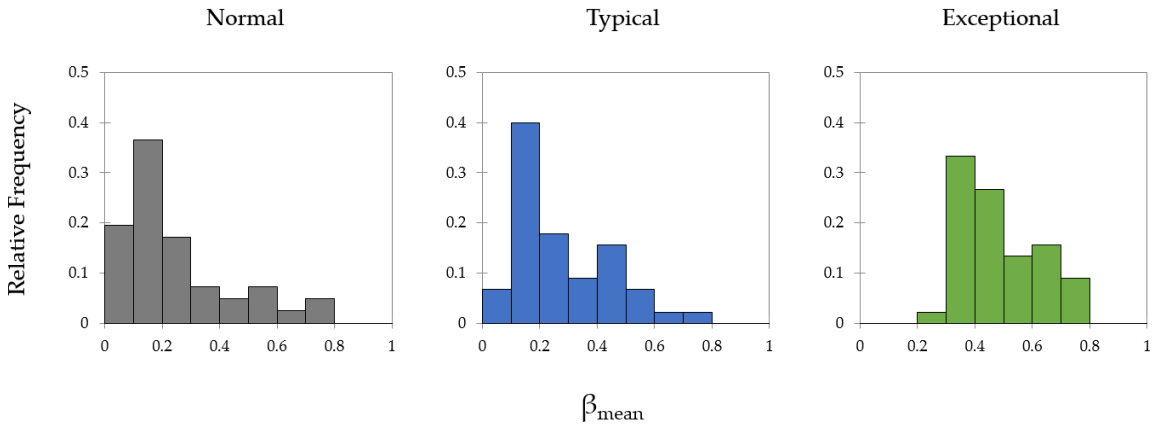

Figure 7: Histograms of β values in normal glial cells and each response group. The response group histograms include mean β values for the 45 CpG sites with Δβ values larger than 0.2, and the normal histogram includes β for 41 of those sites (the remaining 4 were not assessed in the normal arrays). The distribution of β values in typical responders closely resembles the normal distribution, while the exceptional responders are characterized by a shift towards larger β values.

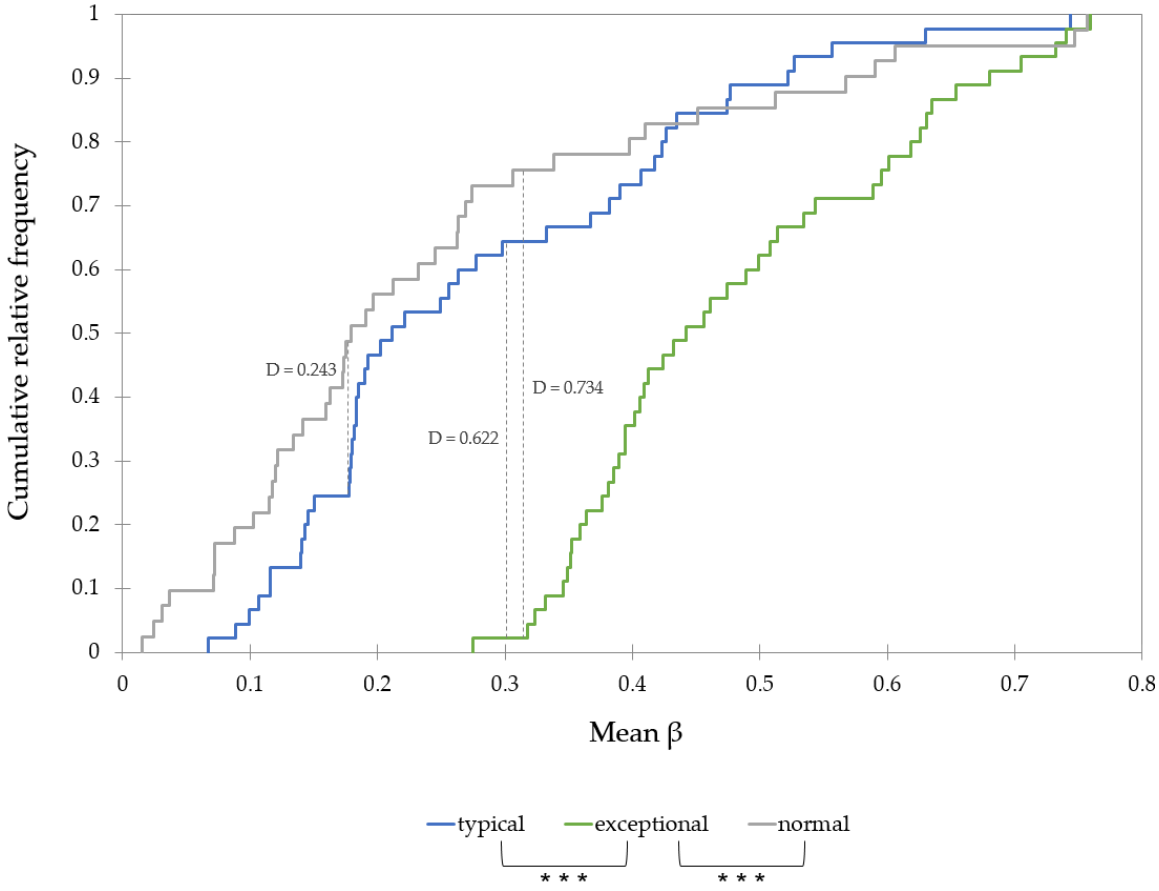

Figure 8: KS tests and cumulative distribution plots of β values. KS tests indicate that the distribution of β values for the CpG sites with $\Delta\beta > 0.2$ is significantly different from both the typical and normal distributions ($p < 0.0001$). There is no difference between the distributions for the typical and normal groups. Cumulative distribution plots are shown for each of the three groups, indicating a clear shift in exceptional responders toward higher β values. The D statistic, a measure of the magnitude of the difference between two datasets, is shown for each comparison at the point of greatest difference. The distance between the exceptional and normal distributions is slightly larger than the distance between the typical and normal distributions.

In the case of methylation, this was not possible due to an insufficient number of samples. Instead, sex chromosomes were removed from the analysis in an effort to control for methylation differences between sexes. When this step was not taken, 44% of significant sites and 57% of significant promoters were on the X chromosome, which was likely due to X-inactivation and the large difference in the number of female patients between the response groups. This method seems to have been effective as many of the results are associated with GBM specifically or with other cancers and are often prognostic factors.

Determination of an Appropriate Δβ Cutoff

A $\log_2$ fold change cutoff is not appropriate for comparing β values since they are all between 0 and 1 and very large fold change values could result from very small changes in beta values, and vice versa. Instead, a Δβ cutoff was applied. A cutoff value of 0.2 was selected based on the literature and on the 27k assay technology. Most methylation studies that use Δβ values select 0.2 as the cutoff[39–41], and the Infinium I technology used in the HumanMethylation27 arrays can detect a Δβ of approximately |0.2| with 99% confidence.[56]

Differentially Methylated Regions

Although there are many genes implicated in the site-specific analysis that did not appear in the results of the promoter analysis, this does not necessarily indicate that the site-specific results are not important. There are many more CpG sites than the 27k technology assesses, so it is entirely plausible that the promoters for those genes have

more sites that are differentially methylated between the response groups that simply were not interrogated with the 27k arrays. A similar study with data from the Illumina 450k or EPIC (which assesses methylation status at over 850,000 CpG sites) platform would be a significant improvement. Unfortunately, there was not a sufficient number of patients that had 450k results to perform such an analysis.

As expected, MGMT was present in the results and is characterized by a higher level of methylation in exceptional responders. Several other genes present in the top results with a higher degree of methylation in exceptional responders are associated with cancer prognosis and/or treatment response, including LY6K (lymphocyte antigen 6 family member K)[57,58], DPYS (dihydropyrimidinase)[59], and SERPINB9 (serpin family B member 9)[60,61]. The reduced methylation of those genes in typical responders suggests increased transcription, and expression of each of those genes is associated with more aggressive tumors in various cancers.

Only two of the significantly differentially methylated sites had a higher level of methylation in typical responders. Those two sites are associated with the genes CAPZB (capping actin protein of muscle Z-line beta subunit) and LRTM1 (leucine rich repeats and transmembrane domains 1). LRTM1 is not well-characterized, but CAPZB is known to be an actin-capping protein that plays a role in cell morphology and differentiation. CAPZB is a metastasis-suppressor in hepatocellular carcinoma[62] and its lower methylation levels in exceptional responders suggest that it may be more highly expressed than it is in typical responders, possibly providing a protective effect to exceptional responders.

Differential β Value Distribution

Among CpG sites that have a Δβ greater than 0.2, the typical group closely resembles the normal beta value distribution, while the exceptional group is characterized by a higher level of methylation. Nearly all (95%) of the differentially methylated CpG sites and 100% of the differentially methylated promoters have a higher degree of methylation in exceptional responders. Histograms and cumulative distribution plots show a strong shift towards higher β values in exceptional responders, and the KS tests indicate that this difference is statistically significant. This hypermethylation in exceptional responders relative to typical responders and normal glial cells suggests an increased level of transcriptional control that may confer a protective effect to exceptional responders.

CHAPTER 4

GENE EXPRESSION

Introduction

Gene Expression Microarrays

Gene expression levels can be assessed with multiple methods, but the gene expression data in this study was generated with Affymetrix microarrays. This technology enables the analysis of gene expression across the whole genome, including more than 45,000 probe sets to assess relative expression levels of more than 39,000 transcripts and variants.[63]

Analysis Packages

Guanine Cytosine Robust Multi-Array Analysis (GCRMA) is a Bioconductor package that performs normalization of microarrays. This tool implements a background correction, followed by a normalization step to make measurements between different arrays comparable, and then a summarization step to calculate a final expression measurement. It also adjusts for background intensities including optical noise and non-specific binding. It is an improvement upon the commonly used Robust Multi-Array Analysis (RMA) algorithm, which does not adjust well for non-specific binding.[64,65]

Another tool utilized in this analysis is *nsFilter*, which is part of the *genefilter* package and removes non-informative genes to reduce noise in the analysis. This tool removes genes with little variation, consistently low signals, and control probe sets.[66]

Limma is a software package for the analysis of gene expression. It utilizes linear models to assess differential expression for a variety of technologies, including microarrays, RNA sequencing, and quantitative PCR.[67–69]

Aberrant Gene Expression in GBM

The most well-known and most frequently occurring gene expression change in GBM is the overexpression of EGFR. Other hallmark genes in GBM include IGFBP2 (insulin like growth factor binding protein 2), IGFBP5 (insulin like growth factor binding protein 5), VEGF (vascular endothelial growth factor), VCAM1 (vascular cell adhesion protein 1), MCM2 (minichromosome maintenance complex component 2), and TNC (tenascin C).[70]

Methods

Affymetrix HT Human Genome U133 DNA microarray CEL files were obtained from the GDC Legacy Archive on August 5, 2016 for 33 exceptional responders and 34 typical responders (Table 8). Three comparisons were performed: all exceptional responders vs all typical responders, male exceptional responders vs male typical responders, and female exceptional responders vs female typical responders. For each of the three analyses, the arrays were normalized with GCRMA and filtering was performed using the *nsFilter* function of the *genefilter* package in R. All experiments were performed at the same location, which should minimize batch effects. Quality control tests, including boxplots of probe intensities and density vs intensity histograms were generated in R

	Typical	Exceptional
Female	11	17
Male	23	16

Table 8: Sample number by sex and response group in gene expression analysis. Most patients in both response groups had gene expression data available. The distribution of sexes between the two groups is consistent with the full groups, with a male majority in the typical group and a roughly equal number of males and females in the exceptional group.

before and after normalization. Differential expression analysis was performed using *limma*. A $\log_2$ fold change cutoff of 1.5 was applied to the resulting list of probes before the multiple testing correction was performed. Only genes that were implicated in all three analyses or in the full analysis but not in the sex-specific analyses were considered to be significantly differentially expressed between the two response groups. Linear regression models were generated with XLSTAT to investigate the prognostic value of some of the differentially expressed genes. These models were generated using the full TCGA GBM dataset, including 385 patients. Utilizing the same CEL files that were analyzed with *limma*, version 2.2.4 of the Broad Institute's Gene Set Enrichment Analysis (GSEA) tool was used to detect enrichment of gene sets between exceptional and typical responders as well as male and female patients. The CEL files were converted to Gene Count files using the ExpressionFileCreator module found in GenePattern. Normalization was performed with GCRMA in conjunction with quantile normalization. GSEA was run using the c5.all.v6 database, with 1000 permutations performed using "phenotype" as the permutation type.

Results

Quality control assessments indicate that all arrays included in the analysis were normalized properly and none need to be excluded. Box plots of the log intensity distributions for each array (Figure 9) are extremely consistent following normalization, indicating that between-array comparisons can be made with this dataset without removing any arrays. Density plots of log intensity distribution (Figure 10) also show very

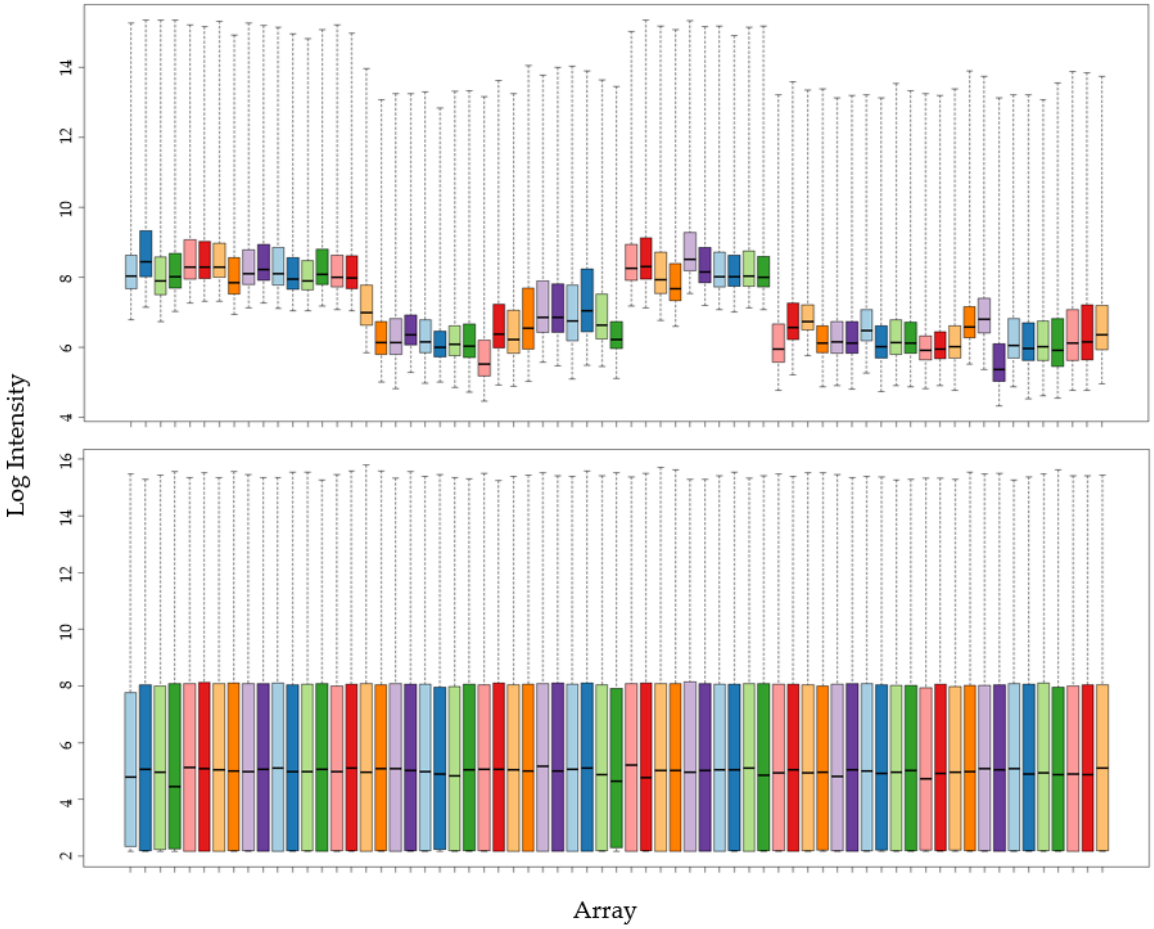

Figure 9: Quality control assessment by log intensity distributions. Log intensity distributions were generated in R before normalization (top) and after normalization (bottom) to determine if between-array comparisons could be made or if any arrays needed to be removed. The consistent distributions following normalization indicate that all arrays in the analysis could be compared and none needed to be removed.

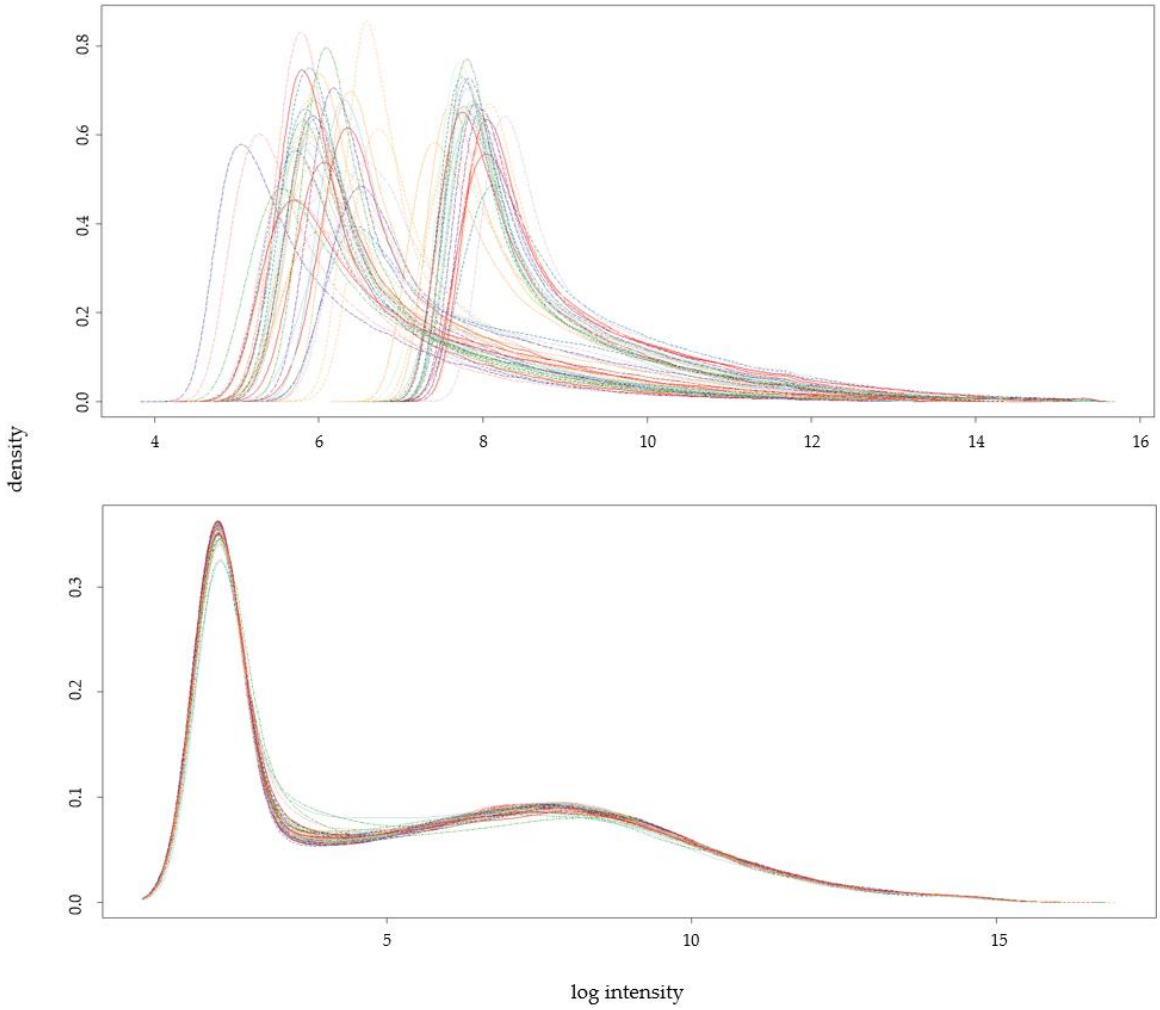

Figure 10: Quality control assessment by density plot. Density plots of log intensity distribution were generated in R before normalization (top) and after normalization (bottom) to identify any arrays with an abnormal distribution. Following normalization, all the arrays had consistent log intensity distributions and none needed to be removed from the analysis.

consistent distributions following normalization, meaning that there are no outliers that need to be excluded.

The analysis identified 4 significantly differentially expressed genes (Table 9). ETNPPL (ethanolamine-phosphate phospho-lyase) and SH3GL2 (SH3 domain containing GRB2 like 2, endophilin A1) were more highly expressed in exceptional responders, while CXCL8 (interleukin 8) and CCL20 (chemokine ligand 20) were more highly expressed in typical responders. There were 13 additional genes determined to be significantly differentially expressed in the full analysis, but they were all also identified in one sex-specific analysis and not the other. These were excluded from the final results because they may only be present due to sex being a confounding variable.

Linear regression models indicate that CXCL8 is predictive of survival time. Increased expression of CXCL8 is associated with reduced survival time. The overall model is statistically significant ($p < 0.001$), as is the CXCL8 term ($p < 0.001$). The equation for the model is: $survival\ time\ (days) = 812 - 40.1 * normalized\ signal\ of\ CXCL8$.

GSEA identified 4 gene sets enriched in exceptional responders and 1 gene set enriched in typical responders (Table 10). The leading edge analysis revealed which genes contributed to the enrichment of which gene sets in the exceptional responders (Figure 11). NLGN1 and STXBP1 contributed to the enrichment of all four gene sets and RAB3A, RIMS3, SNCA, SYN1, RAB5A, RAB3GAP1, and PFN2 contributed to the enrichment of three of them. All of the other implicated genes were associated with only one or two of the gene sets. There were no significantly enriched gene sets identified when the analysis was divided by sex.

Gene	log_2FC (all)	log_2FC (female)	log_2FC (male)	p value (all)	p value (female)	p value (male)
ETNPPL	1.553	1.931	1.659	0.0225	0.0703	0.0560
SH3GL2	1.771	2.301	1.758	0.0018	0.0089	0.0210
CXCL8	-1.735 -1.863	-3.094 -2.345	-1.904	0.0038 0.0045	0.00076 0.0133	0.0330
CCL20	-1.751	-2.383	-1.520	0.0042	0.0155	0.0487

Table 9: Significantly differentially expressed genes. Four genes were determined to be significantly differentially expressed between typical and exceptional responders following multiple testing correction ($q < 0.1$) and comparison to the sex-specific analyses. Log_2 fold change (log_2FC) and p values are shown for the full analysis and for each sex-specific analysis. ETNPPL and SH3GL2 are more highly expressed in exceptional responders and CXCL8 and CCL20 are more highly expressed in typical responders.

Name	Enriched in	Normalized Enrichment Score	p value	FDR q value
regulation of synaptic vesicle transport	exceptional	2.062	p < 0.001	0.124
regulation of neurotransmitter transport	exceptional	2.025	p < 0.001	0.113
positive regulation of calcium ion dependent exocytosis	exceptional	2.008	p < 0.001	0.109
neurotransmitter secretion	exceptional	2.006	p < 0.001	0.078
negative regulation of cytokine biosynthetic process	typical	-2.068	p < 0.001	0.154

Table 10: Enriched gene sets identified by GSEA. GSEA identified five enriched gene sets, four of which are enriched in exceptional responders.

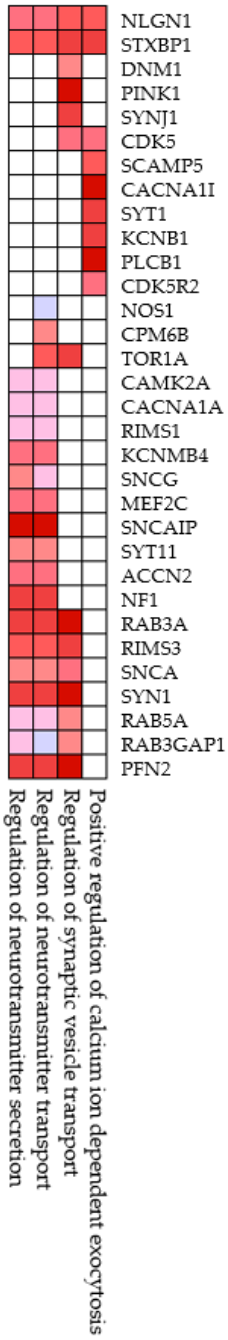

Figure 11: Leading edge analysis of gene sets enriched in exceptional responders. Leading edge analysis of the four gene sets enriched in exceptional responders indicated which genes contributed to each result. The color gradient indicates the range of expression values (red, pink, light blue, and dark blue correspond to high, moderate, low, and lowest expression, respectively). Two genes contributed to all four results and seven genes contributed to three of the four, but most of the genes contributed to just one or two of the enriched gene sets.

Discussion

Addressing the Confounding Variable

A majority of the genes identified as being significantly differentially expressed in the full dataset were excluded because they were also present in the results of only one of the sex-specific analyses and may only appear to be significant due to the confounding variable of sex. In several of the cases, at least, this seems very likely. Some of the genes excluded for potentially being the result of differential expression between males and female include XIST, RPS4Y1, and DDX3Y. XIST is expressed in females as the major effector of the X inactivation process, and RPS4Y1 and DDX3Y are both found on the Y chromosome and should therefore only be expressed in males. That these were excluded from the final results by the methodology for this analysis suggests that this method was successful in controlling for sex-specific results.

Significantly Differentially Expressed Genes

The four genes included in the final results list are CXCL8, CCL20, ETNPPL, and SH3GL2. CXCL8 is an angiogenic factor in GBM, gliomas, and many other cancers and CCL20 promotes malignancy in various cancers and has been implicated in glioma.[71–77] Both of those genes were overexpressed in typical responders relative to exceptional. CXCL8 and CCL20 are often implicated in diseases together, particularly in colorectal cancer, in which they synergize to promote a poor survival outcome via a collaborative induction of the epithelial-mesenchymal transition.[78] Overexpression of these two genes are likely contributing to the poorer prognosis of typical responders. *In vivo* studies of

overexpression of these genes in GBM cell lines would be necessary to confirm that they contribute to a more tumorigenic phenotype.

The other two significant genes were more highly expressed in exceptional responders. ETNPPL is a lyase that is downregulated in hepatocellular carcinoma, and SH3GL2 is a positive prognostic factor in head and neck squamous cell carcinoma.[79] SH3GL2 is targeted by mir330, which promotes malignancy in GBM cell lines, suggesting that reduced expression of SH3GL2 results in more aggressive tumors.[80] The overexpression of these two genes is consistent with a more positive prognosis.

CXCL8 is Predictive of Survival Time

Based on the established relationship between CCL20 and CXCL8 in the literature, these two genes were investigated as possible prognostic factors for GBM with linear regression models. While the term for CCL20 was not statistically significant in the model, when it was removed and the model was regenerated with CXCL8 only, it was found to be significantly predictive of survival. The equation for the model suggests that for every one unit increase in CXCL8 expression, there is an associated 40 day reduction in survival. CXCL8 is a chemokine and potent angiogenic factor that may contribute to tumorigenesis[81], so its overexpression in typical responders relative to exceptional responders is indicative of more aggressive tumors in typical responders, which may explain the significant association between higher CXCL8 expression and shorter survival in the full GBM dataset.

GSEA

GSEA identified 5 enriched gene sets, 4 of which were enriched in exceptional responders and 1 of which was enriched in typical responders. The four gene sets enriched in exceptional responders are all very similar and are related to synaptic transmission, suggesting that exceptional responders have an increased ability to maintain synaptic transmission functions and that it might be granting them some advantage over typical responders.

Interestingly, 6 of the genes contributing to the enriched gene sets in exceptional responders (STXBP1, DNM1, SYNJ1, KCNB1, PLCB1, and CACNA1A) are among a group of genes that have been implicated in early infantile epileptic encephalopathy (EIEE)[82], an extremely debilitating disorder characterized by uncontrollable seizures and severe mental retardation.[83] Mutations in these genes are associated with EIEE, but it appears that overexpression of these genes is associated with a positive prognosis in GBM. All of these genes are associated with synaptic transmission.

CHAPTER 5

miRNA EXPRESSION

Introduction

Micro RNAs (miRNAs) are a class of regulatory molecules that have been implicated as important players in tumorigenesis.[84] A subset of miRNAs have been found to be consistently dysregulated in GBM (Table 11) and some of these can also be used to predict prognosis or therapeutic response. Aberrant expression of these miRNAs can impact tumorigenic pathways in GBM such as induction of angiogenesis, resistance to apoptosis, and sustained proliferation signaling.[85]

The miRNA expression data for this study were generated with the Agilent Human miRNA Microarray 8x15K platform. These microarrays contain probes with high sensitivity and specificity for all human miRNAs reported in the Sanger miRBASE database.[86]

Methods

Files containing the calculated expression values of 534 miRNAs for 38 typical and 34 exceptional responders (Table 12) were retrieved from the GDC Legacy Archive on September 1, 2016. Welch's unequal variances t-tests were performed for each miRNA to identify which ones are significantly differentially expressed between the response groups. The multiple testing correction was performed using the Benjamini-Hochberg

Upregulated				
miR-9	miR-10a	miR-10b	miR-15b	miR-17-5p
miR-25	miR-21	miR-26a	miR-92b	miR-93
miR-106a	miR-130a	miR-155	miR-182	miR-196b
miR-210	miR-221	miR-222	miR-296	miR-451
Downregulated				
miR-7	miR-34a	miR-124*	miR-125b	miR-128*
miR-129-5p*	miR-132	miR-136	miR-137	miR-139-5p*
miR-146b	miR-153	miR-181*	miR-184	miR-218
miR-323	miR-326	miR-328	miR-495	

Table 11: Consistently dysregulated miRNAs in glioblastoma. These miRNAs have been established as frequently upregulated or downregulated in GBM[85] and were used as a guide to reduce multiple testing in the analysis. miRNAs labeled with a * were not assessed by the arrays in this study.

	Typical	Exceptional
Female	11	17
Male	27	17

Table 12: Sample number by sex and response group in miRNA analysis. Nearly all of the patients in both response groups had miRNA expression data available. The distribution of sexes in the two groups is similar to the full groups, with most a majority of typical responders being male and an even number of males and females in the exceptional responders group.

FDR method ($q < 0.1$). The t-tests and corrections were performed again on a reduced dataset after the miRNA list was restricted to only those miRNAs that appear in Table 11.

Results

No miRNAs were significantly differentially expressed between typical and exceptional responders following the multiple testing correction, even when multiple testing was reduced by restricting the results to only miRNAs known to be consistently dysregulated in GBM. Prior to correcting for multiple testing, there were 37 miRNAs with $p < 0.05$ in the full dataset and 2 in the reduced dataset (Table 13).

Discussion

Although quite a few miRNAs have been associated with GBM, including some that are specifically associated with prognosis or treatment response[85], no significantly differentially expressed miRNAs were identified in this study. This does not necessarily mean that the miRNAs analyzed are not important in GBM. They may be differentially expressed in both response groups relative to normal expression, but simply not be differentially expressed between the response groups themselves.

Several of the top results have previously been associated with GBM prognosis or with prognosis and/or treatment response in other cancers. However, these results did not reach statistical significance after a multiple testing correction was applied, so they were not included in any downstream analyses.

Hybridization REF	p value
hsa-miR-621	0.0026
hsa-miR-490	0.0070
hsa-miR-191	0.0094
hsa-miR-330	0.0097
hsa-miR-200b	0.0100
hsa-miR-128a	0.0102
hsa-miR-767-3p	0.0114
hsa-miR-574	0.0124
hsa-miR-200a	0.0154
hsa-miR-128b	0.0156
hsa-miR-510	0.0158
hsa-miR-367	0.0186
kshv-miR-K12-7	0.0192
hsa-miR-429	0.0199
hsa-miR-801	0.0224
hsa-miR-648	0.0233
hsa-miR-204	0.0242
hsa-miR-586	0.0243
hsa-miR-422a	0.0245
hcmv-miR-US5-1	0.0250
hsa-miR-222*	0.0258
hsa-miR-550	0.0305
hsa-miR-200a	0.0315
ebv-miR-BART17-5p	0.0324
kshv-miR-K12-4-3p	0.0330
hsa-miR-548c	0.0336
hsa-miR-603	0.0344
hsa-miR-339	0.0356
ebv-miR-BHRF1-2	0.0378
hsa-miR-604	0.0412
hsa-miR-296*	0.0415
hsa-miR-581	0.0416
hsa-let-7b	0.0434
hsa-miR-20a	0.0444
hsa-miR-19a	0.0455
hsa-miR-520a	0.0455
hsa-miR-345	0.0458

Table 13: Differentially expressed miRNAs. This includes the top results of the miRNA analysis (all miRNAs with $p < 0.05$). miRNAs labeled with a * also appeared in the reduced analysis based on the list in Table 10. None of these results reached statistical significance following a multiple testing correction ($q < 0.1$).

CHAPTER 6

PROTEIN EXPRESSION

Introduction

Reverse Phase Protein Array (RPPA) is a high throughput assay in which antibodies are printed across slides to quantify the amounts of various proteins in multiple samples simultaneously. This method allows for sensitive and accurate quantification of proteins, including phosphoproteins, from a small amount of sample material as long as high quality antibodies are available. The TCGA protein expression data is generated by the MD Anderson RPPA Core Facility, which currently utilizes a panel of 304 antibodies in its experimental protocol.[87,88] The GBM dataset includes results for 222 antibodies.

Methods

Files containing relative protein expression data were obtained from the GDC Legacy Archive for 7 typical and 13 exceptional responders (Table 14) on September 24, 2016. Mann-Whitney U tests were performed for each of the 222 antibodies to identify which proteins are significantly differentially expressed between the response groups. The multiple testing correction was performed using the Benjamini-Hochberg FDR method ($q < 0.1$).

	Typical	Exceptional
Female	2	6
Male	5	7

Table 14: Sample number by sex and response group in protein expression analysis. Very few patients had protein expression data available. The sex distribution across response groups was consistent with the distribution in the full groups, however.

Results

No proteins were significantly differentially expressed between typical and exceptional responders following the multiple testing correction. Without correcting for multiple testing, there are five proteins with significant p values ($p < 0.05$). These proteins are associated with the genes ITGA2, BCL2, BCL2A1, RPS6KB1, and RAD51 (Table 15).

Discussion

Unlike most of the other data types investigated in this work, the protein expression arrays assess a relatively small number of data points. While most of the other analyses were genome-wide, the protein expression data type includes results for just 222 antibodies. Being limited to this comparatively small panel substantially diminishes the likelihood of identifying differential aberrations between typical and exceptional responders for this data type, which may be part of the reason why no statistically significant results were identified.

The protein expression analysis was also characterized by a much smaller sample size than most of the other analyses, with only 7 typical and 13 exceptional responders. Perhaps with a larger sample size, there would be more definitive and statistically significant results.

Of the top results, both BCL2 (B-cell lymphoma 2) and RAD51 (RAD51 recombinase) have been associated with GBM prognosis.[89,90] However, because these results did not reach statistical significance following the multiple testing correction, they were not included in any downstream analyses.

Antibody	Gene	p value
CD49b-M-V	ITGA2	0.0027
Bcl-2-M-V	BCL2	0.0198
Bcl2A1-R-V	BCL2A1	0.0252
P70S6K-R-V	RPS6KB1	0.0324
RAD51-M-C	RAD51	0.0329

Table 15: Differentially expressed proteins. This includes the top results of the protein expression analysis (all antibodies with $p < 0.05$). The gene associated with the protein that each antibody labels is listed. None of these results reached statistical significance following a multiple testing correction ($q < 0.1$).

CHAPTER 7

INTEGRATION AND PATHWAY ANALYSES

Introduction

Pathway analysis techniques are used to help interpret the results of omics studies by identifying genes that play a role in the same cellular process, disease, signaling pathway, or other biological pathway. This can provide a global perspective on the results generated and can help with understanding the results in terms of biological relevance.[91] Two commonly utilized pathway analysis tools are Gene Ontology (GO)[92] and the Kyoto Encyclopedia of Genes and Genomes (KEGG).[93]

The GO project is a tool used to describe gene products in terms of gene product properties, including molecular functions and associated biological processes.[92] *ClueGO* is a Cytoscape plug-in that utilizes Cytoscape's visualization capabilities to generate a GO term network. This tool includes a Fusion feature to reduce redundancy by combining related GO terms into the most representative term.[94]

KEGG is a comprehensive database that represents the current knowledge of molecular interaction and reaction networks and is usually utilized to understand biological pathways and systems, especially in large-scale genomic datasets. This resource can be used to derive a systems-level understanding of molecular-level information and gain insight into the functional significance of the results of high-throughput analyses.[93]

Methods

In order to integrate the results from the individual analyses of this study and derive some functional significance, genes contributing to enriched gene sets, differentially expressed or methylated genes, and genes with differential copy number gains/losses (Table 16) were combined and used as the input for GO and KEGG analyses for each response group. The GO analysis was performed using the Cytoscape plug-in *ClueGO* with all four GO types selected, GO Term Fusion enabled, and results restricted to pathways with $p<0.05$ after Benjamini-Hochberg FDR multiple testing correction. All other parameters were left as the default. In the KEGG analysis, only pathways with at least 3 associated genes were included for further consideration. Heat maps were generated with *Heatmapper*. Both Pearson (r) and Spearman (ρ) correlation coefficients were calculated to assess the degree and direction of correlation between gene expression and copy number as well as gene expression and methylation.

Results

Three enriched GO terms were identified from the gene list that is upregulated in typical responders (Table 17). These include regulation of lipid storage, regulation of interleukin-10 production, and regulation of cytokine biosynthetic process. The exceptional responders, however, had 105 enriched GO terms which *ClueGO* GO Term Fusion reduced to 12 enriched GO terms (Table 18 and Figure 12). Almost all of the enriched GO terms are associated with synapse formation or function.

	Typical Responders	Exceptional Responders
Copy Number	VSTM2A LOC285878 OR4M2 OR4N4	CDKN2A-AS1 CDKN2A
Methylation	SLC15A3 TTC12 LRRC8E SUSD3 LRRC61	
Gene Expression	CXCL8 CCL20	ETNPPL SH3GL2
GSEA	LAG3 SFTPD INHBA TRIB2 KLF4 NMI NFKB1 INHBB IL6 RNF128 BCL3	NLGN1 STXBP1 DNM1 PINK1 SYNJ1 CDK5 SCAMP5 CACNA1I SYT1 KCNB1 PLCB1 CDK5R2 NOS1 CPM6B TOR1A CAMK2A CACNA1A RIMS1 KCNMB4 SNCG MEF2C SNCAIP SYT11 ACCN2 NF1 RAB3A RIMS3 SNCA SYN1 RAB5A RAB3GAP1 PFN2

Table 16: Significantly upregulated genes across all analyses. Genes with statistically significant alterations between the two response groups are included. "Upregulated" for each data type is defined as follows: copy number gains, lower promoter methylation, increased gene expression, and enrichment in GSEA. The only significant results not included are differentially methylated CpG sites. Only promoters were included from the methylation results because promoter methylation is consistently negatively correlated with gene expression, while methylation of sites in other regions is not.

GO Term	p value	Associated Genes
regulation of cytokine biosynthetic process	$1.50x10^{-22}$	BCL3, CCL20, IL6, INHBA, INHBB, KLF4, LAG3, NFKB1, NMI, RNF128, SFTPD, TRIB2
regulation of interleukin-10 production	$1.50x10^{-05}$	BCL3, IL6, TRIB2
regulation of lipid storage	$1.90x10^{-05}$	IL6, NFKB1, VSTM2A

Table 17: Significantly enriched GO terms in typical responders. Three GO terms were identified as significantly enriched in typical responders following Benjamini-Hochberg FDR multiple testing correction. The p values are adjusted. Each group identified with GO Term Fusion only has one associated GO term.

GO Term	p value	Associated Genes
presynaptic process involved in chemical synaptic transmission	$2.9x10^{-31}$	ASIC1, CACNA1A, CAMK2A, CDK5, KCNMB4, MEF2C, NF1, NLGN1, PFN2, RAB3A, RAB3GAP1, RAB5A, RIMS1, RIMS3, SNCA, SNCAIP, SNCG, STXBP1, SYN1, SYNJ1, SYT1, SYT11, TOR1A
regulation of neurotransmitter transport	$4.6x10^{-29}$	ASIC1, CACNA1A, CAMK2A, KCNMB4, MEF2C, NF1, NLGN1, NOS1, PFN2, RAB3A, RAB3GAP1, RAB5A, RIMS1, RIMS3, SNCA, SNCAIP, SNCG, STXBP1, SYN1, SYT11, TOR1A
synaptic vesicle exocytosis	$4.9x10^{-27}$	CACNA1A, CDK5, NLGN1, PFN2, RAB3A, RAB3GAP1, RAB5A, RIMS1, RIMS3, STXBP1, SYN1, SYNJ1, SYT1, SYT11
regulation of synaptic vesicle recycling	$3.0x10^{-24}$	CDK5, DNM1, NLGN1, SNCA, SYT11, TOR1A
regulation of synaptic vesicle exocytosis	$6.1x10^{-22}$	NLGN1, PFN2, RAB3A, RAB3GAP1, RAB5A, RIMS1, RIMS3, STXBP1, SYN1
neuron-neuron synaptic transmission	$2.5x10^{-21}$	CACNA1A, CDK5, DNM1, MEF2C, NF1, NLGN1, PINK1, RAB3GAP1, SNCA, STXBP1, SYT1, TOR1A
regulation of synaptic plasticity	$1.2x10^{-15}$	CAMK2A, CDK5, KCNB1, MEF2C, NF1, NLGN1, RAB3A, RAB3GAP1, SNCA, STXBP1, SYT11
acid secretion	$1.5x10^{-15}$	CACNA1A, NF1, RAB3A, RAB3GAP1, RIMS1, SNCA, STXBP1, SYT1
positive regulation of synaptic transmission	$2.1x10^{-15}$	NF1, NLGN1, PINK1, RAB3GAP1, RIMS1, SNCA, SYT1, SYT11
regulation of amine transport	$6.8x10^{-13}$	CACNA1A, KCNB1, PINK1, RAB3GAP1, SNCA, SNCG, SYT1, TOR1A
positive regulation of protein targeting to membrane	$2.3x10^{-5}$	CACNA1A, CDK5, KCNB1
voltage-gated calcium channel complex	$5.0x10^{-5}$	CACNA1A, CACNA1I, NOS1

Table 18: Significantly enriched GO terms in exceptional responders. Twelve GO term groups were identified as significantly enriched in exceptional responders following Benjamini-Hochberg FDR multiple testing correction. The p values are adjusted and based on the groups identified by GO Term Fusion, not individual GO terms.

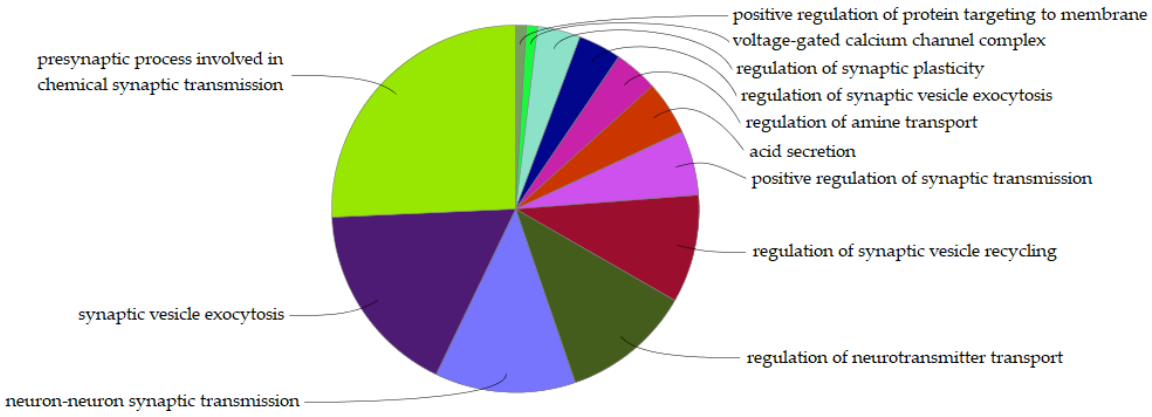

Figure 12: GO Term Fusion results in exceptional responders. The 12 GO term groups identified with GO Term Fusion are associated with 105 enriched GO terms. The pie graph indicates the number of GO terms included in each group.

KEGG analysis for the typical responders identified quite a few pathways based on 5 subsets of the upregulated genes (Table 19). The top results are the IL-17 signaling pathway and the TNF signaling pathway. The KEGG analysis for the exceptional responders gene list yielded 5 enriched pathways, including synaptic vesicle cycle, MAPK signaling pathway, calcium signaling pathway, Alzheimer's disease, and Parkinson's disease (Table 20).

The relationships between the data types with significant results (copy number, methylation, and gene expression) was further investigated with heat maps and correlation analyses (Figure 13 and Table 21). Among only the genes identified in Table 16, there is a weak to moderate positive correlation between copy number and gene expression for both response groups, with similar coefficients resulting from both the Pearson and Spearman tests. A strong negative correlation between methylation and gene expression was indicated by both correlation tests for that same gene list. When comparing methylation and gene expression genome-wide, there is a moderate to strong negative correlation for both the typical responders ($r = -0.356$, $\rho = -0.363$) and the exceptional responders ($r = -0.365$, $\rho = -0.387$).

Discussion

Spearman versus Pearson Correlation

Both the Spearman and the Pearson correlation methods are utilized in the literature to assess the relationship between methylation and gene expression, but most studies include only one or the other. The Pearson test is more suited to continuous

Genes	Pathways
CCL20 IL6 CXCL8 NFKB1	IL-17 signaling pathway
BCL3 CCL20 IL6 NFKB1	TNF signaling pathway
IL6 CXCL8 NFKB1	Pathways in cancer Non-alcoholic fatty liver disease Chagas disease Legionellosis Pertussis Salmonella infection Amoebiasis Kaposi's sarcoma-associated herpesvirus infection Transcriptional misregulation in cancer Toll-like receptor signaling pathway Cellular senescence Hepatitis B NOD-like receptor signaling pathway AGE-RAGE signaling pathway in diabetic complications Influenza A
CCL20 IL6 CXCL8	Rheumatoid arthritis Cytokine-cytokine receptor interaction
CCL20 CXCL8 NFKB1	Chemokine signaling pathway

Table 19: Significantly enriched KEGG pathways in typical responders. KEGG analysis identified multiple pathways enriched in typical responders that are associated with five groups of genes.

Genes	Pathways
CACNA1A DNM1 RAB3A RIMS1 STXBP1 SYT1	Synaptic vesicle cycle
CACNA1A CACNA1I MEF2C NF1	MAPK signaling pathway
CACNA1A CACNA1I NOS1	Calcium signaling pathway
CDK5 NOS1 SNCA	Alzheimer's disease
PINK1 SNCA SNCAIP	Parkinson's disease

Table 20: Significantly enriched KEGG pathways in exceptional responders. KEGG analysis identified enriched pathways in exceptional responders.

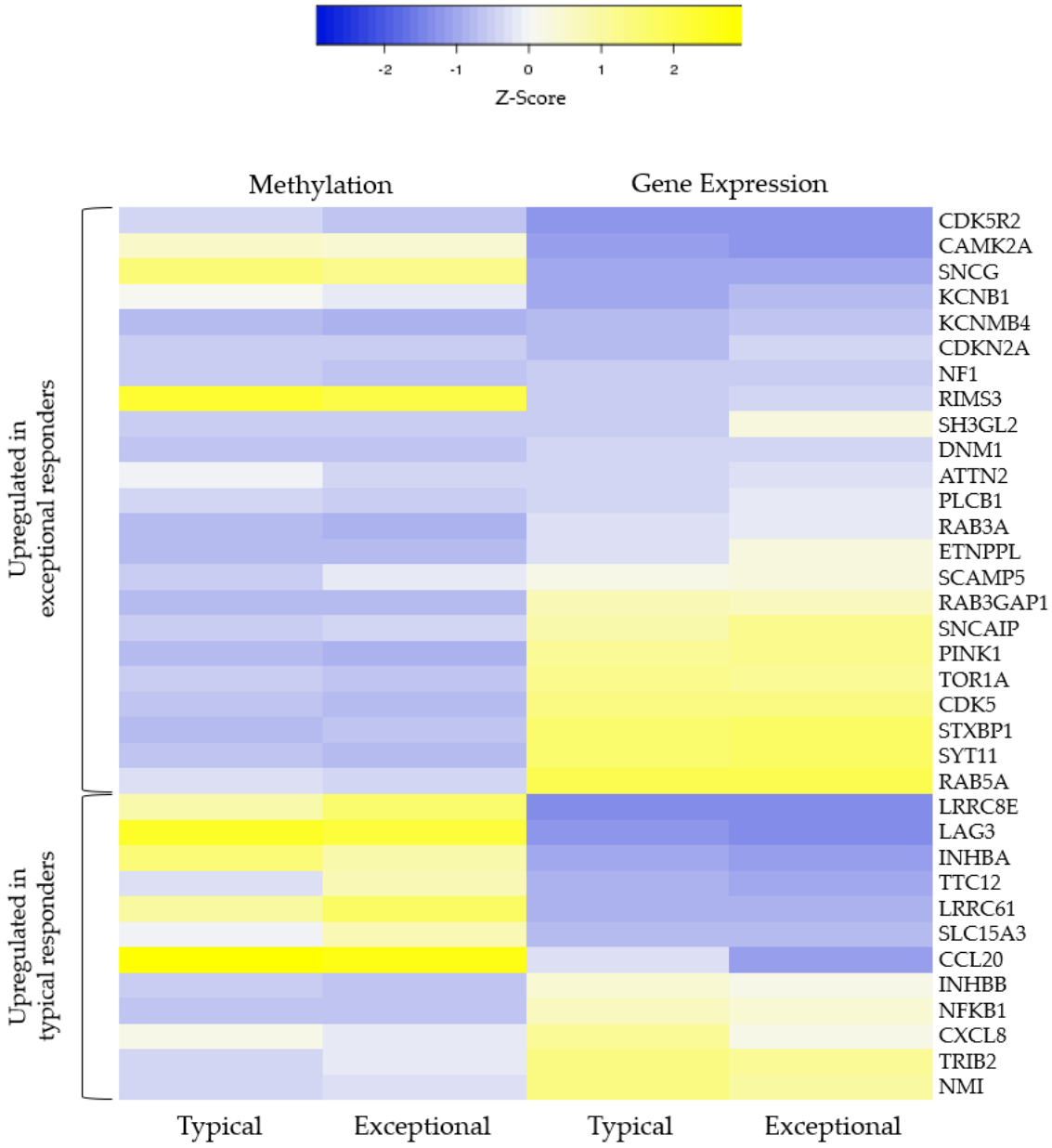

Figure 13: Heat maps for methylation and gene expression. Heat maps were generated with all genes from Table 15 that had data for both methylation and gene expression. Heat maps for both typical and exceptional responders indicate a negative correlation between methylation and gene expression. This correlation appears to be stronger in the list of genes that are upregulated in typical responders.

Typical		
	Copy Number & Gene Expression	**Methylation & Gene Expression**
Pearson (r)	0.2333	-0.4484
Spearman (ρ)	0.2496	-0.5174
Exceptional		
	Copy Number & Gene Expression	**Methylation & Gene Expression**
Pearson (r)	0.2120	-0.6065
Spearman (ρ)	0.2370	-0.5857

Table 21: Correlation between copy number/methylation and gene expression. Each correlation test was performed for the list of genes in Table 15, excluding those that did not have data available for the relevant data type. The results indicate a weak to moderate positive correlation between copy number and gene expression and a strong negative correlation between methylation and gene expression.

variables, while the Spearman test is more suited to categorical variables.[95] β values are a continuous variable, so the Pearson test seems more appropriate, but the Spearman test generally gives results of a larger magnitude for correlations between methylation and gene expression, which is perhaps why it is used so often for those types of studies. It is possible that the Spearman test typically gives better results because methylation may be more categorical than continuous in reality, functioning more like an on/off switch than in a linear manner. In any case, both tests were utilized in this study to provide a clearer understanding of the relationship between methylation and gene expression for this dataset.

Enriched GO Terms

The main contributors to the enriched GO terms for typical responders are IL6 (interleukin 6), which is associated with all three terms, and BCL3 (B-cell lymphoma 3), NFKB1 (nuclear factor kappa B subunit 1), and TRIB2 (tribbles pseudokinase 2), which are associated with two terms each. All four of these genes are associated with NF-κB, a protein complex that acts as a transcription factor and plays a role in cytokine production and cell survival. NF-κB is constitutively active in many cancers, causing cells to proliferate and protecting them from death by apoptosis.[96] Activation of NF-κB in GBM has been shown to contribute to angiogenesis and temozolomide resistance.[97] Typical responders are characterized by a significant upregulation of key players in NF-κB signaling relative to exceptional responders, which may explain their poorer prognosis. An alternative way to view this is that exceptional responders are characterized by less

NF-κB signaling, allowing them to remain more sensitive to temozolomide and resulting in longer survival times.

Nearly all of the GO terms enriched in exceptional responders are associated with synaptic plasticity. The process of autophagy intersects with many of the pathways known to underlie synaptic plasticity, and it has been proposed that autophagy plays a direct role in synaptic plasticity.[98,99] Perhaps the tumors of exceptional responders are more susceptible to cell death by autophagy, or some other consequence of increased synaptic plasticity may confer some advantage.

SNCA (synuclein alpha), which is associated with eight of the enriched GO terms, may be of particular interest as it has been shown to increase the vulnerability of the GBM cell line U373 to cell death.[100] SYT11 is of interest as well, as its depletion is known to block autophagy.[101] The upregulation of SYT11 in exceptional responders may allow autophagy to occur.

Enriched KEGG Pathways

The top enriched pathways in typical responders are IL-17 and TNF. IL-17 is a pro-inflammatory pathway that can contribute to tumor progression and metastasis as well as resistance to chemotherapy.[102] TNF activates the NF-κB pathway and can promote cell growth, proliferation, invasion, and angiogenesis.[103] Enrichment of either of these pathways could contribute to the poorer prognosis seen in the typical response group. Cytokine-cytokine receptor interaction and chemokine signaling pathways are also implicated, much like in the GO analysis.

In exceptional responders, synaptic transmission appears again, along with MAPK signaling, calcium signaling, Parkinson's disease, and Alzheimer's disease, all of which are related to synaptic plasticity.[104–107] Parkinson's disease pathways being enriched in exceptional responders is of particular interest, as several epidemiological studies indicate an inverse association between cancer risk and Parkinson's disease.[108] Like synaptic plasticity, the enrichment of Parkinson's and Alzheimer's pathways also implicates autophagy since both of those diseases are characterized by neurodegeneration in which autophagy plays a role.[109,110] Perhaps the upregulation of these pathways in exceptional responders makes the tumors more susceptible to autophagic cell death.

Correlation Results

No gene was identified as statistically significant among the results of more than one data type. However, the trends seem to match and the correlation results corroborate this. In many cases, the r and ρ values reported in this study would be considered weak to moderate, but in the case of the correlation between methylation and gene expression, correlation coefficients with the magnitudes identified in this study are quite good. Typically, correlation coefficients for methylation and gene expression are rather modest, right around -0.3, which has been attributed to noise, sample heterogeneity, and other regulatory events besides methylation.[111] This means that the relationship between methylation and gene expression in this study is a strong negative correlation, as the correlation coefficients are of a larger magnitude than 0.3, especially when the gene list being assessed is narrowed to only the significant results list. The correlation between

copy number and gene expression is not as strong, but this is not surprising given the large number of other regulatory mechanisms that could be affecting gene expression, as well as the fact that nearly all of the significant copy number aberrations were only gains or losses and not whole duplications or deletions.

DISCUSSION

Sex as a Confounding Variable

Although sex was a confounding variable for survival between the typical and exceptional response groups, it was not confounding in the full dataset. Although the incidence of GBM is higher in males, sex has not been found to be predictive of prognosis or survival in GBM.[112] For these reasons, it seems likely that the identification of sex as a confounding variable is an artifact or just occurred by chance for this particular dataset.

Although it is unlikely that sex is truly predictive of survival in GBM, the distribution of sexes between the two response groups was significantly skewed nonetheless, and this had to be accounted for. When possible, this problem was addressed by performing sex-specific analyses in addition to the full analysis of typical versus exceptional responders (as in the copy number and expression analyses), and then only results identified in all three analyses or only in the full analysis and not in the sex-specific analyses were included in the final results. When there was not a sufficient number of samples available to perform sex-specific analyses (as in the methylation analysis), sex chromosomes were excluded from the analysis. These efforts seem to have been successful in controlling for the differential distribution of sex in the response groups. Prior to performing the methods to control for sex, many results were genes on the X or Y chromosome or were otherwise associated with sex. This was particularly true in the methylation analysis, in which most of the results were on the X chromosome prior to controlling for sex. This was likely due to X inactivation in females, which is the process

by which one X chromosome is silenced via heavy methylation.[113] After the methods to control for the differential distribution of sexes were applied, the final results for all the data types were largely biologically relevant and not associated with any sex-specific biological processes.

Data Types with No Significant Results

Of all the data types analyzed, the miRNA expression analysis and the protein expression analysis were the only ones that did not yield any significantly differentially expressed results. In the case of protein expression, this is likely largely due to both the small sample size and the relatively small number of proteins assessed. While RPPA assesses expression of quite a large number of proteins compared to other techniques such as western blots, it is still only a very small fraction of the proteome (222 antibodies). The only overlap between the significant results from other data types and the list of proteins assessed by RPPA is CDKN2A, which had a greater magnitude of copy number loss in typical responders. Protein expression results indicate that typical responders had a lower mean relative expression of CDKN2A (0.496 for typical and 1.095 for exceptional), as expected based on the copy number analysis, but this result was not statistically significant. Between the low degree of overlap and the small number of patients for whom protein expression data were available, it is not surprising that no significant results were identified. As for the miRNA expression, the sample size was suitable and the number of miRNAs assessed was reasonable, and yet no significant results were identified. While it has been established that aberrant expression of a variety of miRNAs relative to normal

expression is characteristic of GBM overall, it is possible that these are simply not differentially expressed between the response groups in this dataset.

Trend of Disorder in Typical Responders

The results from several analyses suggest that typical responders are characterized by a loss of transcriptional control and aberrations of a greater magnitude than exceptional responders. The copy number analysis revealed that while most copy number alterations are consistent across both response groups, the magnitude is consistently larger in the typical responders. Methylation levels at CpG sites with a large degree of variation between the response groups are almost invariably lower in typical responders, suggesting that exceptional responders have increased gene silencing and transcriptional control. While both response groups tend to be characterized by many of the same alterations, the changes tend to be more severe in typical responders. The somewhat less severe alterations in exceptional responders may be contributing to their better prognosis.

Correlation Trends

No single gene was identified as statistically significant in more than one data type. However, the trends for most of the significant results are concordant. For example, genes with significantly higher promoter methylation tend to have lower expression levels even though the expression comparison does not reach statistical significance. Genes with copy number gains tend to have higher expression, genes with copy number losses tend to have lower expression, and so on. The negative correlation between

methylation and gene expression among the significant results list is particularly convincing, with r and ρ values ranging from -0.4484 to -0.6065, indicating a strong relationship between increased methylation and reduced gene expression among genes that are altered between typical and exceptional responders.

Frequently Affected Pathways and Biological Processes

NF-κB

NF-κB was determined to be upregulated in typical responders relative to exceptional responders in several analyses. Four of the five groups of genes associated with enriched KEGG pathways in typical responders include NFKB1, a key player in the NF-κB pathway. One of these enriched pathways is the TNF signaling pathway, which activates NF-κB signaling, resulting in proliferation and protection from death by apoptosis.[103] NFKB1 was also an important player in two of the three enriched GO terms in typical responders. Several other genes associated with NF-κB signaling were present in the significant results, including IL6, BCL3, and TRIB2. Enrichment of this pathway in typical responders may be partially responsible for their worse prognosis. Upregulation of the NF-κB pathway is common in many cancers, including GBM, so perhaps a better way to view this result is that exceptional responders tend to have less upregulation of NF-κB than most GBM patients, providing them with a survival benefit.

Synaptic Plasticity

The top KEGG pathway result, nearly all of the GO terms, and all of the GSEA gene sets enriched in exceptional responders are directly related to synaptic plasticity, which is the ability of neurons to change the quantity and strength of their synapses. Astrocytes are heavily involved in this process.[114] Autophagy pathways and synaptic plasticity pathways have a lot of overlap, and it has been proposed that autophagy plays a direct role in synaptic plasticity.[98,99] Perhaps the upregulation of genes related to synaptic transmission in exceptional responders leaves their tumor cells more susceptible to autophagy. Glioblastoma cells are more likely to respond to autophagy-inducing therapies than to apoptosis-inducing therapies[115], and it is possible that this characteristic of expectational responders increases this positive response even further. It may also be the case that some other aspect of synaptic transmission and synaptic plasticity confers a benefit to exceptional responders.

Neurodegenerative Diseases

Both Alzheimer's disease and Parkinson's disease pathways were enriched in exceptional responders based on the KEGG analysis. Both of these diseases are characterized by cell death through autophagy and/or apoptosis.[109,110] Some studies indicate that higher risk of Parkinson's disease is inversely associated with cancer risk.[108] While activation of these pathways is certainly detrimental in neurodegenerative diseases, in cancer it may result in increased sensitivity to treatment. Tumors of exceptional

responders with upregulation of these pathways may retain the ability to undergo cell death in response to treatment, granting exceptional responders a better prognosis.

Cytokine Signaling

Cytokines are a group of small proteins that are important in cell signaling and are especially important in the immune system. Pro-inflammatory cytokines are known to promote cancer cell proliferation in many cases.[116] Several pro-inflammatory cytokines and associated pathways were implicated in typical responders throughout this study, including IL6, CCL20, CXCL8, TNF signaling, and IL-17 signaling. The top enriched GO term in typical responders was the regulation of cytokine biosynthetic processes and enriched KEGG pathways in typical responders included cytokine-cytokine receptor interaction and chemokine signaling pathway, in addition to the aforementioned TNF and IL-17 signaling pathways. CXCL8 and CCL20 are perhaps the most prominent of these results, as they were the only two genes with significantly higher expression in typical responders compared to exceptional, and at least one of them is present in every group of genes associated with the enriched KEGG pathways in typical responders. CXCL8 was also found to be predictive of survival outcome in the full TCGA GBM dataset of 408 patients, with increased expression associated with significantly shorter survival times. Typical responders are characterized by upregulation of pro-inflammatory cytokine signaling, and exceptional responders do not share this trait. This lack of pro-inflammatory signaling in exceptional responders may give them a better prognosis.

Final Remarks

The significant results generated in the copy number, methylation, gene expression, and pathway analyses conducted in this study have provided some insight into the molecular differences between typical and exceptional responders in GBM. Upregulated pathways and processes in typical responders are consistently associated with more aggressive tumor phenotypes that may be partially responsible for the poor response to treatment that most GBM patients exhibit. Upregulated pathways and processes in exceptional responders may indicate that the small number of patients who respond very well to treatment have tumors that have retained the ability to undergo cell death by autophagy, which may make the standard GBM treatment of temozolomide more effective for this group.

REFERENCES

1. Adeberg, S. *et al.* A comparison of long-term survivors and short-term survivors with glioblastoma, subventricular zone involvement: a predictive factor for survival? *Radiat. Oncol. Lond. Engl.* **9,** 95 (2014).
2. Davis, M. E. Glioblastoma: Overview of Disease and Treatment. *Clin. J. Oncol. Nurs.* **20,** S2–S8 (2016).
3. Weller, M., Cloughesy, T., Perry, J. R. & Wick, W. Standards of care for treatment of recurrent glioblastoma—are we there yet? *Neuro-Oncol.* **15,** 4–27 (2013).
4. Friedmann-Morvinski, D. Glioblastoma heterogeneity and cancer cell plasticity. *Crit. Rev. Oncog.* **19,** 327–336 (2014).
5. Glioblastoma (GBM) | American Brain Tumor Association. Available at: http://www.abta.org/brain-tumor-information/types-of-tumors/glioblastoma.html. (Accessed: 23rd August 2017)
6. García-Cabezas, M. Á., John, Y. J., Barbas, H. & Zikopoulos, B. Distinction of Neurons, Glia and Endothelial Cells in the Cerebral Cortex: An Algorithm Based on Cytological Features. *Front. Neuroanat.* **10,** (2016).
7. Sofroniew, M. V. & Vinters, H. V. Astrocytes: biology and pathology. *Acta Neuropathol. (Berl.)* **119,** 7–35 (2010).
8. Taylor, T. E., Furnari, F. B. & Cavenee, W. K. Targeting EGFR for Treatment of Glioblastoma: Molecular Basis to Overcome Resistance. *Curr. Cancer Drug Targets* **12,** 197–209 (2012).
9. The Cancer Genome Atlas Home Page. *The Cancer Genome Atlas - National Cancer Institute* Available at: https://cancergenome.nih.gov/. (Accessed: 23rd August 2017)
10. Program Overview. *The Cancer Genome Atlas - National Cancer Institute* Available at: https://cancergenome.nih.gov/abouttcga/overview. (Accessed: 23rd August 2017)

11. The Cancer Genome Atlas - Data Portal. Available at: https://tcga-data.nci.nih.gov/docs/publications/tcga/? (Accessed: 23rd August 2017)

12. GDC. Available at: https://portal.gdc.cancer.gov/. (Accessed: 23rd August 2017)

13. Printz, C. NCI launches exceptional responders initiative: Researchers will attempt to identify why some patients respond to treatment so much better than others. *Cancer* **121,** 803–804 (2015).

14. Exceptional Responders Initiative: Questions and Answers. *National Cancer Institute* Available at: https://www.cancer.gov/about-cancer/treatment/research/exceptional-responders-initiative-qa. (Accessed: 23rd August 2017)

15. Major Initiatives | DCTD. Available at: https://dctd.cancer.gov/majorinitiatives/NCI-sponsored_trials_in_precision_medicine.htm#h06. (Accessed: 23rd August 2017)

16. Exceptional Responders Initiative Moving Forward. *National Cancer Institute* Available at: https://www.cancer.gov/news-events/cancer-currents-blog/2017/exceptional-responders-progress. (Accessed: 23rd August 2017)

17. De La Torre, K., Cohen, E., Loeser, A. & Hurlbert, M. Moonshots and metastatic disease: the need for a multi-faceted approach when studying atypical responses. *NPJ Breast Cancer* **3,** (2017).

18. Ma, J. *et al.* Genomic analysis of exceptional responders to radiotherapy reveals somatic mutations in ATM. *Oncotarget* **8,** 10312–10323 (2016).

19. Perakslis, E. D. & Kohane, I. S. Treating the enigmatic 'exceptional responders' as patients with undiagnosed diseases. *Sci. Transl. Med.* **8,** 340ed8-340ed8 (2016).

20. Analyses of IDH1 mutation and MGMT promoter methylation status for 5 cases of long-term survivors with glioblastoma. *ResearchGate* Available at: https://www.researchgate.net/publication/221781313_Analyses_of_IDH1_mutation_and_M

GMT_promoter_methylation_status_for_5_cases_of_long-term_survivors_with_glioblastoma. (Accessed: 23rd August 2017)

21. Sjöström, S. *et al.* Genetic variations in EGF and EGFR and glioblastoma outcome. *Neuro-Oncol.* **12,** 815–821 (2010).

22. Hartmann, C. *et al.* Long-Term Survival in Primary Glioblastoma With Versus Without Isocitrate Dehydrogenase Mutations. *Clin. Cancer Res.* **19,** 5146–5157 (2013).

23. Kraus, J. A. *et al.* Molecular analysis of the PTEN, TP53 and CDKN2A tumor suppressor genes in long-term survivors of glioblastoma multiforme. *J. Neurooncol.* **48,** 89–94 (2000).

24. Krex, D. *et al.* Long-term survival with glioblastoma multiforme. *Brain J. Neurol.* **130,** 2596–2606 (2007).

25. Ohgaki, H. & Kleihues, P. Genetic Pathways to Primary and Secondary Glioblastoma. *Am. J. Pathol.* **170,** 1445–1453 (2007).

26. Sturm, D. *et al.* Paediatric and adult glioblastoma: multiform (epi)genomic culprits emerge. *Nat. Rev. Cancer* **14,** 92–107 (2014).

27. Hodgson, J. G. *et al.* Comparative analyses of gene copy number and mRNA expression in glioblastoma multiforme tumors and xenografts. *Neuro-Oncol.* **11,** 477–487 (2009).

28. Freire, P. *et al.* Exploratory Analysis of the Copy Number Alterations in Glioblastoma Multiforme. *PLOS ONE* **3,** e4076 (2008).

29. del Mar Inda, M. *et al.* Chromosomal abnormalities in human glioblastomas: Gain in chromosome 7p correlating with loss in chromosome 10q. *Mol. Carcinog.* **36,** 6–14 (2003).

30. Geisenberger, C. *et al.* Molecular profiling of long-term survivors identifies a subgroup of glioblastoma characterized by chromosome 19/20 co-gain. *Acta Neuropathol. (Berl.)* **130,** 419–434 (2015).

31. Wemmert, S. *et al.* Patients with High-Grade Gliomas Harboring Deletions of Chromosomes 9p and 10q Benefit from Temozolomide Treatment. *Neoplasia N. Y. N* **7,** 883–893 (2005).

32. Crespo, I. *et al.* Detailed Characterization of Alterations of Chromosomes 7, 9, and 10 in Glioblastomas as Assessed by Single-Nucleotide Polymorphism Arrays. *J. Mol. Diagn. JMD* **13,** 634–647 (2011).

33. Genome-Wide Human SNP Array 6.0 - Thermo Fisher Scientific. Available at: https://www.thermofisher.com/order/catalog/product/901182. (Accessed: 23rd August 2017)

34. Mayrhofer, M., Viklund, B. & Isaksson, A. Rawcopy: Improved copy number analysis with Affymetrix arrays. *Sci. Rep.* **6,** srep36158 (2016).

35. Verhaak, R. G. W. *et al.* An integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR and NF1. *Cancer Cell* **17,** 98 (2010).

36. Bambury, R. M. *et al.* DNA copy number analysis of metastatic urothelial carcinoma with comparison to primary tumors. *BMC Cancer* **15,** (2015).

37. KANG, J. Genomic alterations on 8p21-p23 are the most frequent genetic events in stage I squamous cell carcinoma of the lung. *Exp. Ther. Med.* **9,** 345–350 (2015).

38. Wong, K. K. *et al.* A comprehensive analysis of common copy-number variations in the human genome. *Am. J. Hum. Genet.* **80,** 91–104 (2007).

39. Nazor, K. L. *et al.* Recurrent Variations in DNA Methylation in Human Pluripotent Stem Cells and their Differentiated Derivatives. *Cell Stem Cell* **10,** 620–634 (2012).

40. Nazarenko, M. S. *et al.* A Comparison of Genome-Wide DNA Methylation Patterns between Different Vascular Tissues from Patients with Coronary Heart Disease. *PLoS ONE* **10,** (2015).

41. Li, Z., Zhuang, X., Zeng, J. & Tzeng, C.-M. Integrated Analysis of DNA Methylation and mRNA Expression Profiles to Identify Key Genes in Severe Oligozoospermia. *Front. Physiol.* **8,** (2017).

42. Young, J. M. *et al.* Extensive Copy-Number Variation of the Human Olfactory Receptor Gene Family. *Am. J. Hum. Genet.* **83,** 228–242 (2008).

43. Gao, X. *et al.* MiR-15a, miR-16-1 and miR-17-92 cluster expression are linked to poor prognosis in multiple myeloma. *Leuk. Res.* **36,** 1505–1509 (2012).

44. Cao, J. *et al.* Correlations between serum trypsinogen-2 and pancreatic cancer. *Hepatogastroenterology.* **62,** 435–440 (2015).

45. Vilen, S.-T. *et al.* Trypsin-2 enhances carcinoma invasion by processing tight junctions and activating ProMT1-MMP. *Cancer Invest.* **30,** 583–592 (2012).

46. Williams, S. J., Gotley, D. C. & Antalis, T. M. Human trypsinogen in colorectal cancer. *Int. J. Cancer* **93,** 67–73 (2001).

47. Lim, D. H. K. & Maher, E. R. DNA methylation: a form of epigenetic control of gene expression. *Obstet. Gynaecol.* **12,** 37–42 (2010).

48. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11,** 587 (2010).

49. Wilhelm-Benartzi, C. S. *et al.* Review of processing and analysis methods for DNA methylation array data. *Br. J. Cancer* **109,** 1394–1402 (2013).

50. Bundo, M. *et al.* A systematic evaluation of whole genome amplification of bisulfite-modified DNA. *Clin. Epigenetics* **4,** 22 (2012).

51. Hegi, M. E. *et al.* MGMT Gene Silencing and Benefit from Temozolomide in Glioblastoma. *N. Engl. J. Med.* **352,** 997–1003 (2005).

52. Bibikova, M. *et al.* Genome-wide DNA methylation profiling using Infinium® assay. *Epigenomics* **1,** 177–200 (2009).

53. Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* **11,** 1138–1140 (2014).

54. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Methodol.* **57,** 289–300 (1995).

55. Kozlenkov, A. *et al.* Differences in DNA methylation between human neuronal and glial cells are concentrated in enhancers and non-CpG sites. *Nucleic Acids Res.* **42,** 109–127 (2014).

56. Bibikova, M. *et al.* High density DNA methylation array with single CpG site resolution. *Genomics* **98,** 288–295 (2011).

57. Kong, H. K. *et al.* Epigenetic activation of LY6K predicts the presence of metastasis and poor prognosis in breast carcinoma. *Oncotarget* **7,** 55677–55689 (2016).

58. Hossiny, M. A. *et al.* Ly6E/K signaling to TGF-β promotes breast cancer progression, immune escape and drug resistance. *Cancer Res.* **76,** 3376–3386 (2016).

59. Vasiljević, N. *et al.* DNA methylation gene-based models indicating independent poor outcome in prostate cancer. *BMC Cancer* **14,** (2014).

60. Rousalova, I., Krepela, E., Prochazka, J., Cermak, J. & Benkova, K. Expression of proteinase inhibitor-9/serpinB9 in non-small cell lung carcinoma cells and tissues. *Int. J. Oncol.* **36,** 275–283 (2010).

61. Fritsch, K., Finke, J. & Grüllich, C. Suppression of granzyme B activity and caspase-3 activation in leukaemia cells constitutively expressing the protease inhibitor 9. *Ann. Hematol.* **92,** 1603–1609 (2013).

62. Li, W. *et al.* Carboxyl-terminal truncated HBx contributes to invasion and metastasis via deregulating metastasis suppressors in hepatocellular carcinoma. *Oncotarget* **7,** 55110–55127 (2016).

63. GeneChip Human Genome U133 Plus 2.0 Array - Thermo Fisher Scientific. Available at: https://www.thermofisher.com/order/catalog/product/900466. (Accessed: 11th September 2017)

64. Wu, Z., Irizarry, R. A., Gentleman, R., Martinez-Murillo, F. & Spencer, F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *J. Am. Stat. Assoc.* **99,** 909–917 (2004).

65. Wu, Z. J. & Irizarry, R. Description of gcrma package. (2010).

66. Gentleman, R. allNA A filter function to determine if all elements of a vector are NA. (2011).

67. Smyth, G. K., Ritchie, M. & Thorne, N. James Wettenhall, Wei Shi and Yifang Hu Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Melbourne, Australia.

68. Phipson, B., Lee, S., Majewski, I. J., Alexander, W. S. & Smyth, G. K. ROBUST HYPERPARAMETER ESTIMATION PROTECTS AGAINST HYPERVARIABLE GENES AND IMPROVES POWER TO DETECT DIFFERENTIAL EXPRESSION. *Ann. Appl. Stat.* **10,** 946–963 (2016).

69. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).

70. Hoelzinger, D. B. *et al.* Gene Expression Profile of Glioblastoma Multiforme Invasive Phenotype Points to New Therapeutic Targets. *Neoplasia N. Y. N* **7,** 7–16 (2005).

71. Gabellini, C. *et al.* Involvement of nuclear factor-kappa B in bcl-xL-induced interleukin 8 expression in glioblastoma. *J. Neurochem.* **107,** 871–882 (2008).

72. Giorgini, S. *et al.* Modulation of bcl-xL in tumor cells regulates angiogenesis through CXCL8 expression. *Mol. Cancer Res. MCR* **5,** 761–771 (2007).

73. Brat, D. J., Bellail, A. C. & Van Meir, E. G. The role of interleukin-8 and its receptors in gliomagenesis and tumoral angiogenesis. *Neuro-Oncol.* **7,** 122–133 (2005).

74. Shen, T. *et al.* CXCL8 induces epithelial-mesenchymal transition in colon cancer cells via the PI3K/Akt/NF-κB signaling pathway. *Oncol. Rep.* **37,** 2095–2100 (2017).

75. Wang, L. *et al.* Overexpression of CCL20 and its receptor CCR6 predicts poor clinical prognosis in human gliomas. *Med. Oncol. Northwood Lond. Engl.* **29,** 3491–3497 (2012).

76. Muscella, A., Vetrugno, C. & Marsigliante, S. CCL20 promotes migration and invasiveness of human cancerous breast epithelial cells in primary culture. *Mol. Carcinog.* (2017). doi:10.1002/mc.22693

77. Lu, E., Su, J., Zhou, Y., Zhang, C. & Wang, Y. CCL20/CCR6 promotes cell proliferation and metastasis in laryngeal cancer by activating p38 pathway. *Biomed. Pharmacother. Biomedecine Pharmacother.* **85,** 486–492 (2017).

78. Cheng, X.-S. *et al.* CCL20 and CXCL8 synergize to promote progression and poor survival outcome in patients with colorectal cancer by collaborative induction of the epithelial-mesenchymal transition. *Cancer Lett.* **348,** 77–87 (2014).

79. Ding, Q. *et al.* AGXT2L1 is down-regulated in heptocellular carcinoma and associated with abnormal lipogenesis. *J. Clin. Pathol.* **69,** 215–220 (2016).

80. Qu, S. *et al.* MicroRNA-330 is an oncogenic factor in glioblastoma cells by regulating SH3GL2 gene. *PloS One* **7,** e46010 (2012).

81. Russo, R. C., Garcia, C. C., Teixeira, M. M. & Amaral, F. A. The CXCL8/IL-8 chemokine family and its receptors in inflammatory diseases. *Expert Rev. Clin. Immunol.* **10,** 593–619 (2014).

82. Early Infantile Epileptic Encephalopathy (EIEE) Sequencing Panel | The University of Chicago Genetic Services. Available at: http://dnatesting.uchicago.edu/tests/early-infantile-epileptic-encephalopathy-eiee-sequencing-panel. (Accessed: 11th September 2017)

83. Tavyev Asher, Y. J. & Scaglia, F. Molecular bases and clinical spectrum of early infantile epileptic encephalopathies. *Eur. J. Med. Genet.* **55,** 299–306 (2012).

84. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15,** 509–524 (2014).

85. Rolle, K. miRNA Multiplayers in glioma. From bench to bedside. *Acta Biochim. Pol.* **62,** 353–365 (2015).

86. Human miRNA Microarrays. Available at: http://www.genomics.agilent.com/en/miRNA-Microarrays/SurePrint-Human-miRNA-Microarrays/?cid=AG-PT-147. (Accessed: 11th September 2017)

87. Charboneau, L. *et al.* Utility of reverse phase protein arrays: applications to signalling pathways and human body arrays. *Brief. Funct. Genomic. Proteomic.* **1,** 305–315 (2002).

88. Functional Proteomics RPPA Core Facility. *MD Anderson Cancer Center* Available at: https://www.mdanderson.org/research/research-resources/core-facilities/functional-proteomics-rppa-core.html. (Accessed: 28th July 2017)

89. Stegh, A. H. & DePinho, R. A. Beyond effector caspase inhibition: Bcl2L12 neutralizes p53 signaling in glioblastoma. *Cell Cycle Georget. Tex* **10,** 33–38 (2011).

90. Welsh, J. W. *et al.* Rad51 Protein Expression and Survival in Patients with Glioblastoma Multiforme. *Int. J. Radiat. Oncol.* **74,** 1251–1255 (2009).

91. Jin, L. *et al.* Pathway-based Analysis Tools for Complex Diseases: A Review. *Genomics Proteomics Bioinformatics* **12,** 210–220 (2014).

92. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25,** 25–29 (2000).

93. Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).

94. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25,** 1091–1093 (2009).

95. Comparison of Values of Pearson's and Spearman's Correlation Coefficients on the Same Sets of Data - ProQuest. Available at: https://search.proquest.com/openview/e6d3a1d8978b8091d16d1d172197120c/1?pq-origsite=gscholar&cbl=2026615. (Accessed: 23rd August 2017)

96. Hayden, M. S. & Ghosh, S. NF- B, the first quarter-century: remarkable progress and outstanding questions. *Genes Dev.* **26,** 203–234 (2012).

97. WANG, X. *et al.* NF-κB inhibitor reverses temozolomide resistance in human glioma TR/U251 cells. *Oncol. Lett.* **9,** 2586–2590 (2015).

98. Shen, W. & Ganetzky, B. Autophagy promotes synapse development in *Drosophila*. *J. Cell Biol.* **187,** 71–79 (2009).

99. Shehata, M. & Inokuchi, K. Does autophagy work in synaptic plasticity and memory? *Rev. Neurosci.* **25,** 543–557 (2014).

100. Stefanova, N. *et al.* Tumor necrosis factor-alpha-induced cell death in U373 cells overexpressing alpha-synuclein. *J. Neurosci. Res.* **73,** 334–340 (2003).

101. Bento, C. F., Ashkenazi, A., Jimenez-Sanchez, M. & Rubinsztein, D. C. The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway. *Nat. Commun.* **7,** 11803 (2016).

102. Fabre, J. *et al.* Targeting the Tumor Microenvironment: The Protumor Effects of IL-17 Related to Cancer Type. *Int. J. Mol. Sci.* **17,** (2016).

103. Wu, Y. & Zhou, B. P. TNF-α/NF-κB/Snail pathway in cancer cell migration and invasion. *Br. J. Cancer* **102,** 639–644 (2010).

104. Thomas, G. M. & Huganir, R. L. MAPK cascade signalling and synaptic plasticity. *Nat. Rev. Neurosci.* **5,** 173–183 (2004).

105. HENNEBERGER, C. & RUSAKOV, D. A. Synaptic plasticity and Ca2+ signalling in astrocytes. *Neuron Glia Biol.* **6,** 141–146 (2010).

106. Picconi, B., Piccoli, G. & Calabresi, P. Synaptic dysfunction in Parkinson's disease. *Adv. Exp. Med. Biol.* **970,** 553–572 (2012).

107. Skaper, S. D., Facci, L., Zusso, M. & Giusti, P. Synaptic Plasticity, Dementia and Alzheimer Disease. *CNS Neurol. Disord. Drug Targets* (2017).

108. Feng, D. D., Cai, W. & Chen, X. The associations between Parkinson's disease and cancer: the plot thickens. *Transl. Neurodegener.* **4,** 20 (2015).

109. Lynch-Day, M. A., Mao, K., Wang, K., Zhao, M. & Klionsky, D. J. The Role of Autophagy in Parkinson's Disease. *Cold Spring Harb. Perspect. Med.* **2,** (2012).

110. Tsujimoto, Y. & Shimizu, S. Another way to die: autophagic programmed cell death. *Cell Death Differ.* **12,** 1528–1534 (2005).

111. VanderKraats, N. D., Hiken, J. F., Decker, K. F. & Edwards, J. R. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic Acids Res.* **41,** 6816–6827 (2013).

112. Filippini, G. *et al.* Prognostic factors for survival in 676 consecutive patients with newly diagnosed primary glioblastoma. *Neuro-Oncol.* **10,** 79–87 (2008).

113. Sharp, A. J. *et al.* DNA methylation profiles of human active and inactive X chromosomes. *Genome Res.* **21,** 1592–1600 (2011).

114. Citri, A. & Malenka, R. C. Synaptic Plasticity: Multiple Forms, Functions, and Mechanisms. *Neuropsychopharmacology* **33,** 18–41 (2007).

115. Jiang, H., White, E. J., Conrad, C., Gomez-Manzano, C. & Fueyo, J. Chapter 13 Autophagy Pathways in Glioblastoma. *Methods Enzymol.* **453,** 273–286 (2009).

116. Turner, M. D., Nedjai, B., Hurst, T. & Pennington, D. J. Cytokines and chemokines: At the crossroads of cell signalling and inflammatory disease. *Biochim. Biophys. Acta BBA - Mol. Cell Res.* **1843,** 2563–2582 (2014).