

# UC Santa Cruz

## UC Santa Cruz Electronic Theses and Dissertations

### Title

Using High-Throughput Sequencing To Study The Progression Of Cancer

### Permalink

<https://escholarship.org/uc/item/4w21q3c3>

### Author

Grifford, Mia Susanne

### Publication Date

2014

### Supplemental Material

<https://escholarship.org/uc/item/4w21q3c3#supplemental>

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

SANTA CRUZ

**USING HIGH-THROUGHPUT SEQUENCING TO STUDY THE  
PROGRESSION OF CANCER**

A dissertation submitted in partial satisfaction of the  
requirements for the degree of

DOCTOR OF PHILOSOPHY

in

BIOMOLECULAR ENGINEERING AND BIOINFORMATICS

by

**Mia Grifford**

December 2014

The Dissertation of Mia Grifford  
is approved:

---

Professor David Haussler, Chair

---

Professor Lindsay Hinck

---

Sofie Salama, Ph.D.

---

Tyrus Miller  
Vice Provost and Dean of Graduate Studies

Copyright © by

Mia Grifford

2014

# Table of Contents

<b>List of Figures</b>	<b>vi</b>
<b>List of Tables</b>	<b>viii</b>
<b>Abstract</b>	<b>ix</b>
<b>Dedication</b>	<b>xi</b>
<b>Acknowledgments</b>	<b>xii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Overview . . . . .	1
1.2 Background . . . . .	3
1.2.1 Detecting alterations in cancer with high-throughput technologies	3
1.2.2 Cancer types . . . . .	5
<b>2 Double Minute Chromosomes in Glioblastoma Multiforme Are Revealed by Precise Reconstruction of Oncogenic Amplicons</b>	<b>11</b>
2.1 Introduction . . . . .	13
2.2 Methods . . . . .	16
2.2.1 Tumor and normal genome sequencing data . . . . .	16
2.2.2 Determining breakpoints related to highly amplified regions . . .	16
2.2.3 Reconstructing amplicons by walking a breakpoint graph . . . .	17
2.2.4 RNA-Seq analysis . . . . .	19
2.2.5 FISH analysis . . . . .	19
2.2.6 Exome analysis to estimate prevalence of double minute chromosomes . . . . .	20
2.2.7 Association of double minute chromosomes/HSR samples with other glioblastoma multiforme tumor features . . . . .	21
2.3 Results . . . . .	21
2.3.1 BamBam: a robust method for identifying tumor-specific variation	21

2.3.2	Reconstruction of candidate double minute chromosomes using whole-genome sequencing . . . . .	22
2.3.3	Transcriptome data reveals a novel double minute chromosome-associated fusion protein . . . . .	28
2.3.4	TCGA-06-0648 and TCGA-06-0145 amplicons exist as double minute chromosomes . . . . .	31
2.3.5	Prevalence of putative double minute chromosomes/HSRs in exome sequencing data . . . . .	32
2.3.6	Validation of exome sequencing-based prediction of double minute chromosomes/HSR-containing samples . . . . .	34
2.4	Discussion . . . . .	36
<b>3</b>	<b>Comprehensive and Integrative Genomic Characterization of Diffuse Lower Grade Gliomas</b>	<b>40</b>
3.1	Introduction . . . . .	41
3.2	Methods . . . . .	43
3.2.1	Patients . . . . .	43
3.2.2	Analytic Platforms . . . . .	43
3.3	Results . . . . .	44
3.3.1	Histology and molecular subtype . . . . .	44
3.3.2	Multiplatform integrative analysis . . . . .	45
3.3.3	Mutational landscape of LGG . . . . .	49
3.3.4	Signaling networks in LGG . . . . .	52
3.3.5	IDHwt LGG and GBM . . . . .	53
3.3.6	Genomic rearrangements and fusion transcripts . . . . .	55
3.3.7	LGG protein expression . . . . .	56
3.3.8	From signatures to pathways . . . . .	57
3.3.9	Clinical characteristics and outcomes of molecular subtypes . . . . .	59
3.4	Discussion . . . . .	60
<b>4</b>	<b>Mutation Independent Activation of The Notch Pathway is Associated With Lapatinib Resistance in Her2+ Breast Cancer Cell Lines</b>	<b>66</b>
4.1	Introduction . . . . .	67
4.2	Methods . . . . .	69
4.2.1	Cell culture and generation of lapatinib resistant cell lines . . . . .	69
4.2.2	Quantitative RT-PCR in drug tolerant persisters . . . . .	70
4.2.3	Immunoblotting . . . . .	71
4.2.4	RNA sequencing library preparation . . . . .	72
4.2.5	RNA sequencing data mapping and analysis . . . . .	73
4.2.6	Gene expression clustering and pathway enrichment analysis . . . . .	74
4.2.7	Accession codes . . . . .	74
4.2.8	Mutation calling with RADIA . . . . .	75
4.3	Results . . . . .	75

4.3.1	Prolonged incubation in lapatinib robustly generates resistant cell lines . . . . .	75
4.3.2	Breast cancer DTPs share features with non-small cell lung cancer mutation-independent resistant cells . . . . .	76
4.3.3	RNA-Sequencing reveals significant changes in gene expression .	78
4.3.4	Resistance is mutation independent . . . . .	80
4.3.5	The Notch pathway is over-expressed in resistant cells . . . . .	82
4.4	Discussion . . . . .	85
<b>5</b>	<b>Conclusions</b>	<b>88</b>
<b>A</b>	<b>Chapter 3 Supplementary Figures</b>	<b>94</b>
<b>B</b>	<b>Chapter 4 Supplementary Figures</b>	<b>98</b>
<b>C</b>	<b>List of supplementary data files</b>	<b>108</b>
	<b>Bibliography</b>	<b>110</b>

# List of Figures

1.1	KM Plot of HER2+ vs. HER2- Patients . . . . .	8
1.2	KM Plot of Patients Grouped by PAM50 Subtype . . . . .	9
2.1	Reconstruction of TCGA-06-0648 double minutes . . . . .	26
2.2	Reconstruction of 06-0145 amplicons. . . . .	29
2.3	Novel CPM C-terminal exon expressed from the TCGA-06-0648 double minute chromosome . . . . .	31
2.4	Visualization of glioblastoma multiforme tumor oncogenic amplicons. . .	32
3.1	Multiplatform analyses point to biologic subtypes defined by IDH muta- tion and 1p/19q codeletion status. . . . .	48
3.2	Muational landscape and unbiased clustering. . . . .	54
3.3	IDHwt LGG resemble IDHwt GBMs . . . . .	58
3.4	Integrative analysis and clinical outcome . . . . .	61
4.1	Lapatinib resistant cell lines were generated by long-term lapatinib treat- ment. . . . .	77
4.2	Breast cancer DTPs share features with non-small cell lung cancer DTPs.	79
4.3	Notch, PI3K and FOS/JUN pathways are over-expressed in DTEPs com- pared to untreated cells. . . . .	83
4.4	Notch activation may lead to altered chromatin state. . . . .	87
A.1	The UCSC Cancer Genomics Browser view of GISTIC thresholded copy number calls for the region of chromosome 9 containing CDKN2A/B. . .	95
A.2	Reconstruction of TCGA-CS-5395 circular amplicon . . . . .	96
A.3	Metrics for cluster of clusters model selection. . . . .	97
B.1	HER2+ lapatinib sensitive breast cancer cell lines used in this study. . .	99
B.2	Drug tolerant expanded persisters begin to grow in 40-70 days. . . . .	100
B.3	DTPs have a distinct chromatin state. . . . .	101
B.4	DE-Seq Finds 6943 Significantly Differentially Expressed Genes. . . . .	102

B.5	DE-Seq reveals that untreated cells, DTPs and DTEPs have significantly different expression patterns, with DTEPs being the most distinct. . . .	103
B.6	Pathways significantly over-expressed in DTEPs are highly interconnected.	104
B.7	Jag1 and Jag2 are under-expressed in HER2+ breast cancer patients. . .	105



# List of Tables

2.1	RNA-Seqbased expression estimates of 06-0648 double minute chromosome-associated genes . . . . .	30
2.2	Summary of amplicons found in TCGA GBM exomes . . . . .	34
3.1	Clinical characteristics of the sample set. . . . .	46
B.1	DE-Seq results confirm that IGF-1R and KDM pathways are over-expressed in resistant cells. . . . .	106
B.2	Genes known to contribute to drug resistance. . . . .	107

## **Abstract**

Using high-throughput sequencing to study the progression of cancer

by

Mia Grifford

Cancer is a deadly disease that affects 1 in 2 men and 1 in 3 women in the United States. The development of high throughput sequencing technologies has allowed for great advancements in cancer therapy, though, we still have a long way to go. Many cancer patients do not receive care soon enough due to late diagnosis, they do not receive the optimum treatment for their specific tumor type or they receive the optimum treatment, but then develop resistance to it. This thesis will address these three problems and describe how high throughput sequencing technology has allowed us to continue to make advancements in cancer treatment. Specifically, I will describe one example of how understanding the mechanism of oncogenesis can lead to new options for early diagnosis and monitoring of disease progression in about a quarter of glioblastoma multiforme and IDH wild-type lower grade glioma patients, harboring double minute chromosomes. I show how comprehensive characterization of lower grade glioma patients has led to the discovery of a new, more accurate method of subtyping patients, which will allow for patients to receive more appropriate treatment for their specific tumor type. This is especially important for one subtype of lower grade glioma patients who have poor

prognosis and appear very similar to glioblastoma multiforme patients, and therefore should receive similar treatment. I will also describe how development of HER2+ breast cancer cell lines that are resistant to the targeted therapy lapatinib, has led to greater understanding of a mutation-independent mechanism of drug resistance involving activation of the Notch pathway. Patients with a similar mechanism of resistance may benefit from treatment with Notch inhibitors, several of which are currently being tested in clinical trials. All of these findings display ways in which high throughput sequencing can be used to make advances in our understanding of cancer, ultimately moving towards better treatment for patients and better patient outcomes.

I dedicate this dissertation to my fiancé Alex,

who continuously loved, supported and encouraged me throughout this process.

## Acknowledgments

I want to thank my committee, Lindsay Hinck, Josh Stuart, Sofie Salama and David Haussler, for their invaluable feedback and support. I especially want to thank Sofie Salama for mentoring me and teaching me so much and David Haussler for his thoughtful insight and encouragement. I want to thank Camilla Forseberg and Scott Boyer for taking me on as a roton in the wetlab, even with no prior experience, and teaching me the fundamentals of wetlab procedures. I also want to thank the many people who guided me and answered all of my questions when I was starting out in the Haussler wet lab, including Bob Sellers, Bryan King, Tracy Ballinger, Dave Greenberg, Courtney Onodera and Frank Jacobs. I want to thank all of the current and past Haussler wet lab members for their camaraderie, guidance and support, including Olena Morozova, Andrew Field, Jimi Rosenkrantz, Mari Olsen and Kristoff Tigyi. I want to thank all of the members of the Haussler/Stuart cancer group and the rest of the Haussler dry lab, especially Zack Sanborn, Steve Benz and Sol Katzman, who've answered my endless questions and continued to guide me and give me invaluable feedback. I want to thank the countless other graduate students who helped me along this journey including Lauren Lui, Arjun Rao, Amie Radenbaugh, Sam Ng, John St. John, Phillip Heller and Wendy Lee. I want to thank all of my teachers and mentors who've truly inspired me both at UCSC and at previous colleges, including Lisa Hunter and Anne Metevier of ISEE, and James Wilson and Heather Shearer in the Writing Program. I lastly want to thank my family and friends who were always there for me and encouraged me to never

give up, without them none of this would have been possible.

# Chapter 1

## Introduction

### 1.1 Overview

Cancer is the second leading cause of death in the United States [40]. Men have a 1 in 2 risk of developing cancer in their lifetime and women have a 1 in 3 risk [133]. Cancer treatment has advanced greatly with the invention of high throughput sequencing technology, which has provided great insights into cancer genomes and allowed for the development of targeted therapies which are more effective and have less harmful side effects than traditional chemotherapy. Though several problems with cancer treatment still exist. This thesis will discuss three key problems with cancer treatment and how high throughput sequencing technology has allowed us to make progress in the areas of diagnostics, subtyping and drug resistance.

One of the biggest problems with cancer treatment is late diagnosis. Many times patients don't know they have cancer until it is too late, and in these cases treatment is less effective than if the tumor had been detected earlier. Chapter 2 of this thesis will discuss one feature of gliomas that occur in about a quarter of glioblastoma multiforme patients that may allow for earlier detection and more effective monitoring of disease progression after diagnosis. These features are small circular amplicons called double minute chromosomes, which may be detectable in the blood. The ability to detect and monitor disease progression through the blood would be especially helpful in patients with tumors in places where surgery and biopsies can be risky, like the brain. The computational method for identifying these double minute chromosomes from whole-genome sequencing data will be discussed, as well as a method for predicting the presence and features of double minute chromosomes using exome sequencing data.

Another problem with cancer treatment is determining which treatments are appropriate for each individual. Patients are often categorized into groups, which determine the optimal treatment. Though, proper treatment can only be given if patients are accurately classified into treatment groups. Chapter 3 will discuss why the current method of categorizing patients in lower grade glioma, by histological type and grade, is not ideal and will propose a new method of subtyping patients that is supported by several independent data types and computational techniques, including high throughput sequencing of both DNA and RNA. This will allow patients to receive the best treatment for their tumor. In particular, this new method of subtyping patients allows for identi-



fication of one group of patients with especially aggressive tumors that behave similarly to glioblastoma multiforme, and should therefore be treated similarly.

When patients are treated with drugs appropriate for their specific tumor type, treatment can be very successful. Often times this involves the use of targeted therapies that directly inhibit proteins critical for tumor growth. These targeted therapies have less side effects than traditional chemotherapy or radiation and can be very effective. Although, there is one critical problem with targeted therapies. Patients often acquire resistance to these drugs and the therapy becomes ineffective. The fourth chapter of this thesis will discuss a mechanism of resistance to the targeted therapy, lapatinib, which was discovered using RNA-sequencing of HER2+ breast cancer cell lines. These cells acquire a mutation independent mechanism of resistance involving activation of the Notch pathway. These findings could explain what is underlying resistance in many patients and provide new treatment options for these patients.

## **1.2 Background**

### **1.2.1 Detecting alterations in cancer with high-throughput technologies**

Cancer occurs when normal functioning of a set of genes goes haywire. Alterations in oncogenes can cause cells to divide uncontrollably and altered tumor suppressor genes

can prevent cells from undergoing apoptosis, or programmed cell death. There are many different types of alterations that can occur in cancer and can either disrupt protein function, cause a protein to signal more than usual or even create novel protein function.

The most well studied genetic alterations are mutations, where a single DNA nucleotide is changed. Alternatively, larger regions of DNA can be either lost (deleted) or duplicated one or more times (amplified). These changes can be as small as a few bases to as large as an entire chromosome, and are referred to as copy number alterations. The DNA can also be broken and reconnected incorrectly. Portions of DNA can be incorrectly connected to other regions of DNA or they can be flipped or reversed. We refer to these types of changes as rearrangements or structural variations. These rearrangements can cause two protein coding genes to connect in novel ways, which are referred to as gene fusions. All of these alterations in DNA can be observed with either whole-genome sequencing (sequencing the entire genome) or exome sequencing (sequencing only the portion of genes that code for proteins). Copy number changes can also be observed using a microarray that is specially designed to measure the amount of DNA throughout the entire genome.

DNA can also be altered epigenetically through histone modification or methylation. DNA methylation typically silences the DNA causing genes to no longer be expressed, while histone modification can cause either silencing or activation of genes. Histone modifications can be observed using ChIP-Seq, combining chromatin immunoprecipita-

tion with sequencing, and DNA methylation is typically measured using methylation arrays.

Instead of looking at the specific genetic or epigenetic alterations, we can also observe the effects of these alterations on both RNA production and proteins. Messenger RNA expression can be observed using RNA sequencing and microRNAs can be observed using microRNA sequencing. We can also observe overall amounts of proteins and phosphorylation of proteins using reverse phase protein arrays.

All of these methods allow us to get a more thorough view of what is happening in the cell and help us to determine what is driving cancer progression.

## **1.2.2 Cancer types**

Several different high-throughput technologies were used in this thesis to study the progression of both glioma and breast cancer.

### **1.2.2.1 Glioma**

Glioma is a type of tumor that arises from glial cells in the brain. Like many other cancer types, they are identified by the cell type that the tumor is thought to have arisen from and tumor aggressiveness is labelled with a grade (I - IV). This thesis will discuss several types of gliomas: astrocytomas, which are thought to have arisen from

astrocytes, oligodendrogliomas, which are thought to have arisen from oligodendrocytes and oligoastrocytomas, which have a mixed histology with features of both astrocytes and oligodendrocytes. Grade IV astrocytoma is the most aggressive type of glioma and is called glioblastoma multiforme.

### **1.2.2.2 HER2+ Breast Cancer**

Another type of cancer discussed in this thesis is HER2+ breast cancer. Breast cancer patients are routinely tested for several markers and classified into subgroups which help guide treatment. Because there are drugs that specifically target human epidermal growth factor receptor 2 (HER2), and therefore only work for HER2+ patients, breast cancer patients are routinely tested for HER2 amplification and/or over-expression using fluorescence in situ hybridization (FISH) and/or immunohistochemistry (IHC). They are also routinely tested for the presence of estrogen receptors (ER status) and progesterone receptors (PR status) using nuclear staining. These tests help stratify patients into subtypes, allowing different subtypes to receive the type of treatment most appropriate for their cancer.

HER2 is also referred to as HER-2/*neu* or ERBB2 and is a member of the ErbB or epidermal growth factor receptor (EGFR) protein family [130]. It is a membrane surface-bound receptor tyrosine kinase involved in signal transduction pathways. Expression of HER2 leads to cell growth and differentiation. In a normal mammary gland, HER2 is

involved in lobuloalveoli development and milk protein secretion [59]. Although HER2 is expressed throughout most developmental stages of the mammary gland, its ligands are only expressed during defined stages [28]. HER2 receptors can either homodimerize or heterodimerize with other HER receptors upon ligand binding, and HER2 is the preferred heterodimeric partner for EGFR, HER3 and HER4 [48]. Amplification of HER2 can lead to overexpression, thereby creating an abundance of receptors on the cell surface, which, because of their close proximity, allows for dimerization in the absence of the ligand [147]. This dimerization leads to autophosphorylation and signaling, and, because HER2 is a proto-oncogene, this excess signaling can cause cancer.

HER2 is overexpressed in about 20% of breast cancer patients [106], and these patients have significantly worse prognosis than other patients [147]. Figure 1.1 is a Kaplan-Meier plot from *The Biology Of Cancer* textbook showing the proportion of breast cancer patients that have not had a recurrence of cancer over time [147]. The patients are grouped into HER2+ and HER2- groups, and the HER2+ patients have significantly worse prognosis than the rest of the patients. This figure was taken from a 1987 paper by Slamon *et. al.* [130], and since then other groups have shown similar results using different methods of grouping patients from various datasets [103, 134]. Figure 1.2 is a figure I made, which also shows that HER2 positive patients (as well as luminal B patients) have worse prognosis than other types of breast cancer, when they are grouped into subtypes using a method developed by Parker *et. al.* [103], called PAM50, which uses the gene expression data from a set of 50 different genes to group the patients into

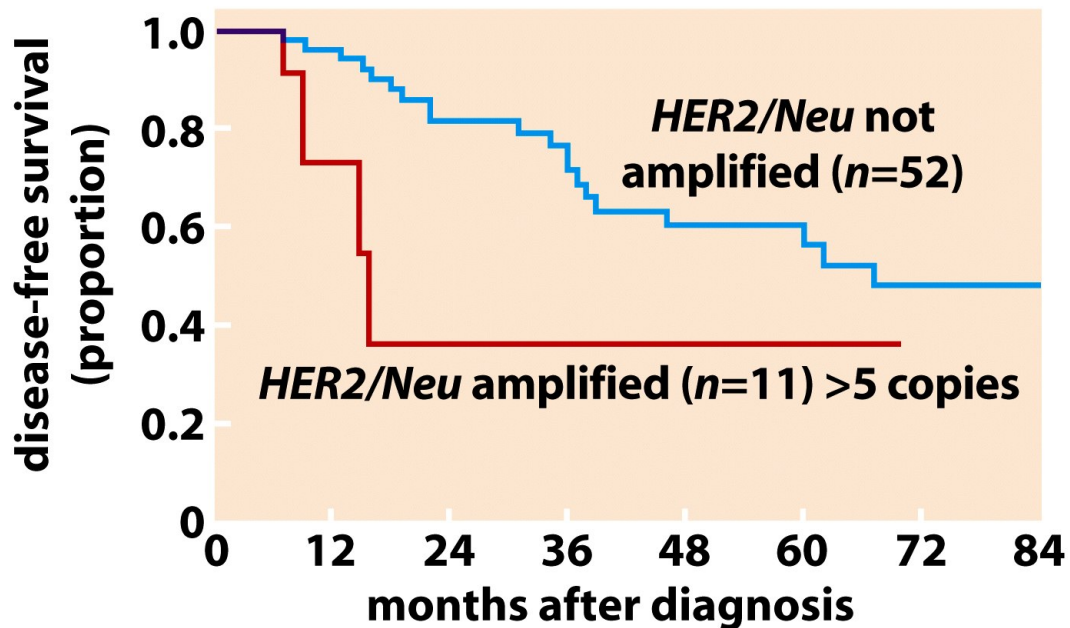


Figure 4-6b The Biology of Cancer (© Garland Science 2007)

Figure 1.1: Kaplan-Meier plot showing disease-free survival for HER2+ and HER2- patients. Figure is from *The Biology Of Cancer* textbook [147].

subtypes. This Kaplan-Meier plot shows the percentage of patients who have not yet had a distant metastasis, meaning the cancer has not yet spread to another location in the body. I generated this plot using a data set from 683 breast cancer patients composed by Christina Yau [157]. These results make it clear that HER2+ breast cancer is an important type of breast cancer to study.

One drug that is commonly used to treat HER2+ patients and has also been used in previous studies is lapatinib. This thesis will discuss how lapatinib was used to generate resistant cell lines. lapatinib, or Tykerb, is a HER2 kinase inhibitor from GlaxoSmithKline used to treat advanced or metastatic HER2+ breast cancer. It is usually given in combination with capecitabine, a chemotherapeutic agent, to patients

Yau Exp Data – PAM50 gene set subtype prediction

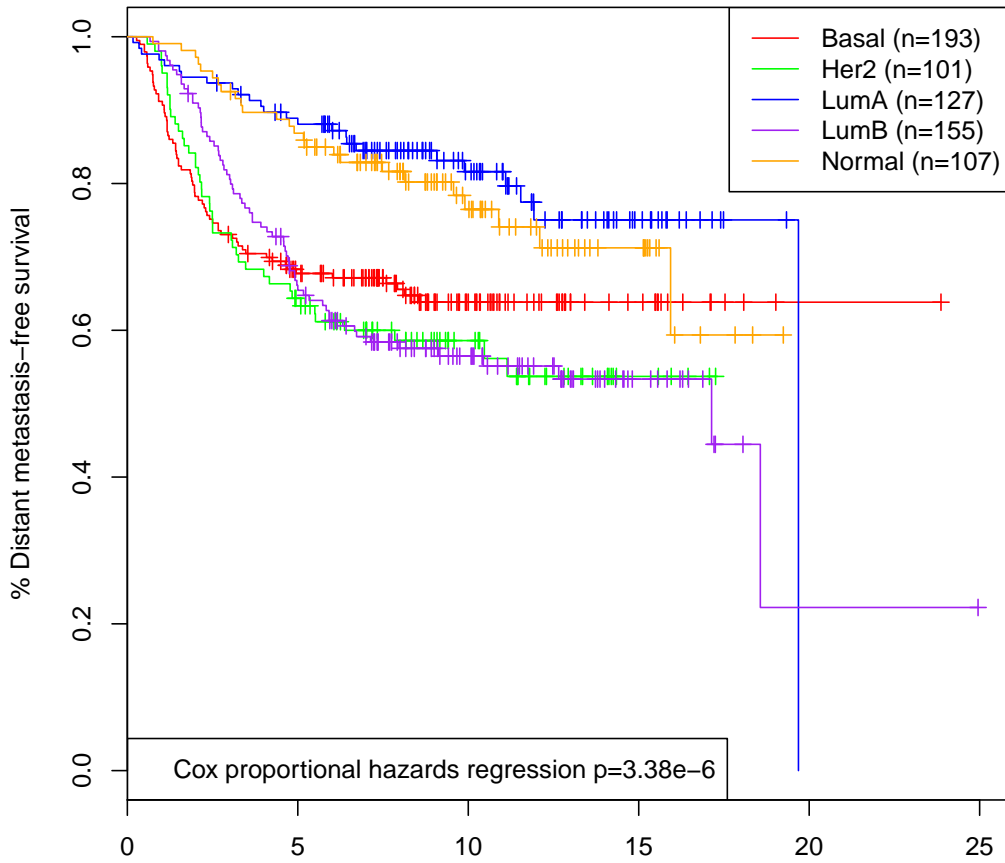


Figure 1.2: **Kaplan-Meier plot showing the percent of patients that have not yet had a distant metastasis over time.** Patients are divided into PAM50 [103] subtypes using gene expression data from 683 breast cancer patients composed by Christina Yau [157].

who have already received prior therapy [43]. It crosses the cell membrane and binds to the tyrosine kinase domain of HER2, preventing phosphorylation of HER2 and therefore preventing signal transduction [154].



## Chapter 2

# Double Minute Chromosomes in Glioblastoma Multiforme Are Revealed by Precise Reconstruction of Oncogenic Amplicons

As discussed in the introduction, high-throughput sequencing technology has enabled many discoveries in cancer research, helping scientists better understand the mechanisms behind cancer growth. Tools developed by Zachary Sanborn that compute copy number and detect structural variants allowed us to gain insights into the mechanisms driving tumor growth in glioblastoma multiforme (GBM), the most aggressive form of glioma. One of these mechanisms, which occurs in over 20% of patients, involves oncogenic

circular amplicons called double minute chromosomes. Because these double minute chromosomes may be detectable in the blood, they may be used for easier diagnosis and monitoring of disease progression. This chapter describes these amplicons in detail for two GBM patients with whole-genome sequencing data, and describes the estimation of the presence of these amplicons in the larger cohort using exome sequencing data.

This chapter contains text from the paper “Double Minute Chromosomes in Glioblastoma Multiforme Are Revealed by Precise Reconstruction of Oncogenic Amplicons” by J. Zachary Sanborn, Sofie R. Salama, Mia Grifford, Cameron W. Brennan, Tom Mikkelsen, Suresh Jhanwar, Sol Katzman, Lynda Chin and David Haussler that was published in *Cancer Research* in 2013. Much of this work was completed by Zachary Sanborn. I contributed to the RNA-Seq analysis, with the help of Sol Katzman in table 2.1 and section 2.2.4. I performed the exome analysis to estimate prevalence of double minute chromosomes and the association of double minute chromosomes/HSR samples with other glioblastoma multiforme tumor features in tables 2.2 and C.1 and sections 2.2.6, 2.3.5 and 2.2.7. I characterized the novel exon, which was discovered by Zachary Sanborn, in section 2.3.3. I also performed the validation of exome sequencing-based prediction of double minute chromosomes/HSR-containing samples in table C.2 and section 2.3.6.

## 2.1 Introduction

High-throughput methods for whole-genome sequencing have provided researchers with an unprecedented ability to measure the complex state of genomic rearrangements characteristic of most cancers. Numerous methods for inferring structural variation from paired-end sequencing data have been developed [127, 24, 9], but the structural variants called by such methods are often considered only in isolation, and used primarily to identify potential fusion genes. The difficulty in discovering all true structural variants and filtering out false positives makes it hard to use the output of such methods to reassemble large regions of the tumor genome. However, such tumor genome assemblies can reveal the complex structure of the tumor genome and can be used to infer the mechanism by which somatic alterations critical to cancer progression occur, such as amplifications of oncogenes and deletions of tumor suppressors. Ideally, these tumor assemblies will also reveal unique features of the cancer that can be used as diagnostic features to monitor disease progression in the patient.

It is well documented that one mechanism by which genes become highly amplified in tumors is via the circularization of double-stranded DNA into what are known as double minute chromosomes [8]. Double minute chromosomes have been shown to confer resistance to certain drugs, as well as pass along this resistance nonuniformly to daughter cells. They have been observed up to a few megabases in size, and contain chromatin similar to actual chromosomes, but lack the centromere or telomeres found in normal

chromosomes. Because double minute chromosomes lack centromeres, they are, like mitochondria, randomly distributed to daughter cells during cell division [81]. They are generally lost in future generations unless there is some selective pressure to maintain them. For example, when they confer selective growth advantage to tumor cells, they are readily retained at high copy number. In particular, double minute chromosomes containing oncogenes may serve to amplify these genes to hundreds of copies per cell. Double minute chromosomes are common in several types of cancer, including brain cancers, with an estimated 10% of glioblastoma multiforme tumors bearing double minute chromosomes [11].

Homogeneously staining regions (HSR) represent another common mode of extreme gene amplification in cancer, observed in both solid and hematologic cancers [6, 159, 137, 136]. An HSR arises from high copy number tandem duplication of a genomic segment such that it expands the affected chromosome. Because they are embedded within larger chromosomes, fluorescent in situ hybridization (FISH) probes specific to genomic sequences within the HSR give a broad band of staining in a specific position within the larger chromosome, distinct from the focal staining usually observed with a locus specific FISH probe. It is believed that double minute chromosomes and HSRs are related, in that double minute chromosomes can derive from HSRs as well as create HSRs via chromosomal reinsertion [6, 16, 112]. FISH is used to distinguish whether HSRs, double minute chromosomes, or both are present in a given tumor sample. Whatever their form, these highly amplified oncogenes are often key drivers of their respective

tumors, and may be vital in their detection, diagnosis, and treatment.

Here, we present methods that analyze high-throughput whole-genome sequencing data from tumor and matched normal samples to detect these extremely amplified and rearranged regions of the tumor genome. We show that these low-level results can be synthesized to construct accurate local genome assemblies that are circular in nature, suggesting the existence of double minute chromosomes and/or HSRs in the tumor. We use FISH analysis to independently identify double minute chromosomes and HSRs in two tumors samples. In addition, we show that RNA-seq data further supports our circular assemblies, in one case identifying a novel isoform of the gene CPM that co-opts intergenic DNA to create a novel exon. Further analysis of the amplicons is done to distinguish likely driver genes, such as MDM2, from likely passenger genes, such as those disrupted by the amplicon structure. We show that this is made possible by the configuration of the rearranged regions represented by the predicted circular assembly. Finally, our analysis of a much larger set of samples with whole-exome sequencing data suggests that 20% of glioblastoma multiforme samples harbor highly amplified oncogenic amplicons.

## 2.2 Methods

### 2.2.1 Tumor and normal genome sequencing data

Tumor and matched-normal whole-genome BAM files were downloaded from CGHub (cghub.ucsc.edu) under the following sample identifiers: The Cancer Genome Atlas (TCGA)-06-0648-01A-01D-0507-08, TCGA-06-0648-10A-01D-0507-08, TCGA-06-0145-01A-01D-0507-08, and TCGA-06-0145-10A-01D-0507-08. In addition, 264 exome BAM files were downloaded from CGHub, with sample identifiers given in Supplementary Table C.1. Whole-genome and exome BAM files were indexed by samtools [76] and processed by BamBam (See Supplementary Materials and Methods) to determine relative copy number, allele fraction, and structural variants. Supplementary Tables C.1 and S2 list the copy number and rearrangement breakpoints detected by BamBam for exome and whole-genome sequencing data, respectively.

### 2.2.2 Determining breakpoints related to highly amplified regions

The read support (number of overlapping reads) for a given breakpoint is directly proportional to the copy number of the regions it connects. Thus, by requiring breakpoints to have a high level of read support, we can filter out breakpoints that are part of a copy-neutral rearrangement, and breakpoints that led to low-copy amplifications and deletions. We can then focus on the breakpoints that are part of highly amplified regions

in the tumor. The particular read support threshold is selected such that breakpoints that have read support expected of copy neutral regions of the tumor genome are removed.

### **2.2.3 Reconstructing amplicons by walking a breakpoint graph**

Similar to a recently published method [49], we construct a breakpoint graph by defining a node for each side of an amplified segment of the tumor genome, connecting these two nodes by a segment edge that represents the amplified segment, and defining a set of bond edges, each of which represents a pair of segment sides that are adjacent in the tumor genome. Which amplified segments of the tumor genome are included as segment edges in the breakpoint graph is determined by relative copy number. The bond edges are the highly supported breakpoints found in the manner described earlier. If an amplified segment is interrupted by a breakpoint, then that segment will be split into two segment edges.

We visualize the graph with the segments laid out according to genomic position. Assuming all segments have the same relative copy number and each segment side has exactly one bond edge attached to it, we determine the arrangement of the amplified segments in the tumor genome by starting at the node for the left side of the first segment and traversing the segment edge to the node at the other side. The path then follows a bond edge attached to that node over to the segment side to which it is con-

nected, then traverses the new segment edge, continuing in this manner of alternatively traversing segment and bond edges until we have returned to the starting node. If there are segment edges remaining that we have not traversed, then we select a new starting node among them and repeat this procedure, until all segment edges are accounted for. Each cyclic path we determine defines a separate circular chromosome via the concatenation of its segments in the direction of traversal. This interpretation of the graph is unambiguous, as the overall direction of traversal of any circular chromosome is immaterial, as is the order in which we discover them. When the number of bond edges per segment side (i.e., node) in the graph is not uniformly one, or not all segments have the same relative copy number, some segments may require more than one traversal to account for the extra copies, and some walks may terminate at segment sides that have no bond. Whenever multiple walks can be made through the set of bond edges and segments, we enumerate the different solutions, including potentially circular solutions that cannot close the loop due to one or more missing bond edges. A single solution that features multiple independent cyclic paths occurs when multiple circular walks are required in the procedure described earlier. Here, distinct sets of bond edges and segments are connected in independent closed loops that lack any bond edges that could fuse the independent loops together into a single, larger closed loop. A toy example breakpoint graph and its solution are described in Supplementary Fig. S1.

When there are multiple solutions, the optimal path(s) through the graph are taken to be those that most closely agree with the observed relative copy number. The number of



times a solution traverses a given segment produces an estimate of that segment's copy number. The root mean square deviation (RMSD) of the segment traversal counts to the observed relative copy number for each solution is calculated, and then the solution(s) with the smallest RMSD value are labeled as optimal.

#### **2.2.4 RNA-Seq analysis**

RNA-Seq reads were mapped using BWA [75] and coverage was calculated using BED tools [109]. We calculated the total coverage per transcript per million uniquely mapped read pairs for each gene within the TCGA-06-0648 double minute chromosomes. For CPM, the coverage over a truncated version of the gene was used as described in the text. These normalized expression metrics of TCGA-06-0648 were compared with a set of nine other TCGA glioblastoma multiforme samples that comprised the original RNA-Seq cohort for glioblastoma multiforme (Table 2.1).

#### **2.2.5 FISH analysis**

FISH was conducted for EGFR/Cep7 and MDM2/Cep12 using prelabeled probes (Vysis, Abbott Molecular). Charged slides with 4-mm paraffin sections were dewaxed, rinsed, microwaved in 10 mmol/L sodium citrate buffer, then digested in pepsin-HCl (40 mg/mL, 10 minutes at 37°C), rinsed, and dehydrated. Probe and slides were codenatured using a HYBrite automated hybridizer at 80°C for eight minutes then hybridized

for two to three days at 37 °C.

### **2.2.6 Exome analysis to estimate prevalence of double minute chromosomes**

Additional samples that potentially bear double minute chromosomes were identified in TCGA glioblastoma multiforme exome datasets as follows. Tumor and matched normal exomes were processed by BamBam to compute relative coverage and identify somatic rearrangements. High copy number peaks were defined as regions with 5-fold increased relative coverage versus their matched normal, a threshold selected to conservatively filter out peaks caused by low-level amplifications and noise (see Supplementary Materials and Methods for details). Glioblastoma multiforme tumors with multiple such high copy number peaks were manually analyzed to discover any oncogenes within peaks, associate peaks with nearby somatic rearrangements, and determine if a sample exhibits multiple peaks with similar copy number levels. Samples were scored as having possible double minute chromosomes if they either contain multiple distinct high copy number peaks with similar copy number levels and at least one peak contained an oncogene, or a single distinct peak containing an oncogene, spanning approximately 1 Mb, and having an associated rearrangement.

### **2.2.7 Association of double minute chromosomes/HSR samples with other glioblastoma multiforme tumor features**

Samples containing likely chromothripsis events were identified from the set of 26 samples with whole-genome sequencing data determined to have possible double minute chromosome(s) by selecting the subset that had multiple ( $> 3$ ) high copy number peaks (18 samples), and were compared with samples containing no high copy number peaks as determined by whole-exome data (112 samples). There after, t tests were conducted in R comparing several features between the two groups, including molecular subtype, survival, and mutation in PTEN, TP53, KEL, IDH1, PIK3R1, PIK3CA, POTEB, NF1, RB1, and EGFR, amplification of PRDM2, MET, MDM2, EGFR, CDK4, CCNE1, and CCND2, and deletion of RB1, PTEN, PRDM2, MET, and CCND2. Bonferroni-adjusted P values were reported. TP53 mutations within the two groups were further evaluated using a two-tailed Fisher's exact test.

## **2.3 Results**

### **2.3.1 BamBam: a robust method for identifying tumor-specific variation**

Considering that a BAM file storing a single patient's whole-genome sequence at high coverage ( $> 30$ ) can be hundreds of gigabytes in compressed form, a serial analysis of

two sequencing datasets requires researchers to store intermediate results that must be merged to conduct a comparative analysis, such as identifying mutations found only in the tumor sample. To overcome this problem, we developed BamBam, a tool that conducts a comparative analysis of a patient’s tumor genome versus his/her germline by simultaneously processing the tumor and matched-normal short-read alignments stored in SAM/BAM-formatted files [76]. Simultaneous processing of both BAM files enables BamBam to efficiently calculate tumor relative coverage and allele fraction, discover somatic mutations and germline SNPs, and infer regions of structural variation. The relative coverage and allele fraction estimates made by BamBam can be used to estimate tumor copy number and normal contamination in the sequenced tumor sample (See Supplementary Methods for details).

### **2.3.2 Reconstruction of candidate double minute chromosomes using whole-genome sequencing**

We focused on three samples from the initial set of 19 glioblastoma multiforme samples from TCGA subjected to whole-genome sequence analysis where we detected highly amplified segments overlapping oncogenes, suggestive of double minute chromosomes. We applied these methods to samples designated TCGA-06-0152, TCGA-06-0648, and TCGA-06-0145. In each case, whole-genome sequencing of a tumor biopsy was available separately from whole-genome sequencing of a blood sample (matched normal tissue sample). For two of these samples (TCGA-06-0152 and TCGA-06-0648), multiple seg-

ments had similar levels of amplification whereas the third sample (TCGA-06-0145) had one large amplified region with further rearrangements internal to the region.

Sample TCGA-06-0648 had a striking pattern of genome amplification in which 16 distinct segments (15 from chromosome 12 and a small fragment from chromosome 9) had similarly high levels of amplification ( $> 10$  copies) and also appeared to be linked to each other by high confidence rearrangement events identified by BamBam (Fig. 2.1A). One of the chromosome 12 segments contains the MDM2 oncogene. Out of the total of 701 putative somatic breakpoints called, 97 breakpoints met the filtering criteria specified in Supplemental Methods. Only 16 of these 97 breakpoints further met or surpassed a chosen minimum read support threshold of 100 to identify breakpoints likely associated with highly amplified regions, including two breakpoints that did not have split read evidence. All of these highly supported breakpoints are proximate to the boundaries of the highly amplified segments clustered on chromosome 12, suggesting that the highly supported breakpoints and the amplifications are directly associated and may represent the rearranged configuration of one (or multiple) amplicons in the tumor's genome.

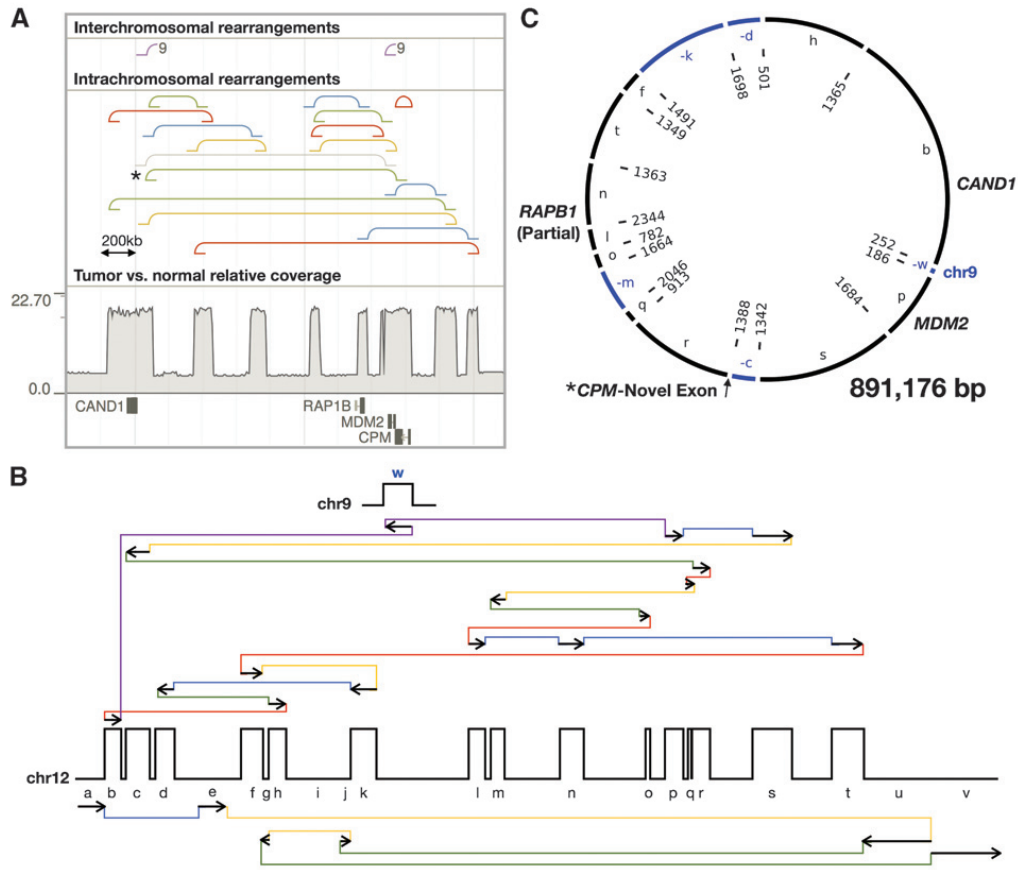
Figure 2.1B shows a schematic of the amplified segments of TCGA-06-0648 and their associated breakpoints. This diagram predicts a circular path that completely accounts for all observed breakpoints and amplified segments, resulting in a single 891-kb circular amplicon containing a single copy of MDM2. Because all but two breakpoints were refined by split read solutions, the estimated size of this amplicon is accurate to within

approximately 100 bp. This reconstructed circle is consistent with either an array within a larger chromosome of precise tandemly duplicated copies of an initial circular amplicon formed from these 16 genomic segments (assuming single-copy breakpoints joining this tandem array to non-repeated DNA were not sufficiently covered to be observed) or an extra chromosomal circular DNA (double minute chromosomes; [67]) with average copy number of approximately 84 in the tumor sample. As we can assume neither a perfectly clonal tumor nor a stable number of double minute chromosome copies in every generation of tumor cells, the average copy number of 84 should be considered the number of double minute chromosome copies in the average tumor cell sequenced. We specifically searched for breakpoints with lower read support within the amplicon region connecting it to other single-copy genomic locations, as these could provide evidence for an insertion site of a tandem array, but found none. Thus, the presence of a double minute chromosome is the more parsimonious explanation of these data, as it does not require us to postulate the existence of one or more pairs of unobserved breakpoints where one or more HSRs each containing multiple exact copies of this 891-kb region are inserted into larger chromosomes.

Double minute chromosomes have been identified in many tumor types where they often contain oncogenes important for that cancer, such as EGFR in the case of glioblastoma multiforme [146]. The TCGA-06-0648 double minute chromosome contains several protein-coding genes from chromosome 12, including intact copies of the MDM2 oncogene and CAND1, which encodes an inhibitor of cullin ring-ubiquitin ligase com-

plexes [36]. In addition, it includes a truncated allele of carboxypeptidase M (CPM), a membrane-bound and secreted protease that cleaves the C-terminal residue of epidermal growth factor (EGF; [35]) . The amplified allele of CPM lacks the last exon, which should not affect the catalytic or major structural domains of the protein, but removes the amino acids necessary for GPI anchoring of CPM to the plasma membrane and would be expected to result in an exclusively secreted form of CPM [114]. A partial allele of the ras-family gene RAP1B is also present, but as it lacks the promoter and first exon, it is unlikely that this allele is expressed. It seems likely that MDM2 drove the high copy maintenance of this double minute chromosome in TCGA-06-0648, but CPM and CAND1 could contribute as well.

This region of the tumor genome has all of the hallmarks of a chromothripsis event, suggesting that the double minute chromosome was created by connecting shattered fragments of chromosome 12 and a small region of chromosome 9 into a single circular episome [135]. The observation that multiple regions with uniformly high read depths are connected by a set of structural variants with similarly high-read support suggests that these alterations likely occurred together during a single event, such as chromothripsis, instead of a series of independent focal rearrangements. Furthermore, all breakpoints lack homology or exhibit 2 to 6 bp microhomologies at their junctions, indicating that nonhomologous end-joining (NHEJ) and microhomology-mediated end joining (MMEJ) are the primary DNA double-stranded break-repair mechanisms responsible for constructing the double minute chromosome [88].



**Figure 2.1: Reconstruction of TCGA-06-0648 double minutes.** A, tumor browser view of a region of chromosome 12 from TCGA-06-0648. The bottom track shows protein-coding genes overlapping the amplified segments. The track above shows relative copy number of the tumor DNA compared with the normal, showing distinct blocks of elevated copy number with similar total copy number between blocks. The next two tracks show intra- and interchromosomal rearrangement breakpoints. All rearrangements shown are supported by at least 100 discordant reads. The type of rearrangement is indicated by the color of the line: duplication (red), deletion (blue), inversion (yellow and green), and interchromosomal rearrangement (purple). B, a diagram of the amplified segments and structural variants identified on chromosome 12 and 9 is shown. Walking through this diagram results in a circular solution, suggesting the double minute chromosomes diagrammed in C where segments inverted relative to their orientation in the reference genome are colored blue. The letters inside the circle correspond to the segments in B. The numbers inside the circle indicate the number of sequencing reads supporting each breakpoint.



We applied these same methods to sample TCGA-06-0152, which also had several highly amplified segments, including segments containing the MDM2, CDK4, and EGFR oncogenes (TCGA; [15]). This analysis predicted two amplicons, in which each amplicon harbored at least one oncogene (Supplementary Fig. S2).

The final case, TCGA-06-0145, exhibits an extreme level of amplification ( $> 50$ -fold) of a single approximately 800 kb genomic segment including EGFR that could indicate the presence of an EGFR-double minute chromosome or HSR (Fig. 2.2). In contrast with the other samples, the amplified region of TCGA-06-0145 contains significant variations in the major and minor allele frequency as well as deletion and duplication events with lower read support, which are more compatible with an HSR interpretation. However, again, we were unable to find evidence of breakpoints that would link this amplicon to another genomic region, which argues against an HSR.

The solution to the breakpoint graph of TCGA-06-0145 shows the possibility of three distinct paths that incorporate all breakpoints and explain the observed copy number, and each path predicts a different form of EGFR. EGFR is intact in the dominant path (seven of every nine copies). The remaining two paths, each present in one of nine copies, feature breakpoints that are internal to the EGFR gene, with one path producing a nonfunctional form of EGFR and the other deleting exons 2 to 7 of EGFR. This form of EGFR is known as EGFRvIII, a highly oncogenic, constitutively active form of EGFR that is expressed in multiple tumor types [89]. This is interesting because it suggests two scenarios: (i) EGFRvIII emerges after wild-type EGFR is significantly amplified

or (ii) EGFRvIII is created early but cells with more copies of wild-type EGFR have a selective advantage in the tumor population. The former scenario seems most plausible, as the increased EGFRvIII mutant will emerge. Regardless of the true scenario, the ratio of EGFRvIII to wild-type EGFR suggests that a high copy number of oncogenic EGFRvIII may not be necessary to provide significant advantage over the wild-type amplification to the growing tumor cell.

### **2.3.3 Transcriptome data reveals a novel double minute chromosome-associated fusion protein**

For one of the three tumor samples examined in this study (TCGA-06-0648), RNA sequencing was also conducted by TCGA. We analyzed these data together with that from the nine other samples in the initial glioblastoma multiforme RNA-Seq batch to examine the expression of alleles associated with the TCGA-06-0648 double minute chromosome. As expected from the absence of the promoter and first exon, RAP1B expression was half that of the other glioblastoma multiforme samples that do not have amplifications in this region, suggesting that only the intact copy of RAP1B on chromosome 12 is expressed and the double minute chromosome allele is not expressed. In contrast, MDM2, CAND1, and the first eight exons of CPM were expressed at more than 15-fold higher levels than was observed in the glioblastoma multiforme samples lacking amplification of these genes (Table 2.1).

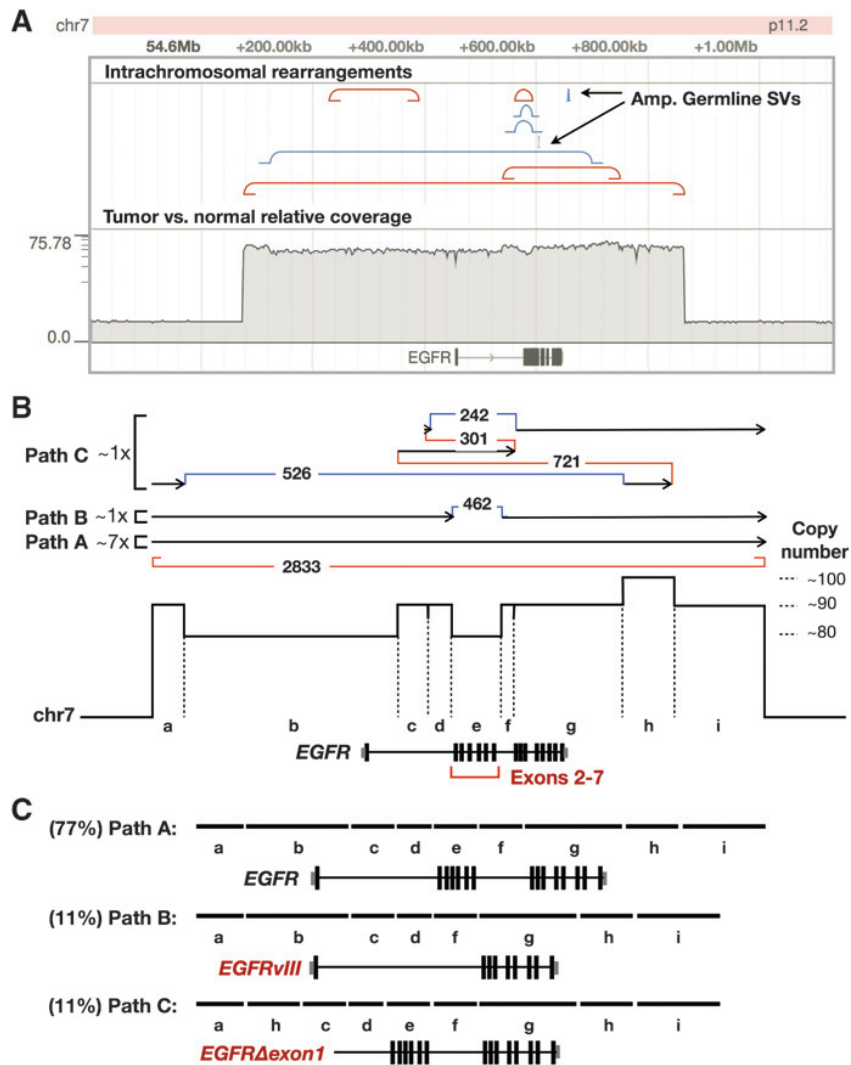


Figure 2.2: **Reconstruction of 06-0145 amplicons.** A, browser shot of the amplicon on chromosome 7. Tracks as described for Fig. 2.1. The two “Amp. Germline SVs” are known germline breakpoints present on the amplicon. B, diagram of the three paths that together form a potential solution of the breakpoint graph for sample TCGA-06-0145 that accounts for all observed, highly supported breakpoints. The location of the gene EGFR is noted. Note that all paths, A, B, and C, can create circular solutions by using the encompassing tandem duplication shown in red to connect the end of each path to the beginning. The tandem duplication may also connect path A to path B, path C to path B, and so on. C, the effects of each path on the EGFR gene, showing that path B creates an oncogenic form of EGFR, EGFRvIII, through specific deletion of exons 2 to 7.

Sample	Expression <sup>a</sup>			
	<i>MDM2</i>	<i>CPM</i> (exons 1–8)	<i>CAND1</i>	<i>RAP1B</i>
TCGA-GBM-02-0033	2,849	1,180	18,717	22,098
TCGA-GBM-06-0124	3,504	662	15,256	14,654
TCGA-GBM-06-0126	3,247	906	14,808	10,403
TCGA-GBM-06-0155	1,955	911	17,741	32,690
TCGA-GBM-06-0214	5,382	2,783	15,000	15,429
TCGA-GBM-06-0216	3,050	143	20,966	14,595
TCGA-GBM-06-0648	57,126	9,888	480,414	8,493
TCGA-GBM-06-0879	1,235	498	12,833	14,488
TCGA-GBM-12-0692	909	224	14,716	11,022
TCGA-GBM-14-0786	2,305	514	17,484	13,911
Avg <sup>b</sup>	2,382	630	16,565	16,733
0648 fold increase	23.98	15.70	29.00	0.51

<sup>a</sup>Expression reported as normalized transcript coverage (coverage per transcript per million uniquely mapped paired-end reads).

<sup>b</sup>Excludes 06-0214 and 06-0648, which have *MDM2* amplification.

Table 2.1: **RNA-Seq-based expression estimates of 06-0648 double minute chromosome-associated genes**

For CPM, we observed that many reads originating in exon 8 terminated in a region 1.47 Mb away from it in the normal version of chromosome 12, but only 13.5 kb away in the double minute chromosome. Closer analysis of this region revealed that the 5' end of these reads are just downstream of a canonical splice-site acceptor sequence that generates a new exon encoding a novel 30 amino acid carboxy terminus for the double minute chromosome-derived CPM allele (Fig. 2.3). This region is not part of any known transcript and the resulting protein sequence has no strong homology to any other proteins. This sequence is unlikely to provide a GPI anchor site; therefore, we anticipate that the double minute chromosome-derived CPM protein would be secreted. It is not clear what the functional effect of expressing this altered CPM gene would be, although it may affect EGF metabolism as both membrane bound and secreted forms of CPM are known to cleave the carboxy-terminal arginine of EGF [86].

In summary, from the point of view of gene expression, the double minute chromosome

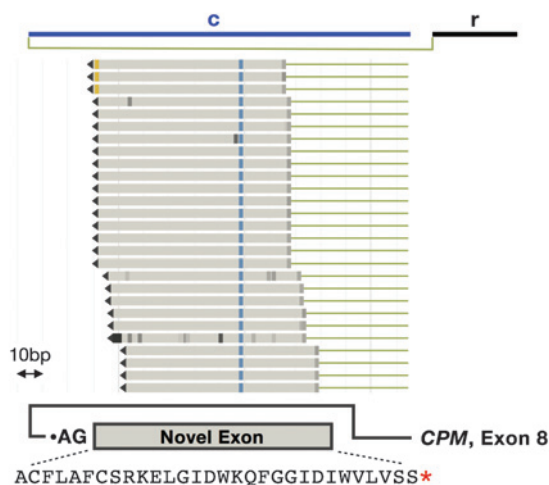


Figure 2.3: **Novel CPM C-terminal exon expressed from the TCGA-06-0648 double minute chromosome.** The top line shows the appropriate region from Fig. 2.1C, with sequencing reads where the other pair maps to exon 8 of CPM below. The bottom panel shows the orientation of the new exon relative to CPM and its amino acid sequence.

results in overexpression of MDM2, CAND1, and a novel form of CPM in this tumor sample.

### 2.3.4 TCGA-06-0648 and TCGA-06-0145 amplicons exist as double minute chromosomes

The ability to distinguish HSRs from double minute chromosomes from short-read sequencing data is limited. To independently assess the nature of the amplification events in these tumors, we conducted FISH analysis on paraffin sections derived from tumors TCGA-06-0648 and TCGA-06-0145 using probes to MDM2 (amplified in TCGA-06-0648) and EGFR (amplified in TCGA-06-0145; Fig. 2.4). Material was unavailable

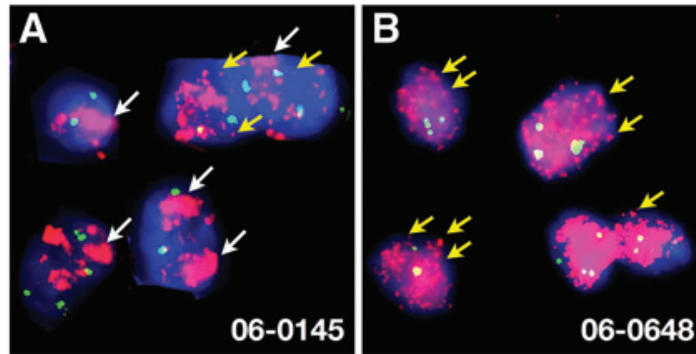


Figure 2.4: **Visualization of glioblastoma multiforme tumor oncogenic amplicons.** FISH analysis of formalin-fixed paraffin-embedded blocks matching samples 06-0145 (A) and 06-0648 (B). Probes for the centromeric region of chromosome 7 (A) and chromosome 12 (B) shown in green gave two to four foci per cell. Probes for EGFR (A) and MDM2 (B) are shown in pink. White arrows point to broad, HSR-like staining patterns, and yellow arrows point to discrete extrachromosomal spots.

for TCGA-06-0152. In 0648, the MDM2 probe gave a punctate pattern throughout the nucleus, indicating many nonchromosomal sites of MDM2, typical of double minute chromosomes. In contrast, the EGFR probe gave a broad pattern of staining in addition to punctate spots for TCGA-06-0145, consistent with a combination HSR and double minute chromosomes.

### 2.3.5 Prevalence of putative double minute chromosomes/HSRs in exome sequencing data

Exome sequencing is cheaper than whole-genome sequencing, and samples of tumors that have been subjected to exome sequencing are more plentiful. Evidence for double minute chromosomes can be obtained from exome-sequencing by searching for patterns

of multiple regions of high-level amplification overlapping at least one oncogene, a pattern common to the double minute chromosome-containing samples we analyzed. In contrast, broad or chromosome arm-level amplification events as well as focal events with a modest level of amplification (e.g., duplications) are not expected to be part of a double minute chromosome. To estimate the prevalence of double minute chromosomes/HSRs in glioblastoma multiforme, we searched a set of 264 TCGA samples with tumor and matched-normal exome sequencing data for signatures of double minute chromosomes. As detailed in the Materials and Methods, we first conducted a computational survey of samples to identify samples exhibiting focal extreme amplification(s), prioritizing those samples with multiple distinct peaks on one or more chromosomes. A careful manual review of these samples was conducted to assess the likelihood that the sample contains a double minute chromosome, looking for the quality of relative copy number calls, any evidence of structural variation associated with the amplified peaks, and the presence of potential oncogenes.

As described in Table 2.2 and Supplementary Table C.1, 61 samples (23%) have features suggesting the presence of a double minute chromosome. A total of 121 oncogenes were amplified across these 61 samples, with at least one oncogene identified in every putative double minute chromosome. EGFR was the oncogene most frequently associated with these high-level amplicons, followed by CDK4 and MDM2. MYCN, which has been identified in a number of glioblastoma multiforme double minute chromosomes [11], was associated with amplicons in two samples. As a group, the putative double minute

Total samples	264
Total potential double minutes	61
Potential DM with <i>EGFR</i>	46
Potential DM with <i>CDK4</i>	18
Potential DM with <i>MDM2</i>	14

Table 2.2: **Summary of amplicons found in TCGA GBM exomes**

chromosomes/HSR-containing samples showed similar survival to the cohort of samples analyzed by exome sequencing (data not shown). Taken together, these results suggest that almost a quarter of glioblastoma multiforme samples have oncogenic amplicons present at high copy number.

### 2.3.6 Validation of exome sequencing-based prediction of double minute chromosomes/HSR-containing samples

Recently, the TCGA project conducted whole-genome sequencing on an additional 25 tumor/normal pairs from the glioblastoma multiforme cohort. This new dataset contains 23 samples that we predicted to harbor a double minute chromosome/HSR on the basis of our curation of the exome data for these samples. Although complete analysis of this data is being pursued by the Analysis Working Group, we conducted a preliminary analysis using BamBam and the methods described earlier to identify circular amplicons in these samples (summarized in Supplementary Table C.2). For 16 of the 23 samples, we were able to reconstruct at least one circular amplicon. For the remaining 7 of the 23 samples, multiple highly amplified peaks were identified with breakpoints connecting



many, but not all, of the peaks. We could not reconstruct circular amplicons for these samples, although the possibility remains that rearrangement breakpoints allowing for a circular solution were not detected for technical reasons. For 2 of the 25 samples where we did not predict a double minute chromosome/HSR based on the exome sequencing data, we also did not detect highly amplified genomic regions with associated rearrangements in the whole-genome sequencing data.

This larger set of samples with circular amplicons allowed us to look for common features associated with these samples compared with TCGA glioblastoma multiforme samples where exome data suggests no amplicons. Previous studies identifying medulloblastomas with chromothripsis-associated double minute chromosomes and other complex genetic rearrangements noted that such samples harbor p53 mutations [111, 94]. However, there was no enrichment of p53 mutations in the samples with circular amplicons in this sample cohort. The only significant association we observed was with PTEN deletion (Bonferroni-adjusted P value = 0.0052), a common event in glioblastoma multiforme tumors.

Focusing on the 16 samples where we successfully reconstructed circular amplicons with at least one oncogene, a mixture of simple and complex amplicons was observed. One sample harbored two simple circular amplicons, each containing one genomic segment with an oncogene and the remaining five had a single oncogenic amplicon containing one genomic segment (Supplementary Table C.2). Ten samples are complex, harboring multiple highly amplified segments from one or more chromosomes. Of these 10, six

samples have at least two circular amplicons. Together, these results strengthen our estimates of the prevalence double minute chromosomes/HSR amplicons in glioblastoma multiformes and suggest that samples containing such amplicons can often harbor multiple independent amplicons.

## 2.4 Discussion

The ability to integrate relative copy number with breakpoints enables us to understand the topology of vital parts of the cancer genome. By examining the overall pattern of amplification events in the TCGA glioblastoma multiforme whole-genome sequencing data, we found that some samples had multiple highly amplified segments of similar copy number. Surprisingly, for many of these highly amplified tumor regions, we are able to completely explain both the observed copy number and highly supported breakpoints surrounding them by solving a simple breakpoint graph, which describes the order and orientation of the highly amplified segments in the tumor genome. For the three glioblastoma multiforme samples analyzed in detail here, the optimal solutions to the breakpoint graphs of amplified segments are circular amplicons. These circular solutions suggest that the observed amplified regions are double minute chromosomes or HSRs.

Sequence coverage of 30X could be limiting our ability to detect breakpoints associated with the chromosomal integration site of an array of these amplicons, as the copy number

of the breakpoint at the integration site is much lower than that of the amplicon. Thus, at this coverage we cannot reliably distinguish between a double minute chromosome and an HSR with our bioinformatic analysis. The availability of tissue sections derived from these same tumor samples did allow us to directly address this issue for two samples. FISH analysis of one sample, TCGA-06-0648, is consistent with a double minute chromosome, whereas another, TCGA-06-0145, gives a pattern suggestive of a combination of double minute chromosome and HSR.

Much can be learned through precise knowledge of the amplicon structure. Genes whose coding or promoter regions are disrupted by the amplicon structure are obvious passenger gene candidates, provided that transcriptional machinery is unable to create a novel transcript like the one observed with the birth of a new exon for the CPM gene. Highly expressed genes such as MDM2 that remain intact may drive tumor development and/or play a role in tumorigenesis, using the double minute chromosome's ability to rapidly reproduce to significantly increase the oncogenic capacity of these cells. The highly amplified state of these oncogene-harboring amplicons indicates that they have strong oncogenic potential and, thus, confer a selective advantage to the tumor cell. Their formation was likely a key event in the tumorigenesis of these glioblastoma multiforme tumors and they are likely to persist overtime.

Examining the larger set of over 250 TCGA glioblastoma multiforme samples for which there is exome data suggests that oncogenic amplicons are quite prevalent as they are found in almost a quarter of the samples. Using recently generated whole-genome

sequencing data for an additional 23 samples from the set of samples where we had predicted double minute chromosomes/HSRs from exome sequencing data, we were able to confirm all of these samples had highly amplified genomic segments with rearrangements connecting at least some of the amplified segments. Further, we were able to reconstruct circular amplicons for 16 of these samples and discovered that six had multiple amplicons as we had observed for TCGA-06-0152, suggesting that the presence of multiple amplicons is common, and that they persist by conferring a combined selective advantage to the tumor cells. These results bolster the notion that chromothripsis type events occur with reasonable frequency in glioblastoma multiforme and, through amplification of oncogene expression, contribute to tumorigenesis.

The prevalence of these amplicons suggests that tumor specific breakpoints associated with double minute chromosomes amplicons may be a potential diagnostic for the presence of tumor cells in a significant fraction of patients with glioblastoma multiforme, especially if double minute chromosome-derived DNA is present in the blood. Several recent observations suggest the possibility of finding such tumor DNA in the blood of patients who have glioma. Skog and colleagues reported microvesicles that bud off from glioblastoma multiforme tumor cells with lipid bilayers intact, carrying cytoplasmic contents of the tumor cell such as mRNA, miRNA, and angiogenic proteins [129]. These vesicles can deliver their contents to other cells or blood. More recently, Balaj and colleagues have isolated microvesicles containing single stranded DNA (ssDNA) with amplified oncogenic sequences, in particular *c-Myc* [7]. Tumor nucleic acids leaked

into the blood in this manner or via macrophages that engulf necrotic or apoptotic cells have been proposed as possible disease biomarkers in several cancers including glioma [124, 69]. Indeed, a recent study robustly detected tumor-associated rearrangements from patients who have breast and colorectal cancer via high-throughput sequencing of the cell-free, plasma fraction of blood [71]. The methods described in this study could readily be applied to such sequencing data and, thus, potentially provide a noninvasive method to characterize and monitor patients with glioma.

## Chapter 3

# Comprehensive and Integrative Genomic Characterization of Diffuse Lower Grade Gliomas

Another form of gliomas, termed lower-grade gliomas (LGG), have historically been thought of as less aggressive than GBM and often eventually progress to GBM. This chapter describes work by The Cancer Genome Atlas (TCGA) consortium, for which I have been heavily involved, describing a new method of classification of LGG samples. This classification reveals one subtype of LGG which is just as aggressive as GBM and has features very similar to GBM, suggesting these patients may benefit from more aggressive treatment. This work demonstrates how this new method of clinical classification is more accurate and groups patients into more clinically relevant subtypes than

the current standard classification, which is histology and grade.

This chapter contains text from the paper "Comprehensive and Integrative Genomic Characterization of Diffuse Lower Grade Gliomas" by The Cancer Genome Atlas Network submitted to *The New England Journal Of Medicine* in 2014. I performed the cluster of clusters analysis described in figures 3.1 and A.3 and sections 3.2.2, 3.3.2 and 3.3.8, the comparison of copy number alterations between IDHwt LGG and GBM in figure 3.3A and sections 3.3.3, 3.3.5 and 3.3.8 and processed data used to compare mutations, copy number changes and structural variants in figure 3.3B and sections 3.3.3, 3.3.5 and 3.3.8 (the generation of the figure was completed by Olena Morozova). I also analyzed deletions in CDKN2A/B described in figure A.1 and section 3.3.8. I processed and analyzed genomic rearrangement data from the tool, BamBam (developed and run by Zachary Sanborn), described in figure 3.3B, table C.3 and section 3.3.6 and performed the double minute/breakpoint-enriched region analysis described in figures 3.3A and A.2, table C.3 and section 3.3.6. Finally, I performed the initial survival analysis described in figure 3.4B and section 3.3.9, although the final version of the figure was re-made by Laila Poisson.

## **3.1 Introduction**

Diffuse low grade and intermediate grade gliomas (WHO grades II and III, hereafter called lower grade gliomas (LGG)) are infiltrative neoplasms that arise most often in

the cerebral hemispheres of adults and include astrocytomas, oligodendrogliomas and oligoastrocytomas of World Health Organization (WHO) grades II and III [78, 98]. Due to their highly invasive nature, complete neurosurgical resection is nearly impossible, and residual neoplasms eventually undergo malignant progression, yet with highly variable intervals. A subset will progress to glioblastoma (GBM, WHO grade IV) within months, while others remain stable for years. Similarly, survival ranges widely from 1-15 years, with some LGG showing impressive chemosensitivity [18, 82]. Current treatment varies with extent of resection, histology, grade and results from ancillary testing, and includes clinical monitoring, chemotherapy, and radiation therapy, with salvage options available upon treatment failure [14, 120]. Although comprehensive molecular profiling has greatly clarified the pathogenic landscape of GBM, LGG remain less understood [15, 21].

While histopathologic classification of LGG is time-honored, it suffers from high intra- and interobserver variability [78, 31, 143]. Consequently, clinicians increasingly rely on genetic classifications to guide clinical decision-making [120, 139, 140, 148]. Mutations in IDH1 and IDH2 characterize the majority of LGG and define a subtype with favorable prognosis [57, 105, 155]. Those with chromosome 1p/19q co-deletion, most often noted in oligodendrogliomas, have better responses to chemotherapy and longer survival [17, 19, 46]. TP53 and ATRX mutations are more frequent in astrocytomas and are also important markers of clinical behavior [61]. To gain additional insights, we performed a comprehensive analysis of 293 LGG, using multiple advanced molecular platforms, to



provide a deeper understanding of their features, to classify them in a clinically relevant manner, and to shed light on potential targets for therapy.

## **3.2 Methods**

### **3.2.1 Patients**

Tumor samples were from 293 adults with previously untreated LGG (WHO grade II and III) and included 100 astrocytomas, 77 oligoastrocytomas and 116 oligodendrogliomas [149]. Diagnoses were established at contributing institutions and a quality review was performed by TCGA neuropathologists. Patient characteristics are described in Tables 3.1, S1 and S2, and in Supplementary Information. We obtained appropriate consent from relevant Institutional Review Boards. The range of patient ages, tumor locations, clinical histories and outcomes, histologies and grades were typical of patients diagnosed with diffuse glioma [149].

### **3.2.2 Analytic Platforms**

We performed exome sequencing (n= 289), DNA copy number profiling (n=285), mRNA sequencing (n=277), microRNA sequencing (n=293), DNA methylation profiling (n=289), and reverse-phase protein lysate array (RPPA) protein expression profiling (n=255). Complete data for all platforms were available for a core set of 254 LGG. Whole genome

sequencing and low pass whole genome sequencing were performed on 21 and 52 samples, respectively. The molecular data package associated with this report was frozen on 1/31/2014. Clinical data were frozen on 8/25/2014. The complete list of data sets is provided in Table S1. Data are available through the Cancer Genome Atlas (TCGA) data portal (<https://tcga-data.nci.nih.gov/tcga>) and CGHub (<https://cghub.ucsc.edu/>). In addition to single platform molecular characterization, we also performed unsupervised analysis that integrated results from multiple platforms, including Cluster of Cluster (CoC) and OncoSign [27]. Briefly, CoC is a second level clustering of class assignments derived from each individual molecular platform. OncoSign classifies tumor samples based on similarity of recurrent somatic events. Statistical analysis includes Fishers exact test for association between categorical variables; one-way ANOVA for association with continuous outcomes; Kaplan-Meier estimates of survival with log-rank tests between strata; and Cox proportional-hazards regression for multi-predictor models of survival. In depth descriptions of methods are found in the Supplemental text.

## **3.3 Results**

### **3.3.1 Histology and molecular subtype**

In order to compare tumor profiles derived from molecular platforms with both histologic classification and with markers frequently used in clinical practice (IDH mutation and 1p/19q co-deletion status), we classified all LGG in our cohort as either:

1) IDH-mutated, 1p/19q co-deleted (IDHmut-codel); 2) IDH-mutated, 1p/19q non-co-deleted (IDHmut-non-codel); or 3) IDHwildtype (IDHwt). Consistent with prior reports [19, 2, 44, 85], we showed a strong correlation of IDHmut-codel status with oligodendroglioma histology (69/84) (Table 3.1). IDHmut-non-codel samples (n=139; 50% of cohort) represented a mixture of LGG histologies, enriched for astrocytomas and oligoastrocytomas. IDHwt samples were mostly astrocytomas (31/55) and grade III LGG (42/55), yet also included other histologies and grades (Table 3.1). Overall, IDH-1p/19q status correlated well with oligodendroglioma classification, but only modestly with astrocytoma and oligoastrocytoma histologies.

### **3.3.2 Multiplatform integrative analysis**

Grouping of tumors based on molecular patterns has shown robust associations with disease biology in GBM [95, 107, 145]. To assess whether similar associations could be established in LGG, we performed unsupervised clustering of molecular profiles derived from four independent platforms and demonstrated well-defined subtypes based on DNA methylation (5 clusters; Figure S1-5); gene expression (4 clusters; Figure S6-7); DNA copy number (3 clusters; Figure S8); and microRNA (4 clusters; Figure S9-10).

To integrate data from individual molecular platforms and compare profiles to histology and IDH-1p/19q status, cluster group assignments defined by the four individual platforms (DNA methylation, mRNA, DNA copy number, and microRNA) were used

		Total (n=278£)	IDHmut codel (N=84)	IDHmut non-codel (N=139)	IDHwt (N=55)
Histological Type ## and Grade ##	Oligodendroglioma II	65 (23%)	<b>38 (45%)</b>	21 (15%)	6 (11%)
	Oligodendroglioma III	44 (16%)	<b>31 (37%)</b>	6 (4%)	<b>7 (13%)</b>
	Oligoastrocytoma II	41 (15%)	9 (11%)	30 (22%)	2 (4%)
	Oligoastrocytoma III	33 (12%)	4 (5%)	<b>20 (14%)</b>	<b>9 (16%)</b>
	Astrocytoma II	30 (11%)	1 (1%)	24 (17%)	5 (9%)
	Astrocytoma III	65 (23%)	1 (1%)	38 (27%)	<b>26 (47%)</b>
Age (Yrs) at Diagnosis ##	Mean (SD)	42.6 (13.5)	<b>45.4 (13.2)</b>	<b>38.1 (10.9)</b>	<b>49.9 (15.3)</b>
	Min, Max	14, 75	17, 75	14, 70	21, 74
Gender	Male	155 (56%)	45 (54%)	84 (60%)	26 (47%)
Race (self-report, n=274)	White	261 (95%)	79 (98%)	131 (95%)	51 (93%)
Ethnicity (self-report, n=261)	Hispanic/Latino	14 (5%)	5 (6%)	6 (5%)	3 (6%)
Treating Country	United States*	265 (95%)	81 (96%)	131 (94%)	53 (96%)
Year of Diagnosis	Before 2005	38 (14%)	10 (12%)	18 (13%)	10 (18%)
	2005-2009	88 (32%)	30 (36%)	44 (32%)	14 (25%)
	2010-2013	152 (55%)	44 (52%)	77 (55%)	31 (56%)
Family History of Cancer (n=190**)#	No History	108 (56%)	30 (52%)	64 (65%)	13 (38%)
	Primary Brain	11 (6%)	2 (3%)	7 (7%)	<b>2 (6%)</b>
	Other Cancers	72 (38%)	26 (45%)	27 (28%)	<b>19 (56%)</b>
Extent of resection (n=268)	Open Biopsy	6 (2%)	1 (1%)	4 (3%)	1 (2%)
	Subtotal Resection	98 (37%)	31 (38%)	45 (34%)	22 (40%)
	Gross Total Resection	164 (61%)	49 (60%)	83 (63%)	32 (58%)
Tumor Location##	Frontal Lobe	172 (62%)	<b>68 (81%)</b>	<b>84 (60%)</b>	20 (36%)
	Parietal Lobe	23 (8%)	5 (6%)	13 (9%)	5 (9%)
	Temporal Lobe	74 (27%)	9 (11%)	40 (29%)	<b>25 (45%)</b>
	Other+	9 (3%)	2 (2%)	2 (1%)	5 (9%)
	Midline	5 (2%)	2 (2%)	2 (1%)	1 (2%)
Laterality (n=276)	Left	133 (48%)	37 (44%)	69 (50%)	27 (49%)
	Right	138 (50%)	45 (54%)	66 (48%)	27 (49%)
White vs. Grey Matter (n=144)	White Matter	74 (51%)	26 (54%)	37 (51%)	11 (46%)
First Presenting Symptom (n=252)	Headache	64 (25%)	15 (21%)	39 (30%)	10 (20%)
	Mental Status Change	22 (9%)	7 (10%)	10 (8%)	5 (10%)
	Motor/Movement Change	18 (7%)	6 (8%)	7 (5%)	5 (10%)
	Seizure	135 (54%)	38 (53%)	70 (54%)	27 (53%)
	Sensory Change	6 (2%)	3 (4%)	1 (1%)	2 (4%)
	Visual Change	7 (3%)	3 (4%)	2 (2%)	2 (4%)
Use of Preoperative Antiseizure Med. (n=202)	Yes	148 (73%)	48 (75%)	73 (72%)	27 (75%)
Use of Preoperative Corticosteroids (n=207)	Yes	90 (43%)	29 (43%)	46 (44%)	15 (43%)

£ Eleven cases with clinical information do not have IDH/Codel status determined

\* Twelve cases were submitted from Russia (3 IDH<sub>mut-codel</sub>, 7 IDH<sub>mut-non-codel</sub>, 2 IDH<sub>wt</sub>), and three case from Italy (1 IDH<sub>mut-non-codel</sub>, 2 unknown IDH/codel)

\*\* Cases with response to both questions of family history of any cancer (n=192) and of family history of primary brain cancer (n=197)

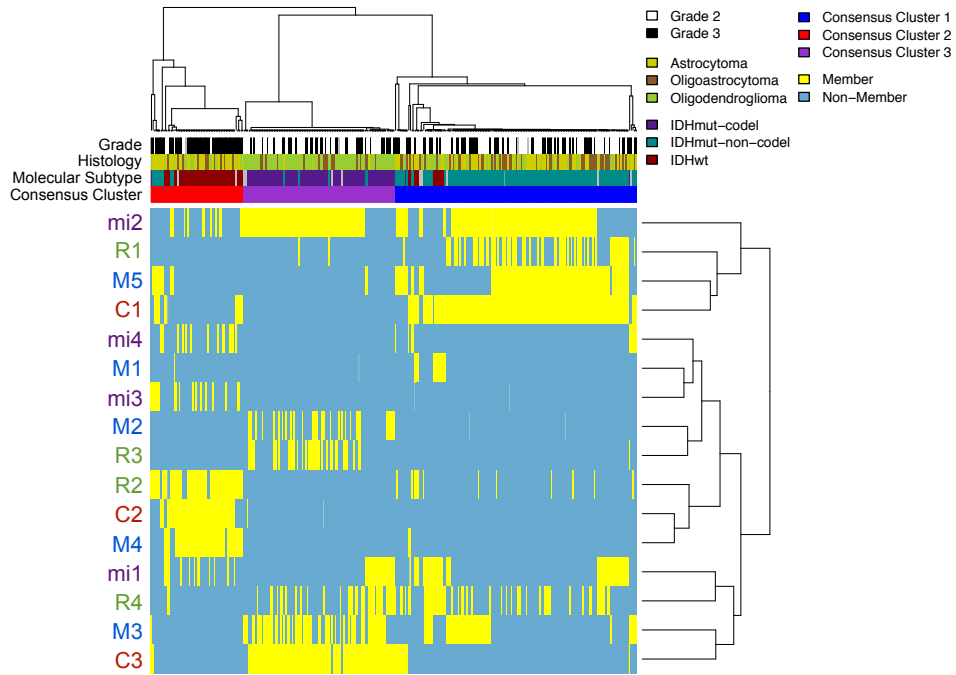
+ One case (IDH<sub>wt</sub>) was in the cerebellum, three cases were in the occipital lobe (2 IDH<sub>mut-codel</sub>, 1 IDH<sub>mut-non-codel</sub>) and five cases were listed as "supratentorial, not otherwise specified" (1 IDH<sub>mut-non-codel</sub>, 4 IDH<sub>wt</sub>)

Table 3.1: Clinical characteristics of the sample set. Clinical characteristics are presented in total and by IDH mutant status and 1p/19q co-deletion status within the IDH mutant group. The number of samples with known information is given when complete data are not available. Count, and percentage within group, is given for categorical variables and distributions are compared by Fishers exact test. Mean, SD, and range are given for age at diagnosis and ANOVA was used to compare groups. Significance is noted as ‡ 0.10 < p < 0.05, # p < 0.05, ## p < 0.01.

for a second level cluster of clusters (CoC) analysis. CoC defines a consensus clustering of the 254 LGG with molecular data from all 4 platforms [90] and resulted in 3 CoC clusters representing distinctive biological themes (Figures 3.1 and A.3). We found a strong correlation between CoC cluster assignment and molecular subtype as defined by IDH mutation and 1p/19q co-deletion status: the large majority of IDHwt LGG are in the CoC cluster with overlapping mRNA R2, microRNA Mi3, DNA methylation M4 and DNA copy number C2 groups.

Another CoC cluster contains almost all IDHmut-codel LGG and includes primarily R3, M2 and M3, and C3 clusters. The third CoC cluster was highly enriched for IDHmut-non-codel LGG and includes R1, M5, C1 and Mi1 classes.

To determine the relative strength of clinical LGG classification schemes in capturing the biologic subsets revealed by CoC analysis, we next compared IDH-co-deletion subtype to histology for correlation with CoC cluster assignment. Among 282 samples with both IDH/co-deletion status and CoC cluster assignments, 127/141 (90%) of IDHmut-non-codel tumors mapped to CoC1; 44/56 (79%) of IDHwt tumors to CoC2; and 82/85 (96%) of IDHmut-codel tumors to CoC3. In contrast, comparison of histology to CoC (293 samples) showed that 62/100 (62%) of astrocytomas mapped to CoC1; 51/77 (66%) of the oligoastrocytomas also mapped to CoC1; and 72/116 (62%) of the oligodendrogliomas mapped to CoC3. None of the histologies showed predominant mapping to CoC2. Importantly, the concordance between IDH-codel status and CoC cluster assignment was much greater than that of histology (adjusted Rand indices



**Figure 3.1: Multiplatform analyses point to biologic subtypes defined by IDH mutation and 1p/19q codeletion status.** Cluster of clusters analysis uses the cluster assignments derived from individual molecular platforms to stratify tumors, thereby integrating data from analysis of mRNA (green), miRNA (purple), DNA methylation (blue) and copy number (red). For each sample, membership in a particular cluster is indicated by a yellow tick; non-membership by a blue tick.

0.79 versus 0.20 Table S2E), suggesting that IDH-codel status provides a simple and accurate reflection of specific LGG biologic subsets. Taken together, integrated analysis of 4 genome-wide platforms resulted in three robust and concordant LGG biologic subtypes whose membership could be largely defined by their IDH and 1p/19q co-deletion status; correlation with histologic classes was poor by comparison.

### 3.3.3 Mutational landscape of LGG

We generated a consensus mutation set based on three mutation calling algorithms (see Methods in Supplemental text), yielding 9,885 mutations detected in 289 samples and corresponding to a frequency of 0.66 mutations per Mb in coding regions (median, 29 mutations/sample; range, 0-597). IDHwt LGG exhibited a higher number of mutations per sample (median, 45) than IDHmut-codel LGG (median, 27;  $p < 0.001$ ) or IDHmut-non-codel LGG (median, 28;  $p < 0.001$ ) (Figures S11-13). LGG mutation frequencies were lower than those for GBM; higher than those for medulloblastoma; and intermediate in the spectrum of mutation frequencies reported for TCGA malignancies (see Figure S11) [70]. Mutations in LGG subsets were dominated by C > T transitions, as shown in Figure S12. One IDHwt sample had a relatively high number of mutations (597), possibly associated with somatic missense mutations in DNA repair genes POLE and MLH1.

We identified statistically significant DNA copy number alterations and gene mutations

for each of the three molecular subtypes (Figures 3.2A, 3.3, S14A, A.1; Tables S3, S4). Consistent with prior findings [10] we found CIC mutations in 62% and FUBP1 in 29% of LGG in the IDHmut-codel group, but not in other molecular subtypes, and also observed mutations in PI3 kinase pathway genes, PIK3CA (20%) and PIK3R1 (9%). Other mutations in this subtype included NOTCH1 (31%) [10, 64, 158], and novel mutations in ZBTB20 (9%), and ARID1A (6%)(Figure 3.2A). Among IDHmut-codel samples, TERT promoter mutations were present in 96%; ATRX mutations were extremely rare in this subset, consistent with prior studies showing mutual exclusivity of these mutations [65]. TERT expression was significantly higher in IDHmut-codel than IDHmut-non-codel LGG, as expected from an activating mutation (Figure 3.2B). The IDHmut-codel subtype showed focal amplification 19p13.3 (Figures 3.2B, S14A), but few other recurring whole arm copy number alterations aside from 1p and 19q loss (Figure S8B). Differences in mutational frequency and copy number alterations in grades II and III were modest (Figures 3.2, 3.3A, S21A). Overall, integrated genomic analysis suggests that one route of LGG initiation starts with IDH mutation, followed by 1p/19q co-deletion and TERT activation with deactivation of CIC, FUBP1, and activating mutations/amplifications in the PI3K pathway. NOTCH1 mutations in IDHmut-codel LGG were most likely inactivating, since they occurred in sites similar to inactivating mutations in lung, head and neck and cervical cancers (Figure S15), but not at activation sites, as described in T cell acute lymphoblastic leukemia [150] (Figure S15). PARADIGM-Shift [92] analysis (Figure S16), which evaluates downstream targets to assess pathway status, also suggested NOTCH1 mutations were inactivating in LGG



[22]. Prior studies have identified NOTCH1 mutations in both oligodendroglioma and anaplastic astrocytoma [10, 64, 158]. We identified them most frequently in IDHmut-codel LGG, yet also rarely in IDHmut-non-codel and IDHwt LGG (Figure 3.2A).

Among 141 IDHmut-non-codel cases, 94% harbored inactivating TP53 mutations and 86% had inactivating alterations of ATRX, including mutations (79%), deletions (3%), gene fusion (2%) or two aberrations (2%). TERT promoter mutations were rare (4%) in this group consistent with the alternative lengthening of telomeres (ALT) associated with ATRX mutations [65]. Concurrent somatic alteration of IDH1, ATRX and TP53 suggests the combined deregulation of metabolism, chromatin organization and apoptosis pathways, respectively, as a mechanism of LGG development in this subtype. The mutually exclusive mutational profiles observed between the IDHmut-codel and IDHmut non-codel subtypes likely reflect distinct molecular mechanisms of LGG oncogenesis and suggest a near requirement for either 1p/19q co-deletion or TP53 mutation following IDH mutation. While prior studies have shown enrichment of oligodendrogliomas for 1p/19q codeletion and of astrocytomas for TP53 mutations, with oligoastrocytomas showing weaker associations with these markers, we demonstrate that IDH mutant LGG are either 1p/19q codeleted or TP53 mutated, with minimal overlap or gaps, indicating a strict molecular dichotomy in IDH mutant gliomas [4].

Two significantly mutated genes in the IDHmut-non-codel LGG have not been previously reported: the SWI/SNF chromatin remodeler SMARCA4 (6%), implicated in glioma progression [58]; and the translation initiation factor EIF1AX (<1%), for which

somatic mutations have been found in uveal melanoma [84] (Figure 3.2A; Table S4). IDHmut-non-codel LGG showed focal gains of 4q12, a locus harboring the receptor tyrosine kinase (RTK) PDGFRA; 12q14, involving cell cycle regulator CDK4; and 8q24, possibly targeting MYC (Figure S14A), analogous to prior findings in IDH1-mutant (MYC) and IDH1-wt (CDK4/PDGFR) proneural GBM [15]. Histologic grade III tumors in this subset had greater frequencies of chromosome 9p and 19q losses, as well as 10p gains, suggesting these events may occur with progression (Figure 3.3A). Mutational profiles in this subset did not differ substantially between grades II and III (Figure S21B).

### 3.3.4 Signaling networks in LGG

To use a multiplatform approach for classifying LGG that also incorporates their mutational landscape, we performed OncoSign analysis [27] using 64 selected functional genetic and epigenetic events. This approach identified four dominant tumor subtypes (OSC1, 2, 3 and 4), which again largely recapitulated those defined by IDH mutation and 1p/19q status (adjusted Rand index, 0.83) (Figure 3.2B; Table S2E). OSC1 showed strong correlation with IDHmut-non-codel and OSC4 contained exclusively IDHwt LGG. The IDHmut-codel group contained both OSC2 and OSC3 LGG, which differed by the presence of mutations in CIC, FUBP1 and NOTCH1, yet were similar in terms of tumor grade and patient outcome. The concordance of IDH mutation and 1p/19q status with two multiplatform genomic data integration approaches (CoC

and OncoSign) is striking and contrasts sharply with the much weaker relationship between histology and unsupervised multiplatform classes (Table S2E). Given the high inter-observer variability of histopathologic diagnosis [1, 144], the finding that clinically available markers (IDH, 1p/19q) are capable of classifying LGG in a manner similar to the unsupervised stratification of genome-wide data suggests that IDH and 1p/19q status should be implemented as clinical classifiers of LGG on a routine basis.

### 3.3.5 IDHwt LGG and GBM

Seven gene mutations were highly associated with IDHwt LGG and five of these seven were previously reported in the TCGA analysis of GBMs [15]: PTEN (23%), EGFR (27%), NF1 (20%), TP53 (14%) and PIK3CA (9%). We also found novel mutations in protein tyrosine phosphatase PTPN11 (7%) and phospholipase C gamma-1 (PLCG1; 5%) (Figure 3.2A). Similarly, copy number alterations in IDHwt tumors were distinct from those of IDH-mutant tumors, and instead resembled IDHwt GBMs [15] (Figure 3.3A). In particular, chromosome 7 gain and 10 loss co-occurred in > 50% of tumors in this subtype (chr7: 56% and chr10: 63%), yet were absent in IDHmut groups. Recurring focal amplifications containing EGFR, MDM4 and CDK4 (38%, 13% and 7% respectively) and focal deletions targeting CDKN2A and RB1 (63% and 25%) were the most frequent acquired copy number variants in the IDHwt subtype, and also mirrored frequencies in IDHwt GBM (Figure 3.3B). These findings raise the possibility that some clinically identified IDHwt LGG represent surgically undersampled GBMs; conversely,



the genomic similarity of IDHwt LGG to IDHwt GBM could also indicate that IDHwt LGG represent an immediate precursor. In either case, IDHwt status in a surgically sampled diffuse glioma lacking histologic criteria for GBM identifies an aggressive tumor, most often with clinical and genetic attributes of GBM; this represents a dramatic improvement over a diagnostic practice based on histology alone.

### 3.3.6 Genomic rearrangements and fusion transcripts

We investigated 20 samples with high level whole genome sequencing, 50 samples with low-pass whole genome sequencing and 311 samples with sequencing of coding regions only, for the presence of structural chromosomal variants such as translocations and inversions. This analysis uncovered 250 high-confidence chromosomal rearrangements, of which 110 had both ends of the breakpoint within a protein-coding gene (Table C.3). Additionally, 19 samples showed evidence of extrachromosomal DNA structures known as double minute chromosomes/breakpoint-enriched regions (DM/BER) (Table C.3, Figure A.2). Of these, 15 occurred in IDHwt samples (27%) (Figure 3.3A), similar in frequency to that in GBM (23%) [15, 121, 161]. Analysis of RNA sequencing data [141, 87] identified fusion transcripts in 265 LGG tumors (Table S6). Correlating fusion events to structural genomic variants suggested chimeric transcription for 44% of the high-confidence chromosomal rearrangements including two fusions involving EGFR (Figure 3.3B), and 58% of the DM/BER events. Thirty-seven transcript fusion events involved protein kinases, and eleven affected significantly mutated genes.

Several genes (EGFR, FGFR3, NOTCH1, ATRX, and CDK4) were affected by fusions in multiple samples (Figure 3.3C). Activating RTK fusions of EGFR and FGFR3, previously found in GBMs [15, 128, 101], were restricted to IDHwt LGG at frequencies comparable to GBM (7% and 3%, respectively; Figures 3.3C, S18, S19) [15]. A chimeric and novel FGFR3-ELAVL3 transcript involved the same breakpoint as previously reported FGFR3-TACC3 fusions and was highly expressed, suggesting similar effects on FGFR3 function. Three samples had fusions between EGFR and intergenic or intronic regions on chromosome 7 that are predicted to remove the EGFR autophosphorylation domain and are likely oncogenic (Figure 3.3C) [26]. Gene fusions involving RTKs were predominantly a feature of IDHwt LGG, as only two IDHmut-non-codel tumors harbored RTK fusions (PDGFRA and MET), and none were identified in IDHmut-codel samples.

### **3.3.7 LGG protein expression**

Reverse-phase protein lysate array (RPPA) analysis resulted in protein expression profiles that demonstrated a striking segregation of IDHwt from IDHmut LGG, and a consistent pattern of RTK pathway activation within IDHwt tumors, including EGFR phosphorylation, providing additional support that IDHwt LGG are biologically similar to GBM (Figures 3.3D, S20). The overexpression of HER2 in this subset represents a potential therapeutic target. Among the IDHmut-LGG, higher expression of hematopoietic markers Syk, E-cadherin and Annexin1 was noted in the non-codel

group, whereas higher phosphorylated HER3 (PY1289) was identified in the IDHmut-codel group. HER3 overexpression may be associated with resistance to PI3K inhibitors [42].

### 3.3.8 From signatures to pathways

To gain insights into signaling pathways, we examined mutations, focal copy number alterations, structural variants and gene fusions affecting RTKs (EGFR, PDGFRA, MET, FGFR), PI3 kinase subunits, MAP kinases, NF1 and BRAF, components of the p53 and RB1 pathways (MDM2, MDM4, MDM1, CDKN2A/B and CDKN2C), and ATRX. Alterations across these loci were remarkably similar in frequency between IDHwt LGG and IDHwt GBMs, but not other LGG subtypes (Figure 3.3B, 3.4A). Forty-three percent of IDHwt LGG and 53% of IDHwt GBM tumors harbored an alteration in EGFR, with EGFR amplification being the most frequent in both (Figure S21). Homozygous CDKN2A/B deletions occurred in 45% of IDHwt LGG, similar in range to IDHwt GBM (55%). This contrasts with IDHmut-non-codel LGG, in which large single copy deletions of chromosome 9p were common, yet CDKN2A/B was homozygously deleted in only 4% (Figure A.1). IDHmut tumors lacked IDHwt-GBM-associated cancer pathway aberrations and instead had their own characteristic cancer pathway alterations that included TP53 and ATRX (IDHmut-non-codel) and TERT, NOTCH1, CIC and FUBP1 (IDHmut-codel) (Figure 3.4A). Grade II IDHwt LGG were uncommon (13 cases), yet differed in their mutational profile from grade III (Figure 3.3A, S21C, D). Grade II





IDHwt LGG were highly enriched with tumors from the small, discrete M1 methylation cluster (Figure 3.1, S1A), which display a low frequency of mutations and copy number alterations, and do not have TERT promoter mutations, potentially indicating they are distinct pathologic entities. TERT promoter mutations were present in 64% of the IDHwt LGG; when M1 LGG are removed, TERT promoter mutations were present in 80% of those remaining, consistent with frequencies in primary GBM [65].

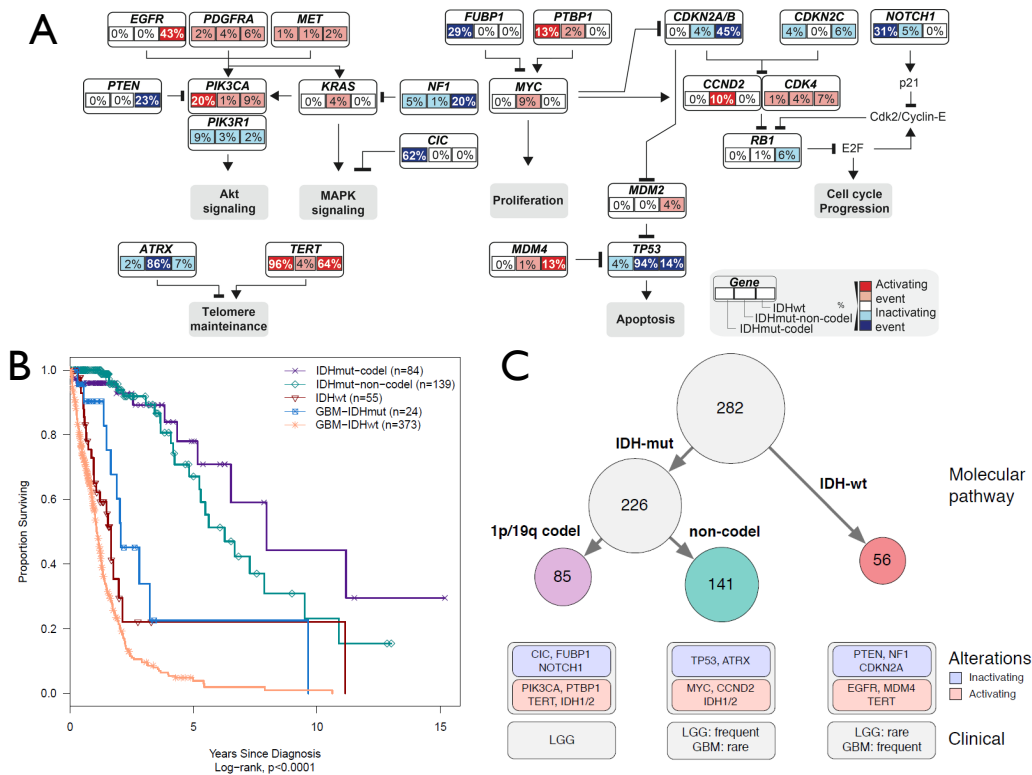
### **3.3.9 Clinical characteristics and outcomes of molecular subtypes**

Patients with IDHwt LGG were older than those with IDHmut LGG and had a more frequent family history of cancer (Table 3.1). Anatomic locations also differed, with IDHmut LGG arising in the frontal lobes more often than IDHwt. Among 289 cases with clinical follow-up, 77/250 (31%) experienced recurrence, and 60/289 (21%) were deceased at the time of analysis. Consistent with molecular findings that showed similarity of IDHwt LGG to GBM [50], survival analysis showed that patients with IDHwt LGG had substantially shorter overall survival than IDHmut LGG (age adjusted hazard ratio 7.4, 95% CI (4.0, 13.8)) and had a prognosis similar to IDHwt GBMs (Figure 3.4B, Table S2D). In comparison, LGG in the IDHmut-codel and non-codel groups were associated with relatively prolonged median survivals of 8.0 years and 6.3 years, respectively. Molecular classification of LGG as IDHwt, IDHmut-non-codel and IDHmut-codel stratifies patient outcomes in multi-predictor models after adjusting for age and extent of resection (Tables S2B-D). Grade, but not histological diagnosis, remains a signifi-

cant predictor of outcome in multivariable models with IDH/codel status and provides additional prognostic value among the molecular subsets (Table S2; Figure S22). Together the results point to three robust LGG subtypes, each with prototypic molecular alterations and distinctive clinical behaviors (Figure 3.4C).

### 3.4 Discussion

This integrative analysis of 293 diffuse LGG employed a comprehensive, multiplatform approach to delineate biological foundations and establish potential therapeutic targets. The findings are striking in their clarity and consistency: 1) unsupervised classification from diverse genome-wide analytic platforms provides strong biologic justification for subtyping LGG based on IDH and 1p/19q status as an improvement over conventional histologic classification; 2) IDH mutant LGG are either 1p/19q codeleted or TP53 mutant, with very few exceptions; and 3) the majority of IDHwt-LGG showed remarkable genomic and clinical similarity to primary GBM. Our results have several immediate clinical ramifications. Current management of LGG patients is based on histologic classification schemes that suffer from significant interobserver variability [31, 143, 1]. Nevertheless, critical decisions regarding therapy depend on these distinctions [14, 120]. Our findings distilled the six possible histology and grade combinations into three robust and non-overlapping molecular subtypes, laying the groundwork for a reproducible and clinically relevant classification that incorporates molecular data into the patho-



**Figure 3.4: Integrative analysis and clinical outcome.** Panel A demonstrates pathway alterations of LGG as a schematic that summarizes genomic alterations-mutations and focal amplifications and homozygous deletions-in LGG across molecular subtypes. Canonical RTK/PI3K/MAPK, RB and p53 regulatory pathways are frequently altered, as is telomere maintenance. However, alterations occur in different pathway components depending on the LGG subtype. Panel B depicts the Kaplan-Meier estimates of overall survival for the LGG samples classified by IDH mutation and 1p/19q co-deletion as well as GBM samples (from previously published TCGA data [15]) classified by IDH mutational status. Age-adjusted analysis can be found in Supplemental Table 2, while further division by histologic grade is present in Figure S22. Overall survival times and IDH mutation status of GBMs in this analysis were as previously described [15]. Panel C summarizes major findings and conclusions, where consensus clustering yields three robust groups based on IDH and 1p/19q co-deletion status with stereotypic and subtype-specific molecular alterations and distinct clinical presentation.

logic diagnosis, as planned for the upcoming revision of the WHO classification of brain tumors [79]. The primary subtype discriminator for LGG was the mutational status of IDH1/IDH2, with all molecular platforms demonstrating a sharp contrast between IDHwt and IDHmut tumors.

Whereas IDHwt tumors showed an enrichment of astrocytomas and IDHmut-codel LGG were predominantly oligodendrogliomas, the largest molecular subclass of LGG (IDHmut-non-codel) contained all histologic types, highlighting the non-specificity inherent to pure histologic classification. By comparison, two unsupervised methods using multiplatform analysis to classify LGG (CoC and OncoSign) yielded strong correlations with IDH/codel status (adjusted Rand index 0.79 and 0.83, respectively), suggesting that such molecular classification captures biological classes more accurately than histology. Additionally, while oligodendrogliomas were typically 1p/19q co-deleted and astrocytomas generally were not, oligoastrocytomas were distributed among the three molecular subtypes with no molecular feature distinguishing them. Thus, while previous WHO classification systems have recognized LGG with mixed histology (oligoastrocytoma), molecular classification indicates that IDH mutant LGG are either 1p/19q co-deleted or TP53 mutant in a nearly mutually exclusive fashion, and does not support the existence of this entity biologically (Figure 3.4C) [152, 118]. Not only will the use of molecular signatures for LGG lead to a practice-altering, biologically-based diagnosis with greatly improved interobserver concordance; a phase-out of the diagnosis oligoastrocytoma and the confusion related to its clinical management is a major advance

stemming from our results.

Another substantial finding was that IDHwt tumors were molecularly and clinically distinct from IDHmutant subtypes, with most showing a striking resemblance to primary GBM on all analytic platforms. The findings suggest that IDHwt LGG may represent immediate precursors of IDHwt GBM that have similar clinical properties; alternatively, such tumors could represent actual GBMs that were incompletely sampled during surgery, precluding definitive histologic classification. Sampling errors represent a significant challenge in surgical neuropathology, regardless of IDH status, class or grade, since a morphologic diagnosis is limited to the findings under the microscope. Thus, molecular classification based on IDH and 1p/19q status represent an improvement in diagnostic practice, since it enables the identification of a clinically aggressive form of LGG (IDHwt) in the absence of morphological criteria for GBM.

Consistent with earlier work [57, 20, 151], analysis of clinical outcomes demonstrated that IDHmut-noncodel LGG performed somewhat less favorably than IDHmut-codel LGG with both subtypes exhibiting substantially longer overall survival than their IDHwt counterparts. Patients with IDHmut-codel LGG had a median survival of 8.0 years, compared to the IDHwt subgroup, which had median a survival of 1.1 years (Figure 3.4B). Compared to outcome prediction based on histologic classification (Figure S22), stratification of clinical risk based on IDH and 1p/19q status is more robust. In addition to reducing inter-observer variability and predicting risk, molecular classification can provide quality control for histopathologic diagnosis. As an example, the small,

discrete DNA methylation cluster M1 in our cohort showed a low frequency of mutations and copy number alterations and contained occasional tumors with BRAF alterations. While not entirely specific, these alterations are more characteristic of grade I circumscribed tumors, such as pilocytic astrocytoma and ganglioglioma, and its finding would lead to a consideration of alternative diagnoses. Given that histopathologic features of diffuse versus circumscribed gliomas can overlap, while clinical management differs greatly, molecular signatures offer the potential to resolve diagnostically challenging cases. Further analysis of our cohorts survival data as it matures will be required to designate precise metrics by which WHO grade may combine with molecular signatures to optimize risk stratification for LGG.

The variable clinical course of LGG patients, coupled with the three sharply contrasting molecular profiles, suggest that distinct therapeutic strategies may be required for effective disease control. Molecular inclusion criteria and stratification in clinical trial design will be necessary for clear interpretation of outcomes from specific treatments. The prevalence of IDH mutations in LGG invites targeting of either the mutant enzymes themselves or the consequences of downstream metabolic and epigenomic derangement, specifically G-CIMP, and such efforts are already underway [115, 142]. ATRX, CIC, and FUBP1 mutations have only recently been implicated in cancer biology; yet their specificity and prevalence in IDHmut LGG support central roles in oncogenesis and argue for thorough characterization of associated signaling networks to facilitate therapeutic development. The remarkable genetic and clinical similarities between IDHwt-LGG and

primary GBM support the potential inclusion of IDHwt LGG within the broad spectrum of GBM-related clinical investigation and treatment protocols. The extent to which IDHwt LGG resemble GBMs in their response to recently developed therapies has yet to be determined. Finally, our integrative analysis has shown that all LGG subtypes rely to some extent on core signaling networks previously implicated in GBM pathogenesis. Many agents targeting these pathways are currently available and in clinical trials and may prove effective against LGG subtypes.

## Chapter 4

# Mutation Independent Activation of The Notch Pathway is Associated With Lapatinib Resistance in Her2+ Breast Cancer Cell Lines

Cancer can not only progress through genomic alterations, but can also progress through non-genomic resistance to treatment. This chapter describes a mutation independent mechanism of resistance to a commonly used HER2 tyrosine kinase inhibitor, lapatinib, in HER2+ breast cancer cell lines. These findings show that the Notch pathway is associated with this resistance and suggest that Notch inhibitors may be beneficial in treating patients with lapatinib resistance.



This chapter contains text from the paper "Mutation Independent Activation of the Notch Pathway is Associated with lapatinib Resistance in Her2+ Breast Cancer Cell Lines" by Mia Grifford, Melanie Plastini, Sol Katzman, Sofie R. Salama and David Haussler in preparation (to be submitted to PLOS One in 2014). I performed the majority of this work with the help of an undergraduate whom I trained, Melanie Plastini, who helped with the cell culture experiments and performed some of the quantitative RT-PCR and Western blot experiments (sections 4.2.1, 4.2.2 and 4.2.3). Sol Katzman performed the mapping of RNA-sequencing, running RSEM and running DE-Seq (figure B.5 and section 4.2.5), and Amie Radenbaugh ran the mutation calling with the tool she developed, RADIA (section 4.2.8), while I analyzed and presented the results.

## 4.1 Introduction

Breast cancer patients are divided into subtypes based on molecular aberrations and activation of key proteins [90]. These subtype classifications help guide treatment. About 20% of patients fall into the HER2+ subtype, which have an activation of the receptor tyrosine kinase human epidermal growth factor 2 (HER2, also known as ERBB2) [106]. These patients are commonly treated with a monoclonal antibody for HER2, trastuzumab, or a HER2 tyrosine kinase inhibitor, lapatinib, in combination with chemotherapy and sometimes other non-chemotherapy drugs [108]. Despite these treatments, HER2+ patients have overall poor survival when compared with other subtypes

[62, 102]. Often poor patient survival is associated with drug resistance. There have been many proposed mechanisms of resistance to targeted therapies like trastuzumab and lapatinib, including activation of alternate survival signaling pathways, such as insulin growth factor (IGF) or phosphatidylinositide 3-kinase (PI3K) [56, 80, 52, 12], but a complete explanation of resistance remains elusive. An understanding of the mechanisms underlying resistance to targeted therapies is critical for increasing their effectiveness and developing strategies to treat resistant tumors.

Many patients will respond to re-treatment of a drug after displaying resistance and being taken off the drug for a period of time [66, 156, 23]. These observations suggest that resistance is reversible in these patients and is not associated with resistance-conferring mutations. Indeed, a reversible, epigenetic mechanism of resistance to a targeted therapy was observed in non-small cell lung cancer (NSCLC) cell lines, which had activated receptor tyrosine kinase epidermal growth factor receptor (EGFR) [126]. EGFR (also known as HER1) is in the same family of receptor tyrosine kinase proteins as HER2 [116] and can be targeted by the EGFR inhibitor erlotinib, which, like lapatinib, is a tyrosine kinase inhibitor. These NSCLC resistant cell lines had active insulin-like growth factor 1 receptor (IGF-1R) signaling, overexpression of the histone demethylase KDM5A, differences in bulk chromatin marks and displayed hypersensitivity to histone deacetylase inhibitors. We wanted to determine if a similar, non-mutational, mechanism of resistance occurs in HER2+ breast cancer cell lines, and further to identify specific signaling pathways involved in this type of resistance.

We found that breast cancer cell lines do indeed acquire lapatinib resistance with features similar to the NSCLC cell line resistance to Erlotinib described in [126]. We could not identify expressed mutations in genes known to be involved in drug resistance, suggesting an epigenetic mechanism for resistance. Transcriptome analysis revealed that the Notch pathway is involved in this mutation-independent mechanism of resistance, similar to what has been shown in other models of resistance [160, 117, 25, 77, 83, 100, 97, 113]. Analysis of downstream targets of Notch suggests that up-regulation of Notch signaling may lead to changes in chromatin structure through activation of the insulin-like growth factor 1 receptor (IGF1R) and lysine (K)-specific demethylase (KDM) family proteins.

## 4.2 Methods

### 4.2.1 Cell culture and generation of lapatinib resistant cell lines

HCC1419 and BT474 (ATCC, CRL-2326: 1/31/11, HTB-20: 4/7/11) were grown in RPMI containing 1% penicillin streptomycin (PenStrep) and 10% Gibco certified fetal bovine serum (FBS) (Invitrogen) (RPMI-complete) or RPMI-complete conditioned with mouse embryonic fibroblasts (MEF). MEF-conditioned media was generated by incubating mitomycin-C inactivated MEFs (55,000 cells/cm<sup>2</sup>) with RPMI-complete for 24 hours before collecting. MEFs were initially plated in DMEM supplemented with 10% FBS and PenStrep (Invitrogen) and were used for a total of 6 days before discarding. MEF-

conditioned RPMI was stored at 4°C until the final day of collection and then pooled, filtered and stored at -80°C. To generate lapatinib resistant lines, cells were plated in MEF-conditioned RPMI-complete. 24 hours after plating, the media was replaced with MEF-conditioned RPMI-complete supplemented with 1  $\mu$ M lapatinib. The media was changed every 3 days. On day 10, drug tolerant persisters (DTPs) were isolated. Two independent resistant lines for HCC1419 (drug tolerant expanded persisters, DTEPs) were grown and isolated at passage numbers 7 and 12 (line 1), and 6 (line 2) for RNA sequencing. For experiments testing sensitivity to trichostatin A (TSA), BT474 cells were grown in 6-well dishes containing 3 wells each of RPMI-complete, RPMI-complete supplemented with 75 nM Trichostatin A (TSA), RPMI-complete supplemented with 1  $\mu$ M lapatinib or RPMI-complete supplemented with both 75 nM TSA and 1  $\mu$ M lapatinib. Pictures were taken at 9 days or 72 days. Cells were counted 3 times for each picture and averaged, the average of 7 pictures was used in figure 2d.

#### **4.2.2 Quantitative RT-PCR in drug tolerant persisters**

RNA from about 500,000 cells each of untreated HCC1419 cells, HCC1419 cells treated with 1  $\mu$ M lapatinib for 24 hours and HCC1419 cells treated with 1  $\mu$ M lapatinib for 9 days (DTPs) was isolated using Trizol following the manufacturers guidelines. Quantitative real-time PCR was performed using the QuantiTect SYBR Green RTPCR Kit (Qiagen), using 20 ng of total RNA per reaction. All reactions were set up in triplicate. Cycling was performed on a Rotor-Gene 6000 (Qiagen) according to the QuantiTect

SYBR Green RTPCR Kit instructions. Primers used are: IGF1R forward: ccattctcatgccttggtct, IGF1R reverse: tgcaagttctggttgctcgag, IGFBP3 forward: gcttctgctggtgtgtgat, IGFBP3 reverse: ggctctacttgctctgcat, KDM5A forward: tccagcctttctacc-caatg, KDM5A reverse: cgtaattgctgccactctga, KDM5B forward: cccacctctccatgatgttc, KDM5B reverse: tagtccctggctgctgttc, KDM5C forward: cttgtctgcctttcccacat, KDM5C reverse: agcccgaaccttcagcttat, GAPDH forward: tgaggccggtgctgagtatg, GAPDH reverse: tggttcacacccatcacacaaac. Ct values were calculated with Corbett 6000 software and the relative enrichment of GAPDH-normalized gene expression levels in lapatinib treated cells was determined as fold change over the gene expression levels in untreated cells.

### **4.2.3 Immunoblotting**

One and a half to two million cells each of untreated HCC1419 cells, HCC1419 cells treated with 1  $\mu$ M lapatinib for 24 hours and HCC1419 cells treated with 1  $\mu$ M lapatinib for 9 days (DTPs) were suspended in NuPAGE LDS sample buffer (Invitrogen) at a concentration of 10,000,000 cells per 1 ml buffer. Histones were extracted by incubating cells in 0.4 N H<sub>2</sub>SO<sub>4</sub> for 30 min, incubating the supernatant in 132  $\mu$ l TCA for 30 min, washing the histone pellet with 500  $\mu$ l acetone and repeating TCA incubation and wash (all steps were performed on ice or in cold room). Histone extractions were loaded on NuPAGE (Invitrogen) 4-12% protein gels for SDS-PAGE. Transfer of proteins onto PVDF membrane was performed as described in the NuPAGE manual. Immunoblotting

was performed using antibodies against H3 (Abcam ab1791, 1:2,500), H3K4me3 (Millipore 07-473, 1:500), H3K27me3 (Millipore 07-449, 1:3,000) and H3ac (Millipore 06-599, 1:1,000). Blots were incubated with SuperSignal West Dura Extended Duration Substrate (Thermo Scientific) according to the manufacturers instructions and visualized on a Biorad Chemidoc MP system.

#### **4.2.4 RNA sequencing library preparation**

RNA sequencing was performed on 2 biological replicates each of untreated parental HCC1419 cells (passage # 31), 2 DTP replicates and 3 DTEP replicates (line 1 passage #s 7 and 12 and line 2 passage # 6). RNA was isolated using Trizol following the manufacturers guidelines. RNA was treated with RQ1 DNaseI (Promega) for 1 hour at 37 °C and total RNA was cleaned up using the RNeasy Mini kit (Qiagen). For each sample, the non-ribosomal fraction of 5  $\mu$ g of total RNA was isolated using a Ribo-Zero rRNA removal Kit (Epicentre) following the manufacturers protocol (Lit. #309-6/2011). For the non-ribosomal fraction of RNA, 200 ng RNA was fragmented by incubating 30 minutes at 98 °C in RNA storage buffer (Ambion) and then double stranded (ds) cDNA was synthesized as described previously [104] using dUTP in the second strand synthesis and USER digest before amplification to retain strand specificity. Clean-up steps were performed using RNA Clean & Concentrator or DNA Clean & Concentrator kits (Zymo Research). Double stranded cDNA was used for library preparation following the Low Throughput guidelines of the TruSeq DNA Sample Preparation kit (Illumina), with the

following additions. Size selections were performed before and after cDNA amplification on an E-gel Safe Imager (Invitrogen) using 2% E-gel SizeSelect gels (Invitrogen). The cDNA fraction of 300-500 bp in size (including adapters) was isolated and purified. For adapter ligations, 1  $\mu$ l instead of 2.5  $\mu$ l of DNA Adapter Index was used. Indexed libraries were pooled and sequenced on the Illumina HiSEQ platform.

#### 4.2.5 RNA sequencing data mapping and analysis

The input Illumina fastq files consisted of paired end reads with each end containing 100 bp. These were reduced to 80x80 bp by trimming 20 bp from the 3' end of each read to improve mapping of fragments shorter than 100 nt. Bowtie2 [68] was used to map the fragments against the set of elements from the repeatMasker [131] human library. After filtering out such fragments, bowtie2 was used to map the remaining fragments against a version of the human transcriptome derived from the UCSC Known Genes track [53] of the GRCh37hg19 human assembly by using the target generation script supplied with RSEM [73]. Only mappings with both read ends aligned were kept. Potential PCR duplicates (mappings of more than one fragment with identical positions for both read ends) were removed with the samtools rmdup function [74], keeping only one of any potential duplicates. The final set of mapped paired end reads for a sample were converted to transcript and gene expression estimates using RSEM [73]. Then, DESeq [3] was used to identify differentially expressed genes between sample types (table C.5). For input to DESeq [3] all genes with non-zero counts in any sample were considered.

#### 4.2.6 Gene expression clustering and pathway enrichment analysis

ConsensusClusterPlus [153] was used to run hierarchical clustering using Pearson distance on RSEM [73] expression values of genes found to be significantly differentially expressed by DESeq [3] (Benjamini-Hochberg adjusted p-value less than 0.1) (fig. B.4). Expression values were logged (base 2), and each row (gene) was mean centered and scaled by the standard deviation. Gene set enrichment analysis (GSEA) [138] was run on RSEM expression values to find all pathways differentially expressed between untreated cells and DTEPs from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [60] or the National Cancer Institute Pathway Interaction Database (PID) [122]. All default parameters were used, except to assess significance genes were permuted 1000 times and the log base 2 fold change was used as the metric for ranking genes. Values displayed in figure 3 are logged (base 2), and each row (gene) is mean centered and scaled by the standard deviation.

#### 4.2.7 Accession codes

The RNA-Seq data has been deposited in NCBI's Gene Expression Omnibus [37] and is accessible through GEO Series accession number GSE62074

(<http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE62074>).



### **4.2.8 Mutation calling with RADIA**

Single Nucleotide Variants (SNVs) were identified by RADIA (RNA and DNA Integrated Analysis), a method that typically combines patient matched normal and tumor DNA whole exome sequencing (DNA-WES) with tumor RNA sequencing (RNA-Seq) for somatic mutation detection [110]. In this case, RADIA was executed on RNA-Seq data alone. Several filters were used to eliminate false positives while maintaining true positive calls. All SNVs that overlap with common SNPs found in at least 1% of the samples in dbSNP were removed. Each SNV had at least 10 total reads, a minimum of 4 reads and at least 10% of the reads supported the variant. All standard filters (e.g. strand bias, positional bias, base quality and mapping quality) were applied. In addition, any SNV that overlapped with pseudogenes or retrogenes were filtered out. Lastly, a BLAT [63] filter was applied to guarantee that each read that supported a variant did not align better to another location in the genome.

## **4.3 Results**

### **4.3.1 Prolonged incubation in lapatinib robustly generates resistant cell lines**

In order to better understand the mechanisms behind lapatinib resistance, lapatinib resistant cell lines were generated through long-term treatment with lapatinib. Two

HER2+ cell lines, HCC1419 and BT474 (fig. B.1), were treated with lapatinib continuously. By day 10, untreated cells became completely confluent, while the vast majority of lapatinib treated cells had died. The remaining cells, termed drug tolerant persisters (DTPs) as in Sharma, et al. [126], were alive but not proliferating. After about 40 to 70 days, these cells began to proliferate and could be expanded through multiple passages, forming lapatinib resistant cell lines termed drug tolerant expanded persisters (DTEPs) [126] (fig. 4.1, B.2).

### **4.3.2 Breast cancer DTPs share features with non-small cell lung cancer mutation-independent resistant cells**

The breast cancer cell line DTPs we generated are similar to non-small cell lung cancer (NSCLC) DTPs, which are characterized by a reversible, chromatin-mediated, drug-tolerant state [126]. Like NSCLC DTPs, breast cancer cell line DTPs have increased IGF-1R and IGFBP3 expression, which has been shown to activate IGF-1R [5, 39] (fig. 4.2a). IGF-1R activation has been shown to be necessary for activation of the histone demethylase KDM5A, which also recruits histone deacetylases [93], and is activated in NSCLC DTPs [126]. We measured the RNA expression levels of many KDM5A paralogs, including KDM5A, KDM5B and KDM5C, which share a common domain structure and function [13] (we excluded KDM5D, because it is not expressed in these cells), and found that all were over-expressed in breast cancer DTPs (fig. 4.2b).

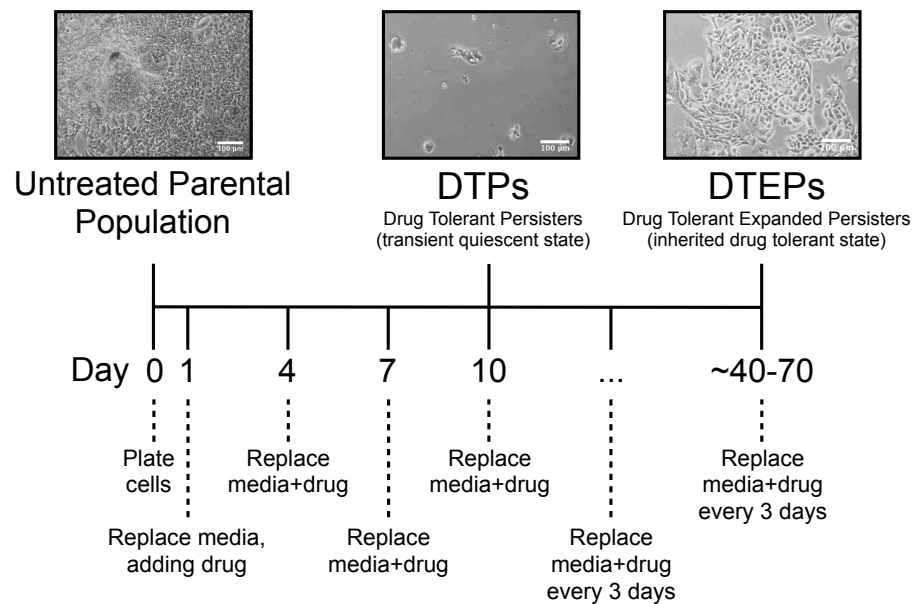


Figure 4.1: **Lapatinib resistant cell lines were generated by long-term lapatinib treatment.** Timeline for generation of lapatinib resistant cells. Lapatinib sensitive HCC1419 or BT474 cells were treated with 1  $\mu$ M lapatinib for 9 days to generate drug tolerant persisters (DTPs, middle), after 40-70 days the cells begin to proliferate and form drug tolerant expanded persisters (DTEPs, right), untreated cells plated at the same initial density proliferate rapidly and become confluent by day 10 (Parental population, left). Micrographs show representative images of the indicated populations of HCC1419 cells.

Also shown in NSCLC DTPs were changes in chromatin marks, consistent with increased amounts of closed chromatin [126]. Breast cancer DTPs do indeed have slight changes in chromatin marks as well, including increased H3K27 tri-methylation, a closed chromatin mark, and decreased H3K4 tri-methylation and total H3 acetylation, both open chromatin marks (fig. B.3).

Because KDM5A, KDM5B and KDM5C are over-expressed in DTPs and H3 acetylation is decreased in DTPs, changes in chromatin organization mediated by histone acetylation state may be critical for lapatinib resistance in this setting. We hypothesized that a histone deacetylase inhibitor may affect DTP sensitivity to lapatinib, as was shown in NSCLC DTPs. We treated resistant cells (DTPs) with the histone deacetylase inhibitor, Trichostatin A (TSA). As was expected, TSA alone did not affect cell growth, but the combination of TSA and lapatinib prevented DTEP formation (fig. 4.2c-d). From these results we can infer that histone deacetylation, and the resulting changes in chromatin, is necessary for resistance.

### **4.3.3 RNA-Sequencing reveals significant changes in gene expression**

We then wanted to determine which genes and pathways were involved in drug resistance. We performed RNA sequencing (RNA-seq) on untreated HCC1419 cells, HCC1419 DTPs, and HCC1419 DTEPs (2 replicates of each, plus an additional time point of one DTEP line - see methods). In order to get gene expression estimates from

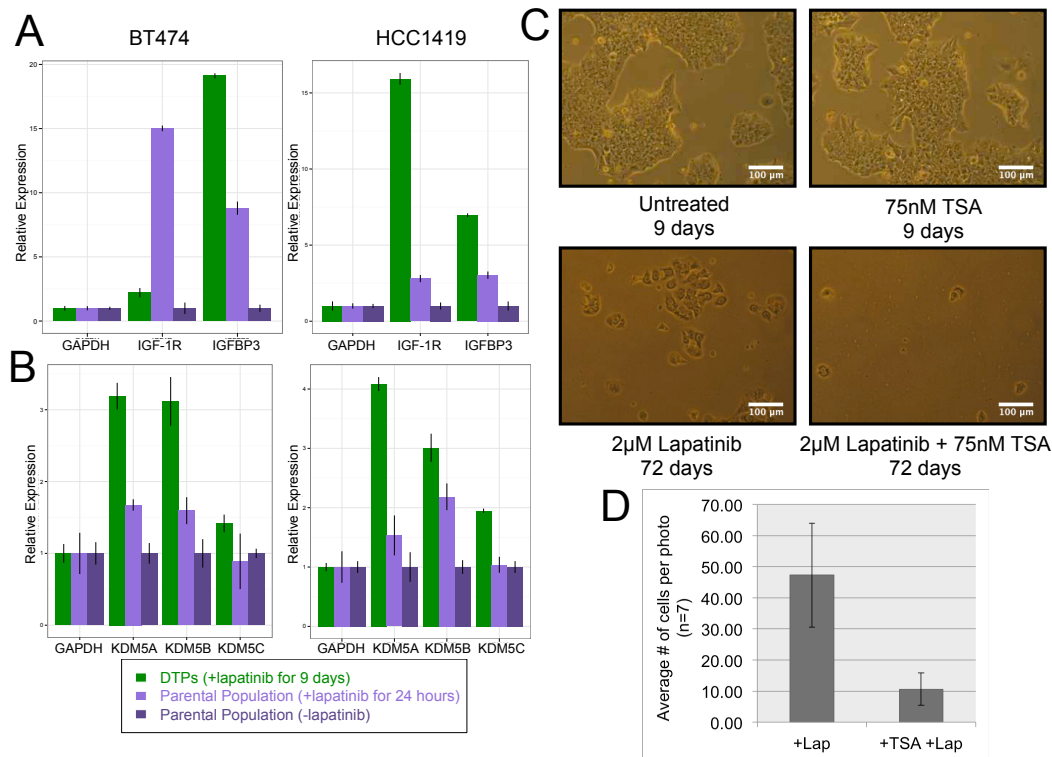


Figure 4.2: **Breast cancer DTPs share features with non-small cell lung cancer DTPs.** A,B: Relative expression measured by quantitative reverse-transcriptase PCR of IGF-1R and IGFBP3 (A) or KDM5A, KDM5B and KDM5C (B) normalized to GAPDH in DTPs (green), parental population treated with lapatinib for 24 hours (light purple) and untreated parental population (dark purple) shows activation of IGF-1R and over-expression of KDM5A paralogs in resistant HCC1419 and BT474 cells. C: Representative micrographs of BT474 cells untreated or treated with 75 nM Trichostatin A (TSA) and/or 2  $\mu$ M lapatinib shows that TSA prevents formation of DTEPs when combined with lapatinib, but TSA alone does not affect cell growth. D: Quantification of cell counts from 7 micrographs as in (C) of cells treated with 2  $\mu$ M lapatinib or 2  $\mu$ M lapatinib and 75 nM TSA for 72 days. Error bars represent standard deviation.

the sequencing data, we used RSEM - RNA-Seq by Expectation Maximization [73]. We then performed differential gene expression analysis using DESeq [3], and found 6,943 significantly differentially expressed genes (fig. B.4). The DESeq results revealed that the replicates were more similar to one another than they were to other samples (fig. B.5). Surprisingly, DTP samples were more similar to untreated samples than they were to DTEP samples, suggesting that significant gene expression changes were necessary for proliferation to occur in resistant cells. In addition, the DESeq results confirmed that IGF-1R pathway genes are over-expressed in resistant cells (Table B.1). We also looked at the KDM family of genes and found that while KDM5A paralogs are transiently over-expressed in DTPs, this does not persist in DTEPs. Other KDM genes (KDM4B and KDM6B) are over-expressed in DTEPs, suggesting a prolonged effect on histone modification.

#### **4.3.4 Resistance is mutation independent**

We next wanted to explore whether genetic mutations were underlying these changes in gene expression. Because we had performed RNA-Seq, we were able to determine if there were any expressed mutations in genes known to be involved in drug resistance. We ran a mutation calling algorithm named RADIA [110] to detect expressed, somatic mutations from the RNA-Seq data. Table C.4 lists all expressed mutations in any DTEP replicate, but unaffected in either untreated replicate. We then searched this list for mutation calls in genes that could be involved in resistance. We searched for

over 75 genes known to be involved in drug resistance (Table B.2) [52, 12, 47]. We also searched the list for any RAS, RAF, IGF, IGFBP, PI3K, or MAPK genes (not including interacting proteins). Only one resistance-related gene, IGFBP5, contained a mutation call, and this only in a single DTEP replicate, and not in either untreated replicate. This gene is lowly expressed in most samples and, upon further investigation, was found to have some evidence of mutation in all 7 samples (including both untreated and both DTP replicates). The position of this mutation call is also toward the end of the 3' UTR, which is a highly repetitive region prone to sequencing and mutation calling errors, and there are no other cancer-related mutations at this location found in the Catalogue of Somatic Mutations In Cancer (COSMIC) [41]. These findings suggest that this mutation call is likely to be an artifact.

Because there were no expressed mutations likely to cause lapatinib resistance, and because DTEP formation was prevented with a histone deacetylase inhibitor and breast cancer DTPs have high IGF-1R, IGFBP3 and KDM5A, KDM5B and KDM5C expression similar to NSCLC DTPs, we conclude that it is likely that the lapatinib resistance is mutation independent and instead due to epigenetic changes as observed in erlotinib resistant NSCLC cell lines [126].

### 4.3.5 The Notch pathway is over-expressed in resistant cells

To identify pathways over-expressed in resistant cells, Gene Set Enrichment Analysis (GSEA) [138] was performed. We found all pathways that were significantly over-expressed in DTEPs when compared with untreated cells from the Kyoto Encyclopedia of Genes and Genomes (KEGG) [60] or the National Cancer Institute Pathway Interaction Database (PID) [122]. Both the KEGG and PID Notch pathways were found to be significantly over-expressed in DTEPs compared to untreated cells. Notch1 and Notch3 are over-expressed in DTEPs, as well as several genes downstream of Notch, including HES1, MYC, FOS and JUN [99, 51]. Other significantly overexpressed pathways are Hedgehog, PDGFRA, IL2 and PI3K. Several genes that are in these pathways have previously been shown to interact with Notch and are involved in several other pathways, including JUN, FOS, MYC, EP300 and PIK3CA. All enriched pathways with a FDR q-value below 0.1 are shown in figure 4.3, and genes involved in 3 or more pathways are highlighted. An alternate method, involving running functional annotation on DESeq [3] significantly differentially expressed gene lists, using the Database for Annotation, Visualization and Integrated Discovery (DAVID) [55, 54], found similar results (data not shown).

In order to determine how closely connected the differentially expressed genes within these pathways were, we used Cytoscape [29, 119, 125, 132] to visualize the pathway connections between all “core” genes in pathways with an FDR q-value below 0.1 (fig.



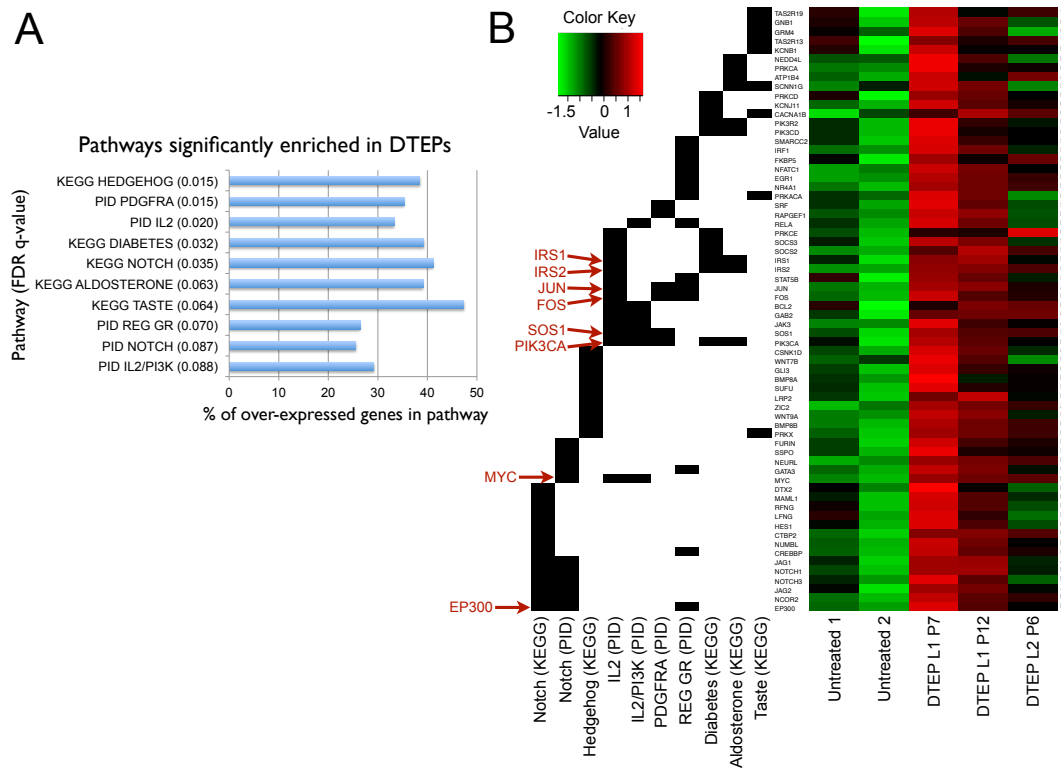


Figure 4.3: **Notch, PI3K and FOS/JUN pathways are over-expressed in DTEPs compared to untreated cells.** A: Pathways found to be over-expressed in DTEPs compared to untreated cells by GSEA [138] with a FDR q-value below 0.1. Bar chart shows the % of genes in the pathway that are over-expressed. B: Heatmap of normalized RSEM [73] expression values of all genes in the 10 pathways found to be significantly over-expressed in DTEPs compared to untreated cells. Expression values were logged (base 2), and each row (gene) was mean centered and scaled by the standard deviation. Black side bars show pathway membership for all genes, highlighted genes occur in 3 or more pathways. DTEP L1P7: drug tolerant expanded persisters line #1 passage #7; DTEP L1P12: drug tolerant expanded persisters line #1 passage #12; DTEP L2P6: drug tolerant expanded persisters line #2 passage #6.

B.6, supplementary data C.6). “Core” genes are those that contribute to the GSEA [138] enrichment score, and are therefore over-expressed in DTEPs. Gene nodes were colored based on the DESeq [3] assigned log fold changes and all colored genes have a DE-Seq [3] Benjamini-Hochberg adjusted p-value less than 0.1. Connections between genes were determined using a superimposed pathway, which incorporates pathway information from the National Cancer Institute Pathway Interaction Database, Biocarta, Reactome and KEGG [60, 122, 91, 32]. This analysis revealed that these pathways are highly interconnected, with only 5 genes left unconnected to the rest of the pathway.

Interestingly, of the 5 ligands for Notch, Jag1 and Jag2 are over-expressed in resistant cells, while the other Notch ligands DLL1, DLL3 and DLL4 are not. This is consistent with previous reports, which show that Jag1 over-expression is associated with tumor recurrence in breast cancer patients [113, 30]. Other groups have found Jag1 and Jag2 to be initially lower in Her2+ tumors when compared with other subtypes, and we also found Jag1 and Jag2 to be lower in Her2+ tumors using patient tumor RNA-Seq data from the The Cancer Genome Atlas and the UCSC Cancer Genome Browser [90, 45, 162] (fig. B.7). Indeed the majority of genes in the Notch pathway that are over-expressed in DTEPs are expressed at lower levels in Her2+ tumors than other tumors in the TCGA breast cancer cohort (fig. B.7).

## 4.4 Discussion

Because resistance to targeted therapies in Her2+ breast cancer patients is common, we set out to better understand the mechanisms behind resistance to a Her2 tyrosine kinase inhibitor, lapatinib. Our experiments suggest that the mechanism behind lapatinib resistance in Her2+ breast cancer cell lines is mutation-independent and involves changes in chromatin, similar to Erlotinib resistance in non-small cell lung cancer cell lines [126]. RNA-sequencing revealed many changes in gene expression between untreated and lapatinib resistant cells (DTEPs), but the most significant pathways that were over-expressed in resistant cells were Notch and other closely related pathways. The Notch pathway may be initially activated through over-expression of the Notch ligands Jag1 and Jag2 in DTPs.

Notch is an integral part of nearly all the over-expressed pathways in resistant cells and Notch has previously been shown to be involved in drug resistance to several different breast cancer drugs (including tamoxifen, aromatase inhibitors, VEGF receptor tyrosine kinase inhibitors, trastuzumab and lapatinib)[160, 117, 25, 77, 83, 100, 97, 113]. Thus, Notch activity may be key to development of lapatinib resistance in this setting. Because Notch1 transcriptionally activates IGF-1R [38], which subsequently activates the KDM family of proteins [126, 72, 33, 34], activation of the Notch pathway may be what leads to the changes in chromatin seen in resistant cells (fig. 4.4). Therefore, treating HER2+ tumors with Notch inhibitors alongside HER2-targeted agents may be of therapeutic

benefit.

Gamma secretase inhibitors (GSIs) have been shown to inhibit Notch signaling and have been shown to prevent drug resistance in other settings [96]. There are also several clinical trials underway combining GSIs with other targeted therapies in breast cancer and other cancers. These results describe a mutation-independent mechanism of resistance to lapatinib involving Notch signaling, suggesting that combining GSIs with lapatinib could prevent this type of resistance and thus improve and/or extend the effectiveness of lapatinib treatments.

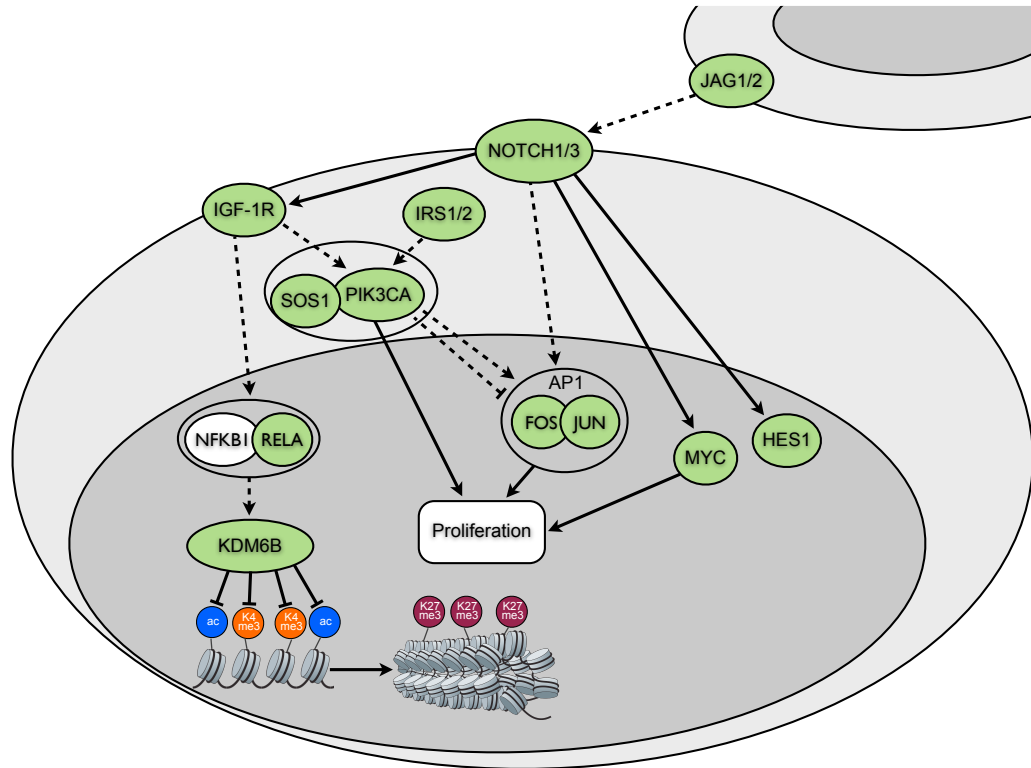


Figure 4.4: **Notch activation may lead to altered chromatin state.** Pathway diagram displaying how the Notch pathway may affect drug resistance through proliferation and chromatin modification. Top is a signal sending cell displaying JAG1/2 and bottom is a signal receiving cell displaying NOTCH. Green nodes are genes over-expressed in DTEPs compared to untreated cells with a DE-Seq [3] Benjamini-Hochberg adjusted p-value less than 0.1. Solid lines represent transcriptional regulation, dashed lines represent post-transcriptional regulation. Ac: histone 3 acetylation; k4me3: histone 3 lysine 4 tri-methylation; k27me3: histone 3 lysine 27 tri-methylation.

## Chapter 5

## Conclusions

High-throughput sequencing technology has allowed for great advances in cancer treatment. Much has been learned about the mechanisms of oncogenesis and new targeted therapies are more successful and have less side effects than traditional chemotherapy and radiation. Though, there are still many problems with cancer treatment and there is still much to learn. Many times cancer treatment is unsuccessful due to late diagnosis, patients not being given the optimum treatment for their particular cancer or drug resistance. These three problems are addressed in this thesis. Each chapter demonstrates another way high-throughput sequencing technology was used in order to make advances in these areas.

In chapter 2, a method for identifying and characterizing double minute chromosomes (DMs) using whole exome sequencing data is described. I used this method to estimate

the prevalence and features of DMs in The Cancer Genome Atlas (TCGA) cohorts of both glioblastoma multiforme (GBM) patients and lower grade glioma patients (LGG). I also used whole genome sequencing to validate predictions made from whole exome sequencing in GBM patients and to reconstruct a DM in an LGG patient. These circular amplicons containing oncogenes explain the mechanism behind cancer progression in about a quarter of all GBM patients and about a quarter of IDH wild-type lower grade glioma (LGG) patients. And, because double minute chromosomes may be detectable in the blood, they present a new opportunity for early diagnosis and monitoring disease progression without needing surgery.

The method I used to detect double minute chromosomes can now be applied to many other types of cancer. In fact, another graduate student in the Haussler lab is working on just that. There are both exome and whole genome sequencing data for large cohorts of patients (generally between 200 to 1,000 patients) for many different cancer types from The Cancer Genome Atlas, and Arjun Rao has been searching for double minute chromosomes in several of these cancers. Initial searches are done using whole exome data, for which there are many more samples, and potential double minute chromosomes are then verified and reconstructed using whole genome data. Once many cancer types have been analyzed, it will give us a greater understanding of how widespread this phenomenon is. We will be able to determine specifically what types of cancers these double minute chromosomes occur in, if and how the features of these DMs vary across cancer types and what genes are involved. This method, if applied more universally, has

the potential to affect diagnosis and monitoring of disease progression in many different types of cancer.

In chapter 3, the comprehensive characterization of LGG patients is described. From which a new method of classification of samples was presented, which more accurately categorizes patients into treatment groups and identifies a group of patients with poor prognosis, who are similar to GBMs and should receive more aggressive treatment. I performed several analyses that support this new method of subtyping patients, including clustering of clusters from various data types, comparing copy number alterations between LGG and GBM patients, comparing alterations in tyrosine kinases between LGG and GBM patients, analyzing genomic rearrangements and analyzing differences in survival between LGG and GBM patients.

Although we have learned a lot about lower grade glioma and how similar it can be to glioblastoma, there is still more to learn. One major question remaining is, if the IDH wild-type LGG patients are so similar to GBM, then why are they being classified as LGGs? What is different between IDH wild-type LGGs and IDH wild-type GBMs? Likewise, we can still ask what are the differences between IDH mutant LGGs and IDH mutant GBMs (which tend to have much better prognosis)? A new TCGA working group was developed to address these questions and many others. They are specifically working on comparing LGG and GBM tumors in more depth than has ever been previously achievable. They are utilizing many different data types, including exome sequencing, whole genome sequencing, RNA-sequencing, copy number, microRNA



sequencing, methylation and reverse phase protein arrays, to analyze cohorts of LGG and GBM patients together. These analyses will lead to a greater understanding of not only the differences between LGG and GBM, but also what causes LGG patients to progress to GBM. Understanding the mechanisms behind progression may allow for new treatments to be used in order to prevent progression to GBM. This could save many lives, because patients can live long, healthy lives with LGG, but once it progresses to GBM, the average survival time is only fourteen months.

In chapter 4, I explain how I developed lapatinib resistant cell lines from HER2+ breast cancer lines, and used these resistant lines to gain a better understanding of the mechanism of resistance to the targeted therapy, lapatinib. I describe a mutation-independent mechanism of drug resistance to lapatinib, which may also be applicable in other settings (different drugs or types of cancer). This mechanism of resistance involves activation of the Notch pathway and therefore may be treatable with gamma secretase inhibitors that are already in clinical trials.

Though, more work will need to be done to determine if combining lapatinib with gamma secretase inhibitors will be successful. We are currently testing if a gamma secretase inhibitor, LY-411575, prevents growth of DTEPs when combined with lapatinib, or if it makes cells more sensitive to lapatinib. We also hope that these findings will encourage other doctors and scientists to continue to study this mechanism of resistance, possibly beginning new clinical trials combining lapatinib with gamma secretase inhibitors. In addition, the RNA sequencing data and the lapatinib resistant cell lines are publicly

available and can be used by other groups who are studying similar mechanisms of resistance. One group in particular, at UCSF, is doing drug screenings with lapatinib resistant DTPs to find drug combinations that would prevent growth of resistant cells. They may be able to utilize the DTEP lines I've generated in their screens, which would model a different type of resistance than the cells they are currently using. If this mechanism of resistance is indeed treatable with Notch inhibitors, it could help save many patients' lives. Also because a similar mechanism of resistance was already shown to occur in non-small cell lung cancer, this mechanism of resistance should be studied in other cancer types with other targeted therapies. This could allow for the discovery of new treatment combinations, potentially affecting a much broader range of patients.

The projects described in this thesis will help to advance current cancer treatment in several ways. We have gained new insight into the mechanism of oncogenesis in a subset of glioblastoma patients which may allow for earlier diagnosis and monitoring of disease progression without the need for surgery. A new method of characterizing lower grade glioma patients, if adopted by the World Health Organization, will make classifying patients more accurate and will lead to more appropriate treatment, especially for the group of LGG patients with poor prognosis. Finally, describing the method of resistance to the targeted therapy, lapatinib, could lead to new, more effective treatment options for patients with resistance. Each of these projects describes one small step toward advancing treatment for cancer patients and ultimately improving patient outcome. In

addition, there are many ways scientists are continuing to build upon this work and utilize these findings to help larger cohorts of cancer patients.

## Appendix A

### Chapter 3 Supplementary Figures

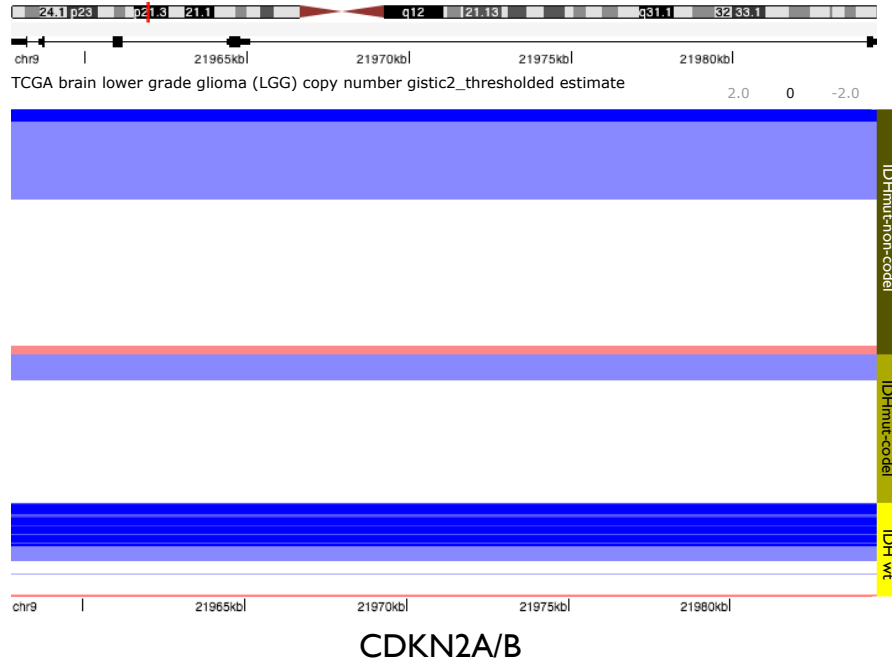


Figure A.1: The UCSC Cancer Genomics Browser view of GISTIC thresholded copy number calls for the region of chromosome 9 containing CDKN2A/B. Samples are sorted based on molecular subtype. About 30% of IDHmut-non-codel samples have deletions in CDKN2A/B, similar to the 29% of samples with chr 9p deletions shown in figure 3.3A, while only 4% passed the cutoff (-2) used for gene-level calls made in figure 3.3B.

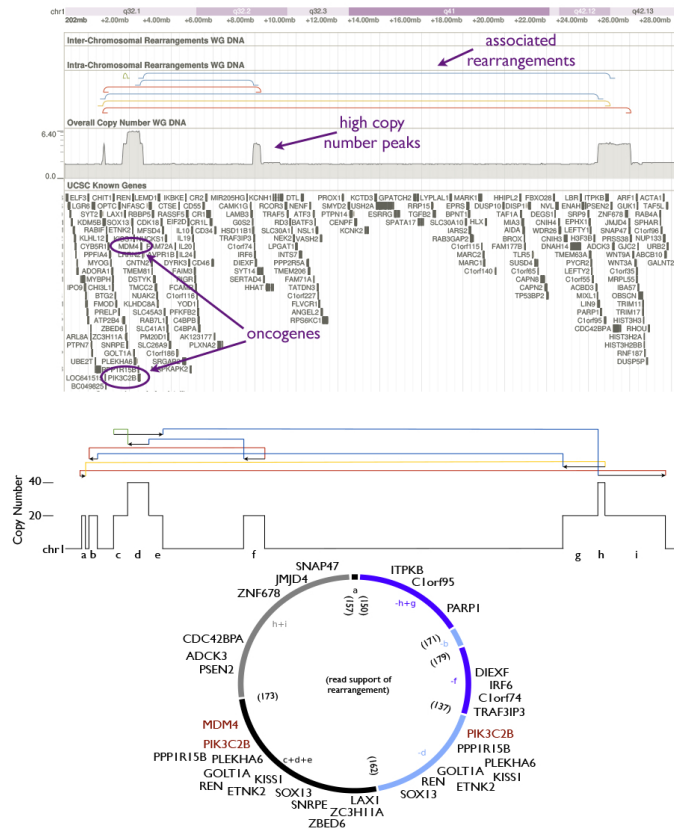


Figure A.2: **Reconstruction of TCGA-CS-5395 circular amplicon** Panel A shows a tumor browser view of the region of chromosome 1 containing the DM/BER. Common features of DMs/BERs are labeled in purple. The bottom track shows protein coding genes within the region. The track above shows relative copy number of the tumor DNA compared with normal, showing distinct blocks of elevated copy number with similar total copy number between blocks. The top two tracks show intra and interchromosomal rearrangement breakpoints. All rearrangements shown are supported by at least 100 discordant reads. The type of rearrangement is indicated by the color of the line: duplication (red), deletion (blue) and inversion (yellow and green). Panel B shows a diagram of the amplified segments and structural variants identified on chromosome 1. Walking through this diagram results in a circular solution, suggesting the double minute chromosome diagrammed in panel C, where segments inverted relative to their orientation in the reference genome are colored blue. The letters inside the circle correspond to the segments in B. The numbers inside the circle indicate the number of sequencing reads supporting each breakpoint. Oncogenes are colored red.

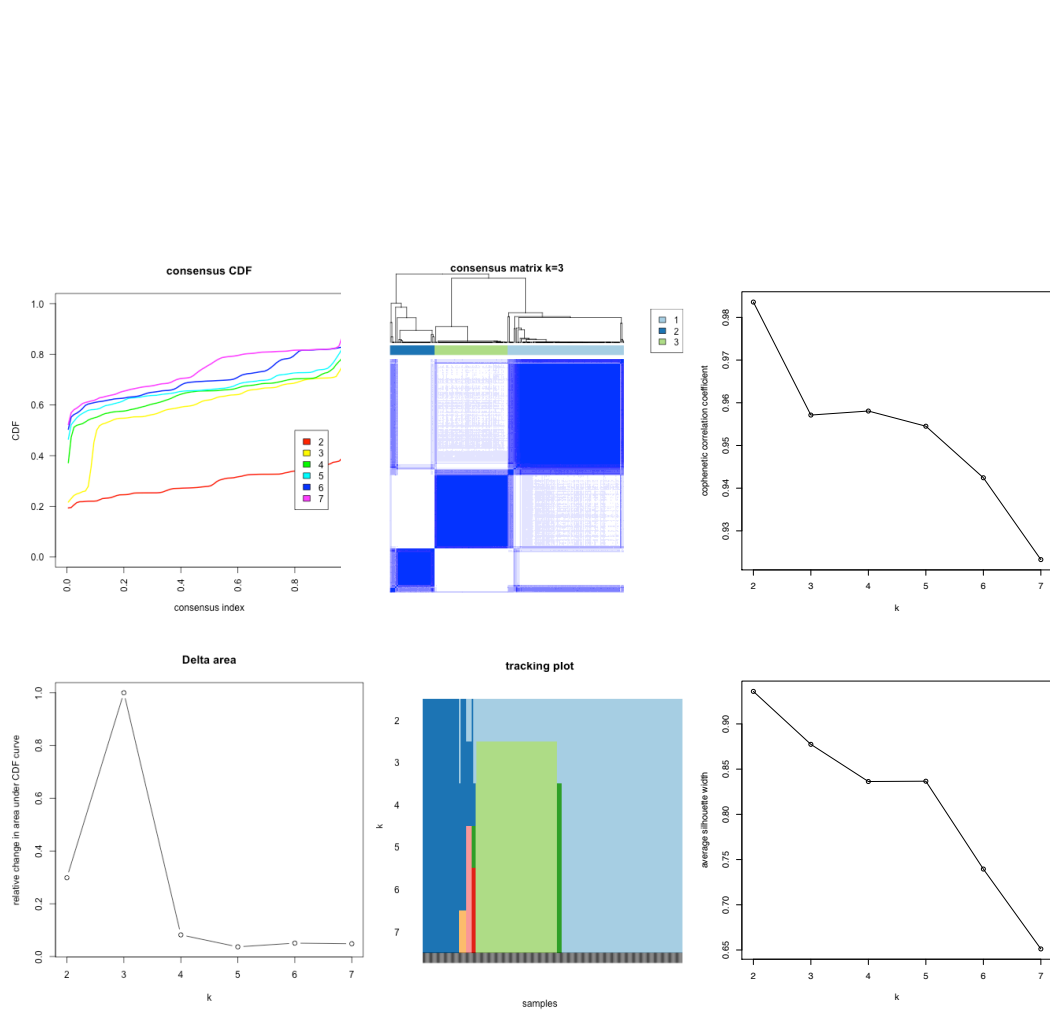


Figure A.3: **Metrics for cluster of clusters model selection.** From top-left: consensus cumulative distribution function (CDF), consensus matrix for  $k=3$ , cophenetic correlation coefficient, relative change in area under the CDF curve, sample tracking plot, average silhouette width.

## Appendix B

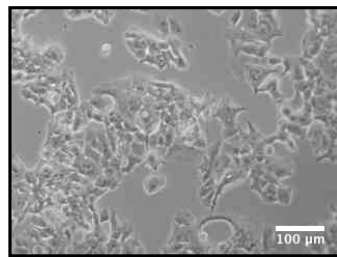
### Chapter 4 Supplementary Figures



A

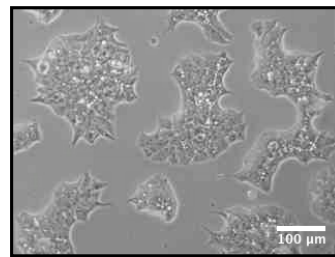
	Derived From	HER2 Status	ER Status	PR Status	P53 Status	Lapatinib GI50
HCC1419	Invasive ductal carcinoma	+	+	+	mutation	0.27 $\mu$ M
BT474	Invasive ductal carcinoma	+	-	-	mutation	0.40 $\mu$ M

B



HCC1419

C



BT474

Figure B.1: **HER2+** lapatinib sensitive breast cancer cell lines used in this study. Characteristics (A) and micrographs of HCC1419 (B) and BT474 (C) cell lines.

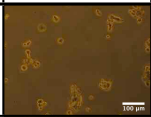
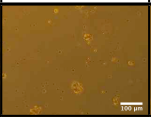
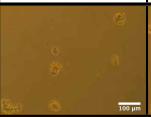
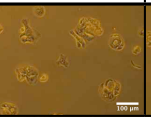
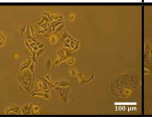
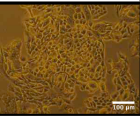
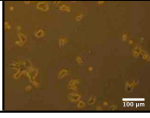
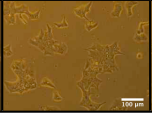
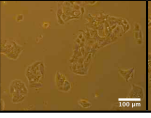
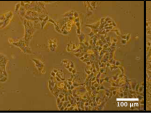
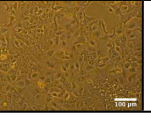

Cell Line	Day 1	Day 13	Day 28	Day 40	Day 70	Day 79
HCC1419						
BT474						

Figure B.2: **Drug tolerant expanded persisters begin to grow in 40-70 days.** Micrographs of HCC1419 and BT474 cells after treatment with 1  $\mu$ M lapatinib at various time points.

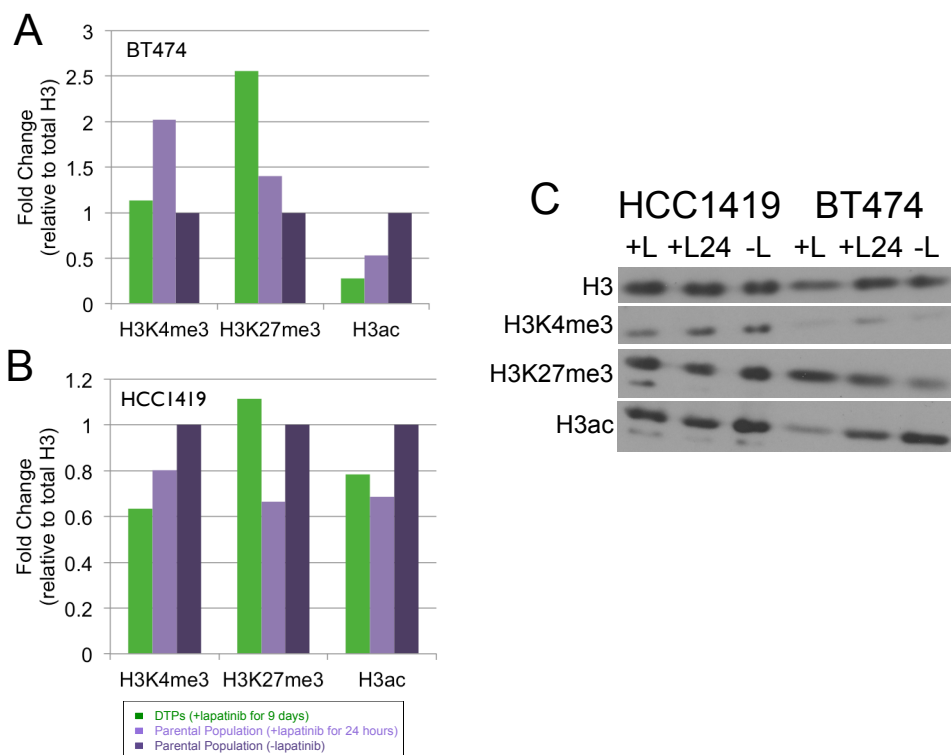


Figure B.3: **DTPs have a distinct chromatin state.** A,B: Quantification of immunoblots shown in (C) using ImageJ [123] for H3K4 tri-methylation (an open chromatin mark), H3K27 tri-methylation (a closed chromatin mark), and H3 total acetylation (an open chromatin mark), normalized to total H3, for DTPs (green), parental population treated for 24 hours (light purple) and untreated parental population (dark purple) in BT474 cells (A) and HCC1419 cells (B) reveals a slightly more closed chromatin state in resistant cells. C: Immunoblots probed with antibodies against H3, H3K4 tri-methylation, H3K27 tri-methylation or H3 total acetylation in DTPs (+L), parental population treated for 24 hours (+L24) and untreated parental population (-L) in HCC1419 and BT474 cells.

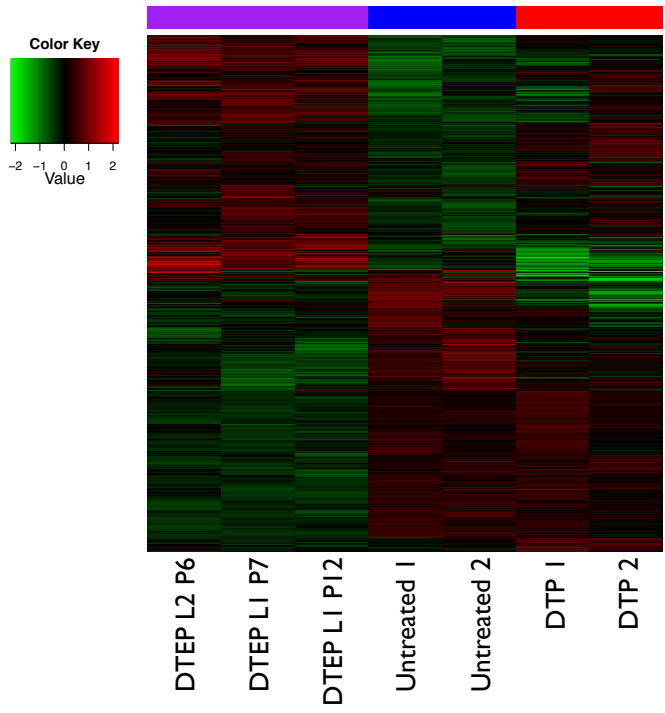


Figure B.4: **DE-Seq Finds 6943 Significantly Differentially Expressed Genes.** Heatmap displaying consensus clustering [153] of all 6,943 genes with a DE-Seq [3] adjusted p-value below 0.1 for any of the 3 possible comparisons (untreated vs. DTP, untreated vs. DTEP, or DTP vs. DTEP). Samples clustered into 3 main clusters by sample type. DTEP L1P7: drug tolerant expanded persisters line #1 passage #7; DTEP L1P12: drug tolerant expanded persisters line #1 passage #12; DTEP L2P6: drug tolerant expanded persisters line #2 passage #6; DTP: drug tolerant persisters.

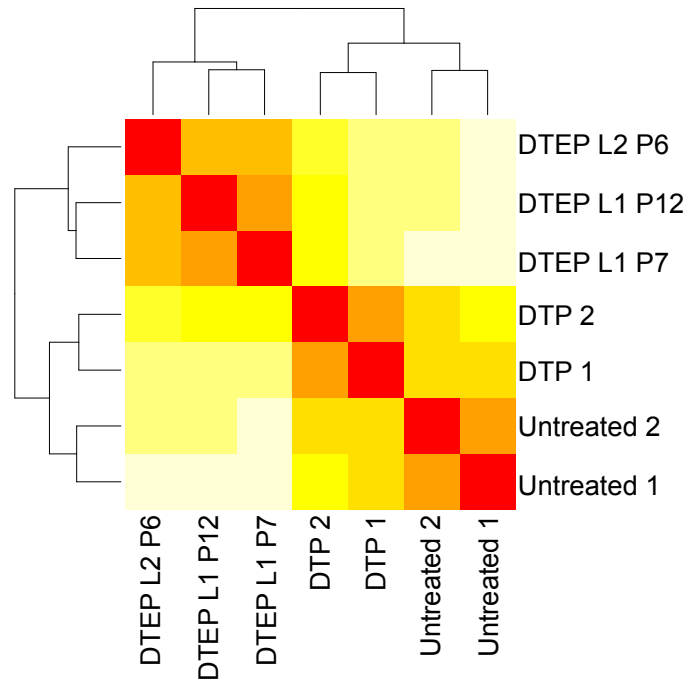


Figure B.5: **DE-Seq reveals that untreated cells, DTPs and DTEPs have significantly different expression patterns, with DTEPs being the most distinct.** Heatmap displaying Euclidian distance between samples, calculated using the variance stabilizing transformation of the DE-Seq [3] count data, red is 0 distance. DTEP L1P7: drug tolerant expanded persisters line #1 passage #7; DTEP L1P12: drug tolerant expanded persisters line #1 passage #12; DTEP L2P6: drug tolerant expanded persisters line #2 passage #6; DTP: drug tolerant persisters.

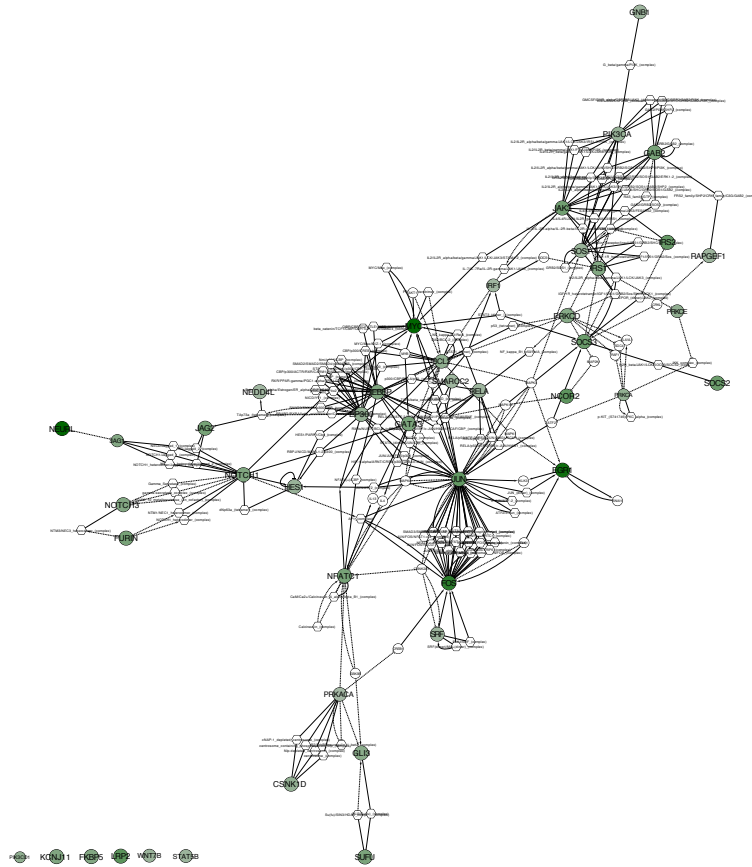


Figure B.6: **Pathways significantly over-expressed in DTEPs are highly interconnected.** Snapshot of Cytoscape [29, 119, 125, 132] visualization of the pathway connections between all “core” genes in pathways GSEA [138] found to be significantly (FDR q-value < 0.1) over-expressed in DTEPs compared to untreated cells. “Core” genes are those that contribute to the GSEA enrichment score, and are therefore over-expressed in DTEPs. Gene nodes were colored based on the DESeq [3] assigned log fold changes and sized based on the DESeq assigned p-value. All colored genes have a DE-Seq Benjamini-Hochberg adjusted p-value less than 0.1. Connections between genes were determined using a superimposed pathway, which incorporates pathway information from the National Cancer Institute Pathway Interaction Database, Biocarta, Reactome and KEGG [60, 122, 91, 32]. Green nodes are over-expressed in DTEPs and purple nodes are over-expressed in untreated cells. Solid lines represent transcriptional regulation, dashed lines represent post-transcriptional regulation, circular nodes represent proteins, triangles represent families, rounded rectangles represent processes and hexagons represent complexes. Cytoscape session available in supplementary data C.6.

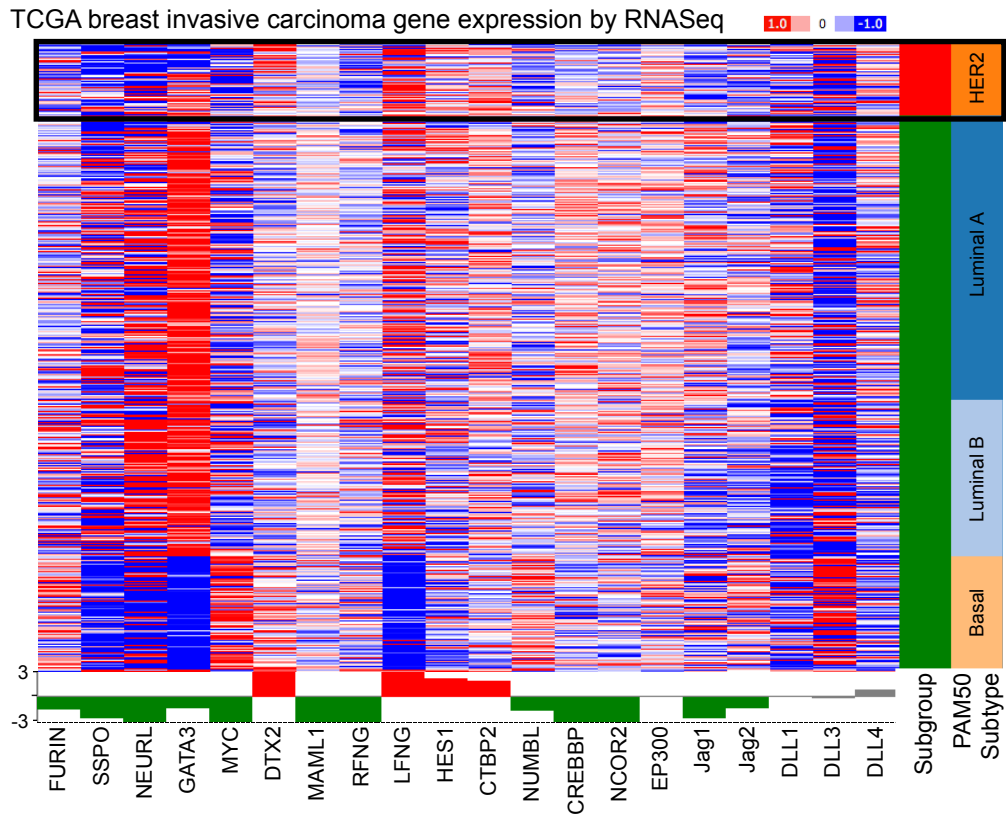


Figure B.7: **Jag1 and Jag2 are under-expressed in HER2+ breast cancer patients.** UCSC Cancer Browser view of RNA-Seq gene expression from TCGA breast cancer patients for all 5 Notch ligands and other Notch pathway genes sorted by PAM50 assigned subtype [90, 45, 162]. Samples are grouped into 2 subgroups, HER2 subtype (red) and Luminal A, Luminal B and Basal subtypes (green), and compared using a t-test. Genes significantly under-expressed in the HER2+ patients are highlighted in green. Similar results are obtained if Bonferroni correction is used.

<b>Up in DTPs and DTEPs vs. untreated</b>	<b>Up in DTEPs vs. untreated</b>	<b>Up in DTEPs vs. untreated and DTPs</b>
IGF-1R IGFBP5 KDM4B	IGF1 IGF-1R IGF-2R IGFBP2 IGFBP4 IGFBP5 IGF2BP3 KDM4B KDM6B	IGF1 IGF-2R IGFBP2 IGF2BP3 KDM4B KDM6B

Table B.1: **DE-Seq results confirm that IGF-1R and KDM pathways are over-expressed in resistant cells.** List of genes in IGF pathway and KDM protein family that are over-expressed in DTPs and DTEPs compared to untreated, over-expressed in DTEPs only compared to untreated or over-expressed in DTEPs compared to both untreated and DTPs. All genes listed have a DE-Seq [3] assigned Benjamini-Hochberg adjusted p-value below 0.1 and a fold change greater than 1.5.



ABCB1 or MDR1	BRAF	IGFBP3	NRAS
ABCC1 or MRP1	COT	IGF...	NRG1
ABCG2 or BCRP	EGFR	IL-6	other integrins (ITG...)
ABL1	ERBB2	KIT	PDGFR
AKT	ERBB3	RAF...	PDGFRA
ALK	ERBB4	KRAS	PDGFRB
AXL	ERCC1	RAS...	PI3K...
BAD	ERK	MAPK...	PIK3CA
BAK	ESR1	MCL1	PTEN
BAX	FGF	MED12	ROS
BCL-2	FGFR	MEK1	SMAC
BCL-W	FLIP STAT2	MET	TGF $\beta$ R
BCL-X	HGF	MLH1	TIMP1
BIM	IGF1	MSH2	XIAP
BIRC2	IGF1R	MTOR	$\beta$ 1-Integrin
BIRC3	IGF2	NEK2	

Table B.2: **Genes known to contribute to drug resistance.** List of all genes found to be involved in drug resistance [52, 12, 47].

# Appendix C

## List of supplementary data files

1. Chapter 2 Supplementary Table S1: BamBam output of peaks and rearrangements  
(TCGA-GBM Exome analysis)
2. Chapter 2 Supplementary Table S3: Data-based validation of exome analysis  
(Whole Genome Sequencing analysis)
3. Chapter 3 Supplementary Table S5: DNA rearrangements and potential double  
minute chromosomes/breakpoint-enriched regions
4. Chapter 4 Supplementary Table S2: Expressed mutations in DTEPs called by  
RADIA
5. Chapter 4 Supplementary Table S4: DE-Seq results: list of differentially expressed

genes

6. Chapter 4 supplementary data: Pathways significantly over-expressed in DTEPs  
(Cytoscape file)

# Bibliography

- [1] K Aldape, M L Simmons, R L Davis, R Miike, J Wiencke, G Barger, M Lee, P Chen, and M Wrensch. Discrepancies in diagnoses of neuroepithelial neoplasms: the san francisco bay area adult glioma study. *Cancer*, 88(10):2342–9, May 2000.
- [2] Kenneth Aldape, Peter C Burger, and Arie Perry. Clinicopathologic aspects of 1p/19q loss and the diagnosis of oligodendroglioma. *Arch Pathol Lab Med*, 131(2):242–51, Feb 2007.
- [3] S. Anders and W. Huber. Differential expression analysis for sequence count data. *Genome Biol*, 11(10):R106, 2010.
- [4] Christina L Appin and Daniel J Brat. Molecular genetics of gliomas. *Cancer J*, 20(1):66–72, 2014.
- [5] A. C. Baeye, G. L. Disbrow, and R. Schlegel. Igfbp-3, a marker of cellular senescence, is overexpressed in human papillomavirus-immortalized cervical cells and enhances igf-1-induced mitogenesis. *J Virol*, 78(11):5720–7, 2004.
- [6] G Balaban-Malenbaum and F Gilbert. Double minute chromosomes and the homogeneously staining regions in chromosomes of a human neuroblastoma cell line. *Science*, 198(4318):739–41, Nov 1977.
- [7] Leonora Balaj, Ryan Lessard, Lixin Dai, Yoon-Jae Cho, Scott L Pomeroy, Xandra O Breakefield, and Johan Skog. Tumour microvesicles contain retrotransposon elements and amplified oncogene sequences. *Nat Commun*, 2:180, 2011.
- [8] P E Barker. Double minutes in human tumor cells. *Cancer Genet Cytogenet*, 5(1):81–94, Feb 1982.
- [9] Adam J Bass, Michael S Lawrence, Lear E Brace, Alex H Ramos, Yotam Drier, Kristian Cibulskis, Carrie Sougnez, Douglas Voet, Gordon Saksena, Andrey

- Sivachenko, Rui Jing, Melissa Parkin, Trevor Pugh, Roel G Verhaak, Nicolas Stransky, Adam T Boutin, Jordi Barretina, David B Solit, Evi Vakiani, Wenlin Shao, Yuji Mishina, Markus Warmuth, Jose Jimenez, Derek Y Chiang, Sabina Signoretti, William G Kaelin, Nicole Spardy, William C Hahn, Yujin Hoshida, Shuji Ogino, Ronald A Depinho, Lynda Chin, Levi A Garraway, Charles S Fuchs, Jose Baselga, Josep Taberero, Stacey Gabriel, Eric S Lander, Gad Getz, and Matthew Meyerson. Genomic sequencing of colorectal adenocarcinomas identifies a recurrent *vtila-tcf712* fusion. *Nat Genet*, 43(10):964–8, Oct 2011.
- [10] Chetan Bettegowda, Nishant Agrawal, Yuchen Jiao, Mark Sausen, Laura D Wood, Ralph H Hruban, Fausto J Rodriguez, Daniel P Cahill, Roger McLendon, Gregory Riggins, Victor E Velculescu, Sueli Mieko Oba-Shinjo, Suely Kazue Nagahashi Marie, Bert Vogelstein, Darell Bigner, Hai Yan, Nickolas Papadopoulos, and Kenneth W Kinzler. Mutations in *cic* and *fubp1* contribute to human oligodendroglioma. *Science*, 333(6048):1453–5, Sep 2011.
- [11] S H Bigner, J Mark, and D D Bigner. Cytogenetics of human brain tumors. *Cancer Genet Cytogenet*, 47(2):141–54, Jul 1990.
- [12] B. G. Blair, A. Bardelli, and B. H. Park. Somatic alterations as the basis for resistance to targeted therapies. *J Pathol*, 232(2):244–54, 2014.
- [13] L. P. Blair, J. Cao, M. R. Zou, J. Sayegh, and Q. Yan. Epigenetic regulation by lysine demethylase 5 (*kdm5*) enzymes in cancer. *Cancers (Basel)*, 3(1):1383–404, 2011.
- [14] T David Bourne and David Schiff. Update on molecular findings, management and outcome in low-grade gliomas. *Nat Rev Neurol*, 6(12):695–701, Dec 2010.
- [15] Cameron W Brennan, Roel G W Verhaak, Aaron McKenna, Benito Campos, Houtan Noushmehr, Sofie R Salama, Siyuan Zheng, Debyani Chakravarty, J Zachary Sanborn, Samuel H Berman, Rameen Beroukhi, Brady Bernard, Chang-Jiun Wu, Giannicola Genovese, Ilya Shmulevich, Jill Barnholtz-Sloan, Lihua Zou, Rahulsimham Vegesna, Sachet A Shukla, Giovanni Ciriello, W K Yung, Wei Zhang, Carrie Sougnez, Tom Mikkelsen, Kenneth Aldape, Darell D Bigner, Erwin G Van Meir, Michael Prados, Andrew Sloan, Keith L Black, Jennifer Eschbacher, Gaetano Finocchiaro, William Friedman, David W Andrews, Abhijit Guha, Mary Iacocca, Brian P O’Neill, Greg Foltz, Jerome Myers, Daniel J Weisenberger, Robert Penny, Raju Kucherlapati, Charles M Perou, D Neil Hayes, Richard Gibbs, Marco Marra, Gordon B Mills, Eric Lander, Paul Spellman, Richard Wilson, Chris Sander, John Weinstein, Matthew Meyerson, Stacey Gabriel, Peter W Laird, David Haussler, Gad Getz, Lynda Chin, and

- TCGA Research Network. The somatic genomic landscape of glioblastoma. *Cell*, 155(2):462–77, Oct 2013.
- [16] R Brookwell and F A Hunt. Formation of double minutes by breakdown of a homogeneously staining region in a refractory anemia with excess blasts. *Cancer Genet Cytogenet*, 34(1):47–52, Aug 1988.
- [17] Gregory Cairncross, Meihua Wang, Edward Shaw, Robert Jenkins, David Brachman, Jan Buckner, Karen Fink, Luis Souhami, Normand Laperriere, Walter Curran, and Minesh Mehta. Phase iii trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of rtog 9402. *J Clin Oncol*, 31(3):337–43, Jan 2013.
- [18] J G Cairncross and D R Macdonald. Oligodendroglioma: a new chemosensitive tumor. *J Clin Oncol*, 8(12):2090–1, Dec 1990.
- [19] J G Cairncross, K Ueki, M C Zlatescu, D K Lisle, D M Finkelstein, R R Hammond, J S Silver, P C Stark, D R Macdonald, Y Ino, D A Ramsay, and D N Louis. Specific genetic predictors of chemotherapeutic response and survival in patients with anaplastic oligodendrogliomas. *J Natl Cancer Inst*, 90(19):1473–9, Oct 1998.
- [20] J Gregory Cairncross, Meihua Wang, Robert B Jenkins, Edward G Shaw, Caterina Giannini, David G Brachman, Jan C Buckner, Karen L Fink, Luis Souhami, Normand J Laperriere, Jason T Huse, Minesh P Mehta, and Walter J Curran, Jr. Benefit from procarbazine, lomustine, and vincristine in oligodendroglial tumors is associated with mutation of idh. *J Clin Oncol*, 32(8):783–90, Mar 2014.
- [21] Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, 455(7216):1061–8, Oct 2008.
- [22] Cancer Genome Atlas Research Network, John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nat Genet*, 45(10):1113–20, Oct 2013.
- [23] S. Cara and I. F. Tannock. Retreatment of patients with the same chemotherapy: implications for clinical mechanisms of drug resistance. *Ann Oncol*, 12(1):23–7, 2001.
- [24] Ken Chen, John W Wallis, Michael D McLellan, David E Larson, Joelle M Kalicki, Craig S Pohl, Sean D McGrath, Michael C Wendl, Qunyuan Zhang, Devin P Locke, Xiaoqi Shi, Robert S Fulton, Timothy J Ley, Richard K Wilson, Li Ding,

- and Elaine R Mardis. Breakdancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods*, 6(9):677–81, Sep 2009.
- [25] E. Chinchar, K. L. Makey, J. Gibson, F. Chen, S. A. Cole, G. C. Megason, S. Vijayakumar, L. Miele, and J. W. Gu. Sunitinib significantly suppresses the proliferation, migration, apoptosis resistance, tumor angiogenesis and growth of triple-negative breast cancers but increases breast cancer stem cells. *Vasc Cell*, 6:12, 2014.
- [26] Jeonghee Cho, Sandra Pastorino, Qing Zeng, Xiaoyin Xu, William Johnson, Scott Vandenberg, Roel Verhaak, Andrew D Cherniack, Hideo Watanabe, Amit Dutt, Jihyun Kwon, Ying S Chao, Robert C Onofrio, Derek Chiang, Yuki Yuza, Santosh Kesari, and Matthew Meyerson. Glioblastoma-derived epidermal growth factor receptor carboxyl-terminal deletion mutants are transforming and are sensitive to egfr-directed therapies. *Cancer Res*, 71(24):7587–96, Dec 2011.
- [27] Giovanni Ciriello, Martin L Miller, Bülent Arman Aksoy, Yasin Senbabaoglu, Nikolaus Schultz, and Chris Sander. Emerging landscape of oncogenic signatures across human cancers. *Nat Genet*, 45(10):1127–1133, Sep 2013.
- [28] Ami Citri, Kochupurakkal Bose Skaria, and Yosef Yarden. The deaf and the dumb: the biology of erbb-2 and erbb-3. *Exp Cell Res*, 284(1):54–65, Mar 2003.
- [29] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P. L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader. Integration of biological networks and gene expression data using cytoscape. *Nat Protoc*, 2(10):2366–82, 2007.
- [30] B. Cohen, M. Shimizu, J. Izrailit, N. F. Ng, Y. Buchman, J. G. Pan, J. Dering, and M. Reedijk. Cyclin d1 is a direct target of jag1-mediated notch signaling in breast cancer. *Breast Cancer Res Treat*, 123(1):113–24, 2010.
- [31] S W Coons, P C Johnson, B W Scheithauer, A J Yates, and D K Pearl. Improving diagnostic accuracy and interobserver concordance in the classification and grading of primary gliomas. *Cancer*, 79(7):1381–93, Apr 1997.
- [32] D. Croft, A. F. Mundo, R. Haw, M. Milacic, J. Weiser, G. Wu, M. Caudy, P. Garapati, M. Gillespie, M. R. Kamdar, B. Jassal, S. Jupe, L. Matthews, B. May, S. Palatnik, K. Rothfels, V. Shamovsky, H. Song, M. Williams, E. Birney, H. Her-

- mjakob, L. Stein, and P. D'Eustachio. The reactome pathway knowledgebase. *Nucleic Acids Res*, 42(Database issue):D472–7, 2014.
- [33] F. De Santa, V. Narang, Z. H. Yap, B. K. Tusi, T. Burgold, L. Austenaa, G. Bucci, M. Caganova, S. Notarbartolo, S. Casola, G. Testa, W. K. Sung, C. L. Wei, and G. Natoli. Jmjd3 contributes to the control of gene expression in lps-activated macrophages. *EMBO J*, 28(21):3341–52, 2009.
- [34] F. De Santa, M. G. Totaro, E. Prosperini, S. Notarbartolo, G. Testa, and G. Natoli. The histone h3 lysine-27 demethylase jmjd3 links inflammation to inhibition of polycomb-mediated gene silencing. *Cell*, 130(6):1083–94, 2007.
- [35] Kathleen Deiteren, Dirk Hendriks, Simon Scharpé, and Anne Marie Lambeir. Carboxypeptidase m: Multiple alliances and unknown partners. *Clin Chim Acta*, 399(1-2):24–39, Jan 2009.
- [36] David M Duda, Daniel C Scott, Matthew F Calabrese, Erik S Zimmerman, Ning Zheng, and Brenda A Schulman. Structural regulation of cullin-ring ubiquitin ligase complexes. *Curr Opin Struct Biol*, 21(2):257–64, Apr 2011.
- [37] R. Edgar, M. Domrachev, and A. E. Lash. Gene expression omnibus: Ncbi gene expression and hybridization array data repository. *Nucleic Acids Res*, 30(1):207–10, 2002.
- [38] S. Elias, S. Liang, Y. Chen, M. A. De Marco, O. Machek, S. Skucha, L. Miele, and M. Bocchetta. Notch-1 stimulates survival of lung adenocarcinoma cells during hypoxia by activating the igf-1r pathway. *Oncogene*, 29(17):2488–98, 2010.
- [39] M. Ernst and G. A. Rodan. Increased activity of insulin-like growth factor (igf) in osteoblastic cells in the presence of growth hormone (gh): positive correlation with the presence of the gh-induced igf-binding protein bp-3. *Endocrinology*, 127(2):807–14, 1990.
- [40] Centers for Disease Control and Prevention. Leading causes of death, 2010.
- [41] S. A. Forbes, G. Bhamra, S. Bamford, E. Dawson, C. Kok, J. Clements, A. Menzies, J. W. Teague, P. A. Futreal, and M. R. Stratton. The catalogue of somatic mutations in cancer (cosmic). *Curr Protoc Hum Genet*, Chapter 10:Unit 10.11, 2008.
- [42] Kinisha Gala and Sarat Chandarlapaty. Molecular pathways: Her3 targeted therapy. *Clin Cancer Res*, 20(6):1410–6, Mar 2014.



- [43] Charles E Geyer, John Forster, Deborah Lindquist, Stephen Chan, C Gilles Romieu, Tadeusz Pienkowski, Agnieszka Jagiello-Gruszczyk, John Crown, Arlene Chan, Bella Kaufman, Dimosthenis Skarlos, Mario Campone, Neville Davidson, Mark Berger, Cristina Oliva, Stephen D Rubin, Steven Stein, and David Cameron. Lapatinib plus capecitabine for her2-positive advanced breast cancer. *N Engl J Med*, 355(26):2733–2743, Dec 2006.
- [44] Caterina Giannini, Peter C Burger, Brian A Berkey, J Gregory Cairncross, Robert B Jenkins, Minesh Mehta, Walter J Curran, and Ken Aldape. Anaplastic oligodendroglial tumors: refining the correlation among histopathology, 1p 19q deletion and clinical outcome in intergroup radiation therapy oncology group trial 9402. *Brain Pathol*, 18(3):360–9, Jul 2008.
- [45] M. Goldman, B. Craft, T. Swatloski, K. Ellrott, M. Cline, M. Diekhans, S. Ma, C. Wilks, J. Stuart, D. Haussler, and J. Zhu. The ucsc cancer genomics browser: update 2013. *Nucleic Acids Res*, 41(Database issue):D949–54, 2013.
- [46] Thierry Gorlia, Jean-Yves Delattre, Alba A Brandes, Johan M Kros, Martin J B Taphoorn, Mathilde C M Kouwenhoven, H J J A Bernsen, Marc Frénay, Cees C Tijssen, Denis Lacombe, and Martin J van den Bent. New clinical, pathological and molecular prognostic models and calculators in patients with locally diagnosed anaplastic oligodendroglioma or oligoastrocytoma. a prognostic factor analysis of european organisation for research and treatment of cancer brain tumour group study 26951. *Eur J Cancer*, 49(16):3477–85, Nov 2013.
- [47] M. M. Gottesman. Mechanisms of cancer drug resistance. *Annu Rev Med*, 53:615–27, 2002.
- [48] D Graus-Porta, R R Beerli, J M Daly, and N E Hynes. ErbB-2, the preferred heterodimerization partner of all erbB receptors, is a mediator of lateral signaling. *EMBO J*, 16(7):1647–1655, Apr 1997.
- [49] Chris D Greenman, Erin D Pleasance, Scott Newman, Fengtang Yang, Beiyuan Fu, Serena Nik-Zainal, David Jones, King Wai Lau, Nigel Carter, Paul A W Edwards, P Andrew Futreal, Michael R Stratton, and Peter J Campbell. Estimation of rearrangement phylogeny for cancer genomes. *Genome Res*, 22(2):346–61, Feb 2012.
- [50] Christian Hartmann, Bettina Hentschel, Wolfgang Wick, David Capper, Jörg Felsberg, Matthias Simon, Manfred Westphal, Gabriele Schackert, Richard Meyer-mann, Torsten Pietsch, Guido Reifenberger, Michael Weller, Markus Loeffler, and Andreas von Deimling. Patients with idh1 wild type anaplastic astrocytomas exhibit worse prognosis than idh1-mutated glioblastomas, and idh1 mutation sta-

tus accounts for the unfavorable prognostic effect of higher age: implications for classification of gliomas. *Acta Neuropathol*, 120(6):707–18, Dec 2010.

- [51] F. E. Henken, J. De-Castro Arce, F. Rösl, L. Bosch, C. J. Meijer, P. J. Snijders, and R. D. Steenbergen. The functional role of notch signaling in hpv-mediated transformation is dose-dependent and linked to ap-1 alterations. *Cell Oncol (Dordr)*, 35(2):77–84, 2012.
- [52] C. Holohan, S. Van Schaeybroeck, D. B. Longley, and P. G. Johnston. Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer*, 13(10):714–26, 2013.
- [53] F. Hsu, W. J. Kent, H. Clawson, R. M. Kuhn, M. Diekhans, and D. Haussler. The ucsc known genes. *Bioinformatics*, 22(9):1036–46, 2006.
- [54] da W Huang, B. T. Sherman, and R. A. Lempicki. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*, 37(1):1–13, 2009.
- [55] da W Huang, B. T. Sherman, and R. A. Lempicki. Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nat Protoc*, 4(1):44–57, 2009.
- [56] N. E. Hynes and H. A. Lane. Erbb receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer*, 5(5):341–54, 2005.
- [57] Yuchen Jiao, Patrick J Killela, Zachary J Reitman, Ahmed B Rasheed, Christopher M Heaphy, Roeland F de Wilde, Fausto J Rodriguez, Sergio Rosemberg, Sueli Mieko Oba-Shinjo, Suelly Kazue Nagahashi Marie, Chetan Bettegowda, Nishant Agrawal, Eric Lipp, Christopher Pirozzi, Giselle Lopez, Yiping He, Henry Friedman, Allan H Friedman, Gregory J Riggins, Matthias Holdhoff, Peter Burger, Roger McLendon, Darell D Bigner, Bert Vogelstein, Alan K Meeker, Kenneth W Kinzler, Nickolas Papadopoulos, Luis A Diaz, and Hai Yan. Frequent atrx, cic, fubp1 and idh1 mutations refine the classification of malignant gliomas. *Oncotarget*, 3(7):709–22, Jul 2012.
- [58] Brett E Johnson, Tali Mazor, Chibo Hong, Michael Barnes, Koki Aihara, Cory Y McLean, Shaun D Fouse, Shogo Yamamoto, Hiroki Ueda, Kenji Tatsuno, Saurabh Asthana, Llewellyn E Jalbert, Sarah J Nelson, Andrew W Bollen, W Clay Gustafson, Elise Charron, William A Weiss, Ivan V Smirnov, Jun S Song, Adam B Olshen, Soonmee Cha, Yongjun Zhao, Richard A Moore, Andrew J Mungall, Steven J M Jones, Martin Hirst, Marco A Marra, Nobuhito Saito, Hiroyuki Aburatani, Akitake Mukasa, Mitchel S Berger, Susan M Chang, Barry S Taylor, and

- Joseph F Costello. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*, 343(6167):189–93, Jan 2014.
- [59] F E Jones and D F Stern. Expression of dominant-negative *erbb2* in the mammary gland of transgenic mice reveals a role in lobuloalveolar development and lactation. *Oncogene*, 18(23):3481–3490, Jun 1999.
- [60] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, 2000.
- [61] Kasthuri Kannan, Akiko Inagaki, Joachim Silber, Daniel Gorovets, Jianan Zhang, Edward R Kastenhuber, Adriana Heguy, John H Petrini, Timothy A Chan, and Jason T Huse. Whole-exome sequencing identifies *atrx* mutation as a key molecular determinant in lower-grade glioma. *Oncotarget*, 3(10):1194–203, Oct 2012.
- [62] H. Kennecke, R. Yerushalmi, R. Woods, M. C. Cheang, D. Voduc, C. H. Speers, T. O. Nielsen, and K. Gelmon. Metastatic behavior of breast cancer subtypes. *J Clin Oncol*, 28(20):3271–7, 2010.
- [63] W. J. Kent. Blat—the blast-like alignment tool. *Genome Res*, 12(4):656–64, 2002.
- [64] Patrick J Killela, Christopher J Pirozzi, Zachary J Reitman, Sian Jones, B Ahmed Rasheed, Eric Lipp, Henry Friedman, Allan H Friedman, Yiping He, Roger E McLendon, Darell D Bigner, and Hai Yan. The genetic landscape of anaplastic astrocytoma. *Oncotarget*, 5(6):1452–7, Mar 2014.
- [65] Patrick J Killela, Zachary J Reitman, Yuchen Jiao, Chetan Bettegowda, Nishant Agrawal, Luis A Diaz, Jr, Allan H Friedman, Henry Friedman, Gary L Gallia, Beppino C Giovanella, Arthur P Grollman, Tong-Chuan He, Yiping He, Ralph H Hruban, George I Jallo, Nils Mandahl, Alan K Meeker, Fredrik Mertens, George J Netto, B Ahmed Rasheed, Gregory J Riggins, Thomas A Rosenquist, Mark Schiffman, Ie-Ming Shih, Dan Theodorescu, Michael S Torbenson, Victor E Velculescu, Tian-Li Wang, Nicolas Wentzensen, Laura D Wood, Ming Zhang, Roger E McLendon, Darell D Bigner, Kenneth W Kinzler, Bert Vogelstein, Nickolas Papadopoulos, and Hai Yan. Tert promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci U S A*, 110(15):6021–6, Apr 2013.
- [66] T. Kurata, K. Tamura, H. Kaneda, T. Nogami, H. Uejima, G. Asai Go, K. Nakagawa, and M. Fukuoka. Effect of re-treatment with gefitinib (‘iressa’, *zd1839*) after acquisition of resistance. *Ann Oncol*, 15(1):173–4, 2004.
- [67] Fabien Kuttler and Sabine Mai. Formation of non-random extrachromosomal

- elements during development, differentiation and oncogenesis. *Semin Cancer Biol*, 17(1):56–64, Feb 2007.
- [68] B. Langmead and S. L. Salzberg. Fast gapped-read alignment with bowtie 2. *Nat Methods*, 9(4):357–9, 2012.
- [69] Iris Lavon, Miri Refael, Bracha Zelikovitch, Edna Shalom, and Tali Siegal. Serum dna can define tumor-specific genetic and epigenetic markers in gliomas of various grades. *Neuro Oncol*, 12(2):173–80, Feb 2010.
- [70] Michael S Lawrence, Petar Stojanov, Paz Polak, Gregory V Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L Carter, Chip Stewart, Craig H Mermel, Steven A Roberts, Adam Kiezun, Peter S Hammerman, Aaron McKenna, Yotam Drier, Lihua Zou, Alex H Ramos, Trevor J Pugh, Nicolas Stransky, Elena Helman, Jaegil Kim, Carrie Sougnez, Lauren Ambrogio, Elizabeth Nickerson, Erica Shefler, Maria L Cortés, Daniel Auclair, Gordon Saksena, Douglas Voet, Michael Noble, Daniel DiCara, Pei Lin, Lee Lichtenstein, David I Heiman, Timothy Fennell, Marcin Imielinski, Bryan Hernandez, Eran Hodis, Sylvan Baca, Austin M Dulak, Jens Lohr, Dan-Avi Landau, Catherine J Wu, Jorge Melendez-Zajgla, Alfredo Hidalgo-Miranda, Amnon Koren, Steven A McCarroll, Jaume Mora, Ryan S Lee, Brian Crompton, Robert Onofrio, Melissa Parkin, Wendy Winckler, Kristin Ardlie, Stacey B Gabriel, Charles W M Roberts, Jaclyn A Biegel, Kimberly Stegmaier, Adam J Bass, Levi A Garraway, Matthew Meyerson, Todd R Golub, Dmitry A Gordenin, Shamil Sunyaev, Eric S Lander, and Gad Getz. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, 499(7457):214–8, Jul 2013.
- [71] Rebecca J Leary, Mark Sausen, Isaac Kinde, Nickolas Papadopoulos, John D Carpten, David Craig, Joyce O’Shaughnessy, Kenneth W Kinzler, Giovanni Parmigiani, Bert Vogelstein, Luis A Diaz, Jr, and Victor E Velculescu. Detection of chromosomal alterations in the circulation of cancer patients with whole-genome sequencing. *Sci Transl Med*, 4(162):162ra154, Nov 2012.
- [72] D. A. Lewis and D. F. Spandau. Uvb-induced activation of nf-kappab is regulated by the igf-1r and dependent on p38 mapk. *J Invest Dermatol*, 128(4):1022–9, 2008.
- [73] B. Li and C. N. Dewey. Rsem: accurate transcript quantification from rna-seq data with or without a reference genome. *BMC Bioinformatics*, 12:323, 2011.
- [74] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup.

- The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–9, 2009.
- [75] Heng Li and Richard Durbin. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 25(14):1754–60, Jul 2009.
- [76] Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, and 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and samtools. *Bioinformatics*, 25(16):2078–9, Aug 2009.
- [77] Y. Lombardo, M. Faronato, A. Filipovic, V. Vircillo, L. Magnani, and R. C. Coombes. Nicastrin and notch4 drive endocrine therapy resistance and epithelial-mesenchymal transition in mcf7 breast cancer cells. *Breast Cancer Res*, 16(3):R62, 2014.
- [78] David N Louis, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvett, Bernd W Scheithauer, and Paul Kleihues. The 2007 who classification of tumours of the central nervous system. *Acta Neuropathol*, 114(2):97–109, Aug 2007.
- [79] David N Louis, Arie Perry, Peter Burger, David W Ellison, Guido Reifenberger, Andreas von Deimling, Kenneth Aldape, Daniel Brat, V Peter Collins, Charles Eberhart, Dominique Figarella-Branger, Gregory N Fuller, Felice Giangaspero, Caterina Giannini, Cynthia Hawkins, Paul Kleihues, Andrey Korshunov, Johan M Kros, M Beatriz Lopes, Ho-Keung Ng, Hiroko Ohgaki, Werner Paulus, Torsten Pietsch, Marc Rosenblum, Elisabeth Rushing, Figen Soylemezoglu, Otmar Wiestler, and Pieter Wesseling. International society of neuropathology-haarlem consensus guidelines for nervous system tumor classification and grading. *Brain Pathol*, 24(5):429–35, Sep 2014.
- [80] Y. Lu, X. Zi, Y. Zhao, D. Mascarenhas, and M. Pollak. Insulin-like growth factor-i receptor signaling and resistance to trastuzumab (herceptin). *J Natl Cancer Inst*, 93(24):1852–7, 2001.
- [81] Gisela Lundberg, Anders H Rosengren, Ulf Håkanson, Henrik Stewénus, Yuesheng Jin, Ylva Stewénus, Sven Pålman, and David Gisselsson. Binomial mitotic segregation of mycn-carrying double minutes in neuroblastoma illustrates the role of randomness in oncogene amplification. *PLoS One*, 3(8):e3099, 2008.
- [82] D R Macdonald, L E Gaspar, and J G Cairncross. Successful chemotherapy for newly diagnosed aggressive oligodendroglioma. *Ann Neurol*, 27(5):573–4, May 1990.

- [83] L. Magnani, A. Stoeck, X. Zhang, A. Lánczky, A. C. Mirabella, T. L. Wang, B. Györffy, and M. Lupien. Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. *Proc Natl Acad Sci U S A*, 110(16):E1490–9, 2013.
- [84] Marcel Martin, Lars Maßhöfer, Petra Temming, Sven Rahmann, Claudia Metz, Norbert Bornfeld, Johannes van de Nes, Ludger Klein-Hitpass, Alan G Hinnebusch, Bernhard Horsthemke, Dietmar R Lohmann, and Michael Zeschnigk. Exome sequencing identifies recurrent somatic mutations in *eif1ax* and *sf3b1* in uveal melanoma with disomy 3. *Nat Genet*, 45(8):933–6, Aug 2013.
- [85] J Matthew McDonald, Siew Ju See, Ivo W Tremont, Howard Colman, Mark R Gilbert, Morris Groves, Peter C Burger, David N Louis, Caterina Giannini, Gregory Fuller, Sandra Passe, Hilary Blair, Robert B Jenkins, Helen Yang, Alicia Ledoux, Joann Aaron, Ulka Tipnis, Wei Zhang, Kenneth Hess, and Ken Aldape. The prognostic impact of histology and 1p/19q status in anaplastic oligodendroglial tumors. *Cancer*, 104(7):1468–77, Oct 2005.
- [86] G B McGwire and R A Skidgel. Extracellular conversion of epidermal growth factor (egf) to des-arg53-egf by carboxypeptidase m. *J Biol Chem*, 270(29):17154–8, Jul 1995.
- [87] Andrew McPherson, Fereydoun Hormozdiari, Abdalnasser Zayed, Ryan Giuliany, Gavin Ha, Mark G F Sun, Malachi Griffith, Alireza Heravi Moussavi, Janine Senz, Nataliya Melnyk, Marina Pacheco, Marco A Marra, Martin Hirst, Torsten O Nielsen, S Cenk Sahinalp, David Huntsman, and Sohrab P Shah. defuse: an algorithm for gene fusion discovery in tumor rna-seq data. *PLoS Comput Biol*, 7(5):e1001138, May 2011.
- [88] Mitch McVey and Sang Eun Lee. Mmej repair of double-strand breaks (director’s cut): deleted sequences and alternative endings. *Trends Genet*, 24(11):529–38, Nov 2008.
- [89] D K Moscatello, M Holgado-Madruga, A K Godwin, G Ramirez, G Gunn, P W Zoltick, J A Biegel, R L Hayes, and A J Wong. Frequent expression of a mutant epidermal growth factor receptor in multiple human tumors. *Cancer Res*, 55(23):5536–9, Dec 1995.
- [90] The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418):61–70, 2012.
- [91] The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature*, 499(7456):43–9, 2013.

- [92] Sam Ng, Eric A Collisson, Artem Sokolov, Theodore Goldstein, Abel Gonzalez-Perez, Nuria Lopez-Bigas, Christopher Benz, David Haussler, and Joshua M Stuart. Paradigm-shift predicts the function of mutations in multiple cancers using pathway impact analysis. *Bioinformatics*, 28(18):i640–i646, Sep 2012.
- [93] G. Nishibuchi, Y. Shibata, T. Hayakawa, N. Hayakawa, Y. Ohtani, K. Shinmyozu, H. Tagami, and J. I. Nakayama. Physical and functional interactions between the histone h3k4 demethylase kdm5a and the nucleosome remodeling and deacetylase (nurd) complex. *J Biol Chem*, 2014.
- [94] Paul A Northcott, David J H Shih, John Peacock, Livia Garzia, A Sorana Morrissy, Thomas Zichner, Adrian M Stütz, Andrey Korshunov, Jüri Reimand, Steven E Schumacher, Rameen Beroukhim, David W Ellison, Christian R Marshall, Anath C Lionel, Stephen Mack, Adrian Dubuc, Yuan Yao, Vijay Ramaswamy, Betty Luu, Adi Rolider, Florence M G Cavalli, Xin Wang, Marc Remke, Xiaochong Wu, Readman Y B Chiu, Andy Chu, Eric Chuah, Richard D Corbett, Gemma R Hoad, Shaun D Jackman, Yisu Li, Allan Lo, Karen L Mungall, Ka Ming Nip, Jenny Q Qian, Anthony G J Raymond, Nina T Thiessen, Richard J Varhol, Inanc Birol, Richard A Moore, Andrew J Mungall, Robert Holt, Daisuke Kawauchi, Martine F Roussel, Marcel Kool, David T W Jones, Hendrick Witt, Africa Fernandez-L, Anna M Kenney, Robert J Wechsler-Reya, Peter Dirks, Tzvi Aviv, Wieslawa A Grajkowska, Marta Perek-Polnik, Christine C Haberler, Olivier Delattre, Stéphanie S Reynaud, François F Doz, Sarah S Pernet-Fattet, Byung-Kyu Cho, Seung-Ki Kim, Kyu-Chang Wang, Wolfram Scheurlen, Charles G Eberhart, Michelle Fèvre-Montange, Anne Jouvét, Ian F Pollack, Xing Fan, Karin M Muraszko, G Yancey Gillespie, Concezio Di Rocco, Luca Massimi, Erna M C Michiels, Nanne K Kloosterhof, Pim J French, Johan M Kros, James M Olson, Richard G Ellenbogen, Karel Zitterbart, Leos Kren, Reid C Thompson, Michael K Cooper, Boleslaw Lach, Roger E McLendon, Darell D Bigner, Adam Fontebasso, Steffen Albrecht, Nada Jabado, Janet C Lindsey, Simon Bailey, Nalin Gupta, William A Weiss, László Bognár, Almos Klekner, Timothy E Van Meter, Toshihiro Kumabe, Teiji Tominaga, Samer K Elbabaa, Jeffrey R Leonard, Joshua B Rubin, Linda M Liau, Erwin G Van Meir, Maryam Fouladi, Hideo Nakamura, Giuseppe Cinalli, Miklós Garami, Peter Hauser, Ali G Saad, Achille Iolascon, Shin Jung, Carlos G Carlotti, Rajeev Vibhakar, Young Shin Ra, Shenandoah Robinson, Massimo Zollo, Claudia C Faria, Jennifer A Chan, Michael L Levy, Poul H B Sorensen, Matthew Meyerson, Scott L Pomeroy, Yoon-Jae Cho, Gary D Bader, Uri Tabori, Cynthia E Hawkins, Eric Bouffet, Stephen W Scherer, James T Rutka, David Malkin, Steven C Clifford, Steven J M Jones, Jan O Korbel, Stefan M Pfister, Marco A Marra, and Michael D Taylor. Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature*, 488(7409):49–56, Aug 2012.
- [95] Houtan Noushmehr, Daniel J Weisenberger, Kristin Diefes, Heidi S Phillips,

- Kanan Pujara, Benjamin P Berman, Fei Pan, Christopher E Pelloski, Erik P Sulman, Krishna P Bhat, Roel G W Verhaak, Katherine A Hoadley, D Neil Hayes, Charles M Perou, Heather K Schmidt, Li Ding, Richard K Wilson, David Van Den Berg, Hui Shen, Henrik Bengtsson, Pierre Neuvial, Leslie M Cope, Jonathan Buckley, James G Herman, Stephen B Baylin, Peter W Laird, Kenneth Aldape, and Cancer Genome Atlas Research Network. Identification of a cpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*, 17(5):510–22, May 2010.
- [96] R. Olsauskas-Kuprys, A. Zlobin, and C. Osipo. Gamma secretase inhibitors of notch signaling. *Onco Targets Ther*, 6:943–55, 2013.
- [97] C. Osipo, P. Patel, P. Rizzo, A. G. Clementz, L. Hao, T. E. Golde, and L. Miele. Erbb-2 inhibition activates notch-1 and sensitizes breast cancer cells to a gamma-secretase inhibitor. *Oncogene*, 27(37):5019–32, 2008.
- [98] Quinn T Ostrom, Haley Gittleman, Paul Farah, Annie Ondracek, Yanwen Chen, Yingli Wolinsky, Nancy E Stroup, Carol Kruchko, and Jill S Barnholtz-Sloan. Cbtrus statistical report: Primary brain and central nervous system tumors diagnosed in the united states in 2006-2010. *Neuro Oncol*, 15 Suppl 2:ii1–56, Nov 2013.
- [99] T. Palomero, W. K. Lim, D. T. Odom, M. L. Sulis, P. J. Real, A. Margolin, K. C. Barnes, J. O’Neil, D. Neuberg, A. P. Weng, J. C. Aster, F. Sigaux, J. Soulier, A. T. Look, R. A. Young, A. Califano, and A. A. Ferrando. Notch1 directly regulates c-myc and activates a feed-forward-loop transcriptional network promoting leukemic cell growth. *Proc Natl Acad Sci U S A*, 103(48):18261–6, 2006.
- [100] K. Pandya, K. Meeke, A. G. Clementz, A. Rogowski, J. Roberts, L. Miele, K. S. Albain, and C. Osipo. Targeting both notch and erbb-2 signalling pathways is required for prevention of erbb-2-positive breast tumour recurrence. *Br J Cancer*, 105(6):796–806, 2011.
- [101] Brittany C Parker, Matti J Annala, David E Cogdell, Kirsi J Granberg, Yan Sun, Ping Ji, Xia Li, Joy Gumin, Hong Zheng, Limei Hu, Olli Yli-Harja, Hannu Haapasalo, Tapio Visakorpi, Xiuping Liu, Chang-Gong Liu, Raymond Sawaya, Gregory N Fuller, Kexin Chen, Frederick F Lang, Matti Nykter, and Wei Zhang. The tumorigenic fgfr3-tacc3 gene fusion escapes mir-99a regulation in glioblastoma. *J Clin Invest*, 123(2):855–65, Feb 2013.
- [102] J. S. Parker, M. Mullins, M. C. Cheang, S. Leung, D. Voduc, T. Vickery, S. Davies, C. Fauron, X. He, Z. Hu, J. F. Quackenbush, I. J. Stijleman, J. Palazzo, J. S. Marron, A. B. Nobel, E. Mardis, T. O. Nielsen, M. J. Ellis, C. M. Perou, and P. S.



- Bernard. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*, 27(8):1160–7, 2009.
- [103] Joel S Parker, Michael Mullins, Maggie C U Cheang, Samuel Leung, David Voduc, Tammi Vickery, Sherri Davies, Christiane Fauron, Xiaping He, Zhiyuan Hu, John F Quackenbush, Inge J Stijleman, Juan Palazzo, J S Marron, Andrew B Nobel, Elaine Mardis, Torsten O Nielsen, Matthew J Ellis, Charles M Perou, and Philip S Bernard. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*, 27(8):1160–1167, Mar 2009.
- [104] D. Parkhomchuk, T. Borodina, V. Amstislavskiy, M. Banaru, L. Hallen, S. Krobitsch, H. Lehrach, and A. Soldatov. Transcriptome analysis by strand-specific sequencing of complementary dna. *Nucleic Acids Res*, 37(18):e123, 2009.
- [105] D Williams Parsons, Siân Jones, Xiaosong Zhang, Jimmy Cheng-Ho Lin, Rebecca J Leary, Philipp Angenendt, Parminder Mankoo, Hannah Carter, I-Mei Siu, Gary L Gallia, Alessandro Olivi, Roger McLendon, B Ahmed Rasheed, Stephen Keir, Tatiana Nikolskaya, Yuri Nikolsky, Dana A Busam, Hanna Tekleab, Luis A Diaz, Jr, James Hartigan, Doug R Smith, Robert L Strausberg, Suely Kazue Nagahashi Marie, Sueli Mieko Oba Shinjo, Hai Yan, Gregory J Riggins, Darell D Bigner, Rachel Karchin, Nick Papadopoulos, Giovanni Parmigiani, Bert Vogelstein, Victor E Velculescu, and Kenneth W Kinzler. An integrated genomic analysis of human glioblastoma multiforme. *Science*, 321(5897):1807–12, Sep 2008.
- [106] Carmen Phillips. Treatment options for her2-positive breast cancer expand and evolve, 2012.
- [107] Heidi S Phillips, Samir Kharbanda, Ruihuan Chen, William F Forrest, Robert H Soriano, Thomas D Wu, Anjan Misra, Janice M Nigro, Howard Colman, Liliana Soroceanu, P Mickey Williams, Zora Modrusan, Burt G Feuerstein, and Ken Aldape. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell*, 9(3):157–73, Mar 2006.
- [108] Sandhya Pruthi. Her2-positive breast cancer: What is it?, 2012.
- [109] Aaron R Quinlan and Ira M Hall. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6):841–2, Mar 2010.
- [110] Amie J Radenbaugh, Singer Ma, Adam Ewing, Joshua M Stuart, Eric A Collisson, Jingchun Zhu, and David Haussler. Radia: Rna and dna integrated analysis for somatic mutation detection. *PLoS One*, 9(11):e111516, 2014.

- [111] Tobias Rausch, David T W Jones, Marc Zapatka, Adrian M Stütz, Thomas Zichner, Joachim Weischenfeldt, Natalie Jäger, Marc Remke, David Shih, Paul A Northcott, Elke Pfaff, Jelena Tica, Qi Wang, Luca Massimi, Hendrik Witt, Sebastian Bender, Sabrina Pleier, Huriye Cin, Cynthia Hawkins, Christian Beck, Andreas von Deimling, Volkmar Hans, Benedikt Brors, Roland Eils, Wolfram Scheurlen, Jonathon Blake, Vladimir Benes, Andreas E Kulozik, Olaf Witt, Dianna Martin, Cindy Zhang, Rinnat Porat, Diana M Merino, Jonathan Wasserman, Nada Jabado, Adam Fontebasso, Lars Bullinger, Frank G Rücker, Konstanze Döhner, Hartmut Döhner, Jan Koster, Jan J Molenaar, Rogier Versteeg, Marcel Kool, Uri Tabori, David Malkin, Andrey Korshunov, Michael D Taylor, Peter Lichter, Stefan M Pfister, and Jan O Korbel. Genome sequencing of pediatric medulloblastoma links catastrophic dna rearrangements with tp53 mutations. *Cell*, 148(1-2):59–71, Jan 2012.
- [112] K S Reddy. Double minutes (dmin) and homogeneously staining regions (hsr) in myeloid disorders: a new case suggesting that dmin form hsr in vivo. *Cytogenet Genome Res*, 119(1-2):53–9, 2007.
- [113] M. Reedijk, D. Pinnaduwege, B. C. Dickson, A. M. Mulligan, H. Zhang, S. B. Bull, F. P. O’Malley, S. E. Egan, and I. L. Andrulis. Jag1 expression is associated with a basal phenotype and recurrence in lymph node-negative breast cancer. *Breast Cancer Res Treat*, 111(3):439–48, 2008.
- [114] David Reverter, Klaus Maskos, Fulong Tan, Randal A Skidgel, and Wolfram Bode. Crystal structure of human carboxypeptidase m, a membrane-bound enzyme that regulates peptide hormone activity. *J Mol Biol*, 338(2):257–69, Apr 2004.
- [115] Dan Rohle, Janeta Popovici-Muller, Nicolaos Palaskas, Sevin Turcan, Christian Grommes, Carl Campos, Jennifer Tsoi, Owen Clark, Barbara Oldrini, Evangelia Komisopoulou, Kaiko Kunii, Alicia Pedraza, Stefanie Schalm, Lee Silverman, Alexandra Miller, Fang Wang, Hua Yang, Yue Chen, Andrew Kernytsky, Marc K Rosenblum, Wei Liu, Scott A Biller, Shinsan M Su, Cameron W Brennan, Timothy A Chan, Thomas G Graeber, Katharine E Yen, and Ingo K Mellinghoff. An inhibitor of mutant idh1 delays growth and promotes differentiation of glioma cells. *Science*, 340(6132):626–30, May 2013.
- [116] R. Roskoski. The erbb/her family of protein-tyrosine kinases and cancer. *Pharmacol Res*, 79:34–74, 2014.
- [117] S. Sahebjam, P. L. Bedard, V. Castonguay, Z. Chen, M. Reedijk, G. Liu, B. Cohen, W. J. Zhang, B. Clarke, T. Zhang, S. Kamel-Reid, H. Chen, S. P. Ivy, A. R. Razak, A. M. Oza, E. X. Chen, H. W. Hirte, A. McGarrity, L. Wang, L. L. Siu, and S. J.

- Hotte. A phase i study of the combination of ro4929097 and cediranib in patients with advanced solid tumours (pjc-004/nci 8503). *Br J Cancer*, 109(4):943–9, 2013.
- [118] Felix Sahn, David Reuss, Christian Koelsche, David Capper, Jens Schittenhelm, Stephanie Heim, David T W Jones, Stefan M Pfister, Christel Herold-Mende, Wolfgang Wick, Wolf Mueller, Christian Hartmann, Werner Paulus, and Andreas von Deimling. Farewell to oligoastrocytoma: in situ molecular genetics favor classification as either oligodendroglioma or astrocytoma. *Acta Neuropathol*, 128(4):551–9, Oct 2014.
- [119] R. Saito, M. E. Smoot, K. Ono, J. Ruschinski, P. L. Wang, S. Lotia, A. R. Pico, G. D. Bader, and T. Ideker. A travel guide to cytoscape plugins. *Nat Methods*, 9(11):1069–76, 2012.
- [120] Nader Sanai, Susan Chang, and Mitchel S Berger. Low-grade gliomas in adults. *J Neurosurg*, 115(5):948–65, Nov 2011.
- [121] J Zachary Sanborn, Sofie R Salama, Mia Grifford, Cameron W Brennan, Tom Mikkelsen, Suresh Jhanwar, Sol Katzman, Lynda Chin, and David Haussler. Double minute chromosomes in glioblastoma multiforme are revealed by precise reconstruction of oncogenic amplicons. *Cancer Res*, 73(19):6036–45, Oct 2013.
- [122] C. F. Schaefer, K. Anthony, S. Krupa, J. Buchoff, M. Day, T. Hannay, and K. H. Buetow. Pid: the pathway interaction database. *Nucleic Acids Res*, 37(Database issue):D674–9, 2009.
- [123] C. A. Schneider, W. S. Rasband, and K. W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nat Methods*, 9(7):671–5, 2012.
- [124] Heidi Schwarzenbach, Dave S B Hoon, and Klaus Pantel. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer*, 11(6):426–37, Jun 2011.
- [125] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–504, 2003.
- [126] S. V. Sharma, D. Y. Lee, B. Li, M. P. Quinlan, F. Takahashi, S. Maheswaran, U. McDermott, N. Azizian, L. Zou, M. A. Fischbach, K. K. Wong, K. Brandstetter, B. Wittner, S. Ramaswamy, M. Classon, and J. Settleman. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell*, 141(1):69–80, 2010.
- [127] Suzanne Sindi, Elena Helman, Ali Bashir, and Benjamin J Raphael. A geometric

- approach for classification and comparison of structural variants. *Bioinformatics*, 25(12):i222–30, Jun 2009.
- [128] Devendra Singh, Joseph Minhow Chan, Pietro Zoppoli, Francesco Niola, Ryan Sullivan, Angelica Castano, Eric Minwei Liu, Jonathan Reichel, Paola Porrati, Serena Pellegatta, Kunlong Qiu, Zhibo Gao, Michele Ceccarelli, Riccardo Riccardi, Daniel J Brat, Abhijit Guha, Ken Aldape, John G Golfinos, David Zagzag, Tom Mikkelsen, Gaetano Finocchiaro, Anna Lasorella, Raul Rabadan, and Antonio Iavarone. Transforming fusions of fgfr and tacc genes in human glioblastoma. *Science*, 337(6099):1231–5, Sep 2012.
- [129] Johan Skog, Tom Würdinger, Sjoerd van Rijn, Dimphna H Meijer, Laura Gainche, Miguel Sena-Esteves, William T Curry, Jr, Bob S Carter, Anna M Krichevsky, and Xandra O Breakefield. Glioblastoma microvesicles transport rna and proteins that promote tumour growth and provide diagnostic biomarkers. *Nat Cell Biol*, 10(12):1470–6, Dec 2008.
- [130] D J Slamon, G M Clark, S G Wong, W J Levin, A Ullrich, and W L McGuire. Human breast cancer: correlation of relapse and survival with amplification of the her-2/neu oncogene. *Science*, 235(4785):177–182, Jan 1987.
- [131] Green P Smit AFA, Hubley R. Repeatmasker open-3.0, 1996-2010.
- [132] M. E. Smoot, K. Ono, J. Ruscheinski, P. L. Wang, and T. Ideker. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27(3):431–2, 2011.
- [133] American Cancer Society. Cancer facts & figures 2011. Atlanta: American Cancer Society, 2011.
- [134] T Sorlie, C M Perou, R Tibshirani, T Aas, S Geisler, H Johnsen, T Hastie, M B Eisen, M van de Rijn, S S Jeffrey, T Thorsen, H Quist, J C Matese, P O Brown, D Botstein, P Eystein Lonning, and A L Borresen-Dale. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*, 98(19):10869–10874, Sep 2001.
- [135] Philip J Stephens, Chris D Greenman, Beiyan Fu, Fengtang Yang, Graham R Bignell, Laura J Mudie, Erin D Pleasance, King Wai Lau, David Beare, Lucy A Stebbings, Stuart McLaren, Meng-Lay Lin, David J McBride, Ignacio Varela, Serena Nik-Zainal, Catherine Leroy, Mingming Jia, Andrew Menzies, Adam P Butler, Jon W Teague, Michael A Quail, John Burton, Harold Swerdlow, Nigel P Carter, Laura A Morsberger, Christine Iacobuzio-Donahue, George A Follows, Anthony R Green, Adrienne M Flanagan, Michael R Stratton, P Andrew Futreal, and Peter J

- Campbell. Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1):27–40, Jan 2011.
- [136] Clelia Tiziana Storlazzi, Angelo Lonoce, Maria C Guastadisegni, Domenico Trombetta, Pietro D’Addabbo, Giulia Daniele, Alberto L’Abbate, Gemma Macchia, Cecilia Surace, Klaas Kok, Reinhard Ullmann, Stefania Purgato, Orazio Palumbo, Massimo Carella, Peter F Ambros, and Mariano Rocchi. Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure. *Genome Res*, 20(9):1198–206, Sep 2010.
- [137] B Streubel, P Valent, U Jäger, M Edelhäuser, H Wandt, T Wagner, T Büchner, K Lechner, and C Fonatsch. Amplification of the mll gene on double minutes, a homogeneously staining region, and ring chromosomes in five patients with acute myeloid leukemia or myelodysplastic syndrome. *Genes Chromosomes Cancer*, 27(4):380–6, Apr 2000.
- [138] A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, 102(43):15545–50, 2005.
- [139] Ghazaleh Tabatabai, Roger Stupp, Martin J van den Bent, Monika E Hegi, Jörg C Tonn, Wolfgang Wick, and Michael Weller. Molecular diagnostics of gliomas: the clinical perspective. *Acta Neuropathol*, 120(5):585–92, Nov 2010.
- [140] Brett J Theeler, W K Alfred Yung, Gregory N Fuller, and John F De Groot. Moving toward molecular classification of diffuse gliomas in adults. *Neurology*, 79(18):1917–26, Oct 2012.
- [141] Wandaliz Torres-García, Siyuan Zheng, Andrey Sivachenko, Rahulsimham Vegesna, Qianghu Wang, Rong Yao, Michael F Berger, John N Weinstein, Gad Getz, and Roel G W Verhaak. Prada: pipeline for rna sequencing data analysis. *Bioinformatics*, 30(15):2224–6, Aug 2014.
- [142] Sevin Turcan, Armida W M Fabius, Alexandra Borodovsky, Alicia Pedraza, Cameron Brennan, Jason Huse, Agnes Viale, Gregory J Riggins, and Timothy A Chan. Efficient induction of differentiation and growth inhibition in idh1 mutant glioma cells by the dnmt inhibitor decitabine. *Oncotarget*, 4(10):1729–36, Oct 2013.
- [143] Martin J van den Bent. Interobserver variation of the histopathological diagnosis

- in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol*, 120(3):297–304, Sep 2010.
- [144] P Varlet, A Jouvet, C Miquel, G Saint-Pierre, F Beuvon, and C Daumas-Duport. [criteria of diagnosis and grading of oligodendrogliomas or oligo-astrocytomas according to the who and sainte-anne classifications]. *Neurochirurgie*, 51(3-4 Pt 2):239–46, Sep 2005.
- [145] Roel G W Verhaak, Katherine A Hoadley, Elizabeth Purdom, Victoria Wang, Yuan Qi, Matthew D Wilkerson, C Ryan Miller, Li Ding, Todd Golub, Jill P Mesirov, Gabriele Alexe, Michael Lawrence, Michael O'Kelly, Pablo Tamayo, Barbara A Weir, Stacey Gabriel, Wendy Winckler, Supriya Gupta, Lakshmi Jakkula, Heidi S Feiler, J Graeme Hodgson, C David James, Jann N Sarkaria, Cameron Brennan, Ari Kahn, Paul T Spellman, Richard K Wilson, Terence P Speed, Joe W Gray, Matthew Meyerson, Gad Getz, Charles M Perou, D Neil Hayes, and Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in *pdgfra*, *idh1*, *egfr*, and *nf1*. *Cancer Cell*, 17(1):98–110, Jan 2010.
- [146] Nicolas Vogt, Sandrine-Hélène Lefèvre, Françoise Apiou, Anne-Marie Dutrillaux, Andrej Cör, Pascal Leuraud, Marie-France Poupon, Bernard Dutrillaux, Michelle Debatisse, and Bernard Malfoy. Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. *Proc Natl Acad Sci U S A*, 101(31):11368–73, Aug 2004.
- [147] RA Weinberg. *The Biology Of Cancer*. Garland Science, Taylor & Francis Group, LLC, 2007.
- [148] Michael Weller, Roger Stupp, Monika E Hegi, Martin van den Bent, Joerg C Tonn, Marc Sanson, Wolfgang Wick, and Guido Reifenberger. Personalized care in neuro-oncology coming of age: why we need mgmt and 1p/19q testing for malignant glioma patients in clinical practice. *Neuro Oncol*, 14 Suppl 4:iv100–8, Sep 2012.
- [149] Patrick Y Wen and Santosh Kesari. Malignant gliomas in adults. *N Engl J Med*, 359(5):492–507, Jul 2008.
- [150] Andrew P Weng, Adolfo A Ferrando, Woojoong Lee, John P Morris, 4th, Lewis B Silverman, Cheryl Sanchez-Irizarry, Stephen C Blacklow, A Thomas Look, and Jon C Aster. Activating mutations of *notch1* in human t cell acute lymphoblastic leukemia. *Science*, 306(5694):269–71, Oct 2004.
- [151] Benedikt Wiestler, David Capper, Tim Holland-Letz, Andrey Korshunov, An-

- reas von Deimling, Stefan Michael Pfister, Michael Platten, Michael Weller, and Wolfgang Wick. Atrx loss refines the classification of anaplastic gliomas and identifies a subgroup of idh mutant astrocytic tumors with better prognosis. *Acta Neuropathol*, 126(3):443–51, Sep 2013.
- [152] Benedikt Wiestler, David Capper, Martin Sill, David T W Jones, Volker Hovestadt, Dominik Sturm, Christian Koelsche, Anna Bertoni, Leonille Schweizer, Andrey Korshunov, Elisa K Weiß, Maximilian G Schliesser, Alexander Radbruch, Christel Herold-Mende, Patrick Roth, Andreas Unterberg, Christian Hartmann, Torsten Pietsch, Guido Reifenberger, Peter Lichter, Bernhard Radlwimmer, Michael Platten, Stefan M Pfister, Andreas von Deimling, Michael Weller, and Wolfgang Wick. Integrated dna methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma. *Acta Neuropathol*, 128(4):561–71, Oct 2014.
- [153] M. D. Wilkerson and D. N. Hayes. Consensusclusterplus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*, 26(12):1572–3, 2010.
- [154] Wenle Xia, Robert J Mullin, Barry R Keith, Lei-Hua Liu, Hong Ma, David W Rusnak, Gary Owens, Krystal J Alligood, and Neil L Spector. Anti-tumor activity of gw572016: a dual tyrosine kinase inhibitor blocks egf activation of egfr/erbb2 and downstream erk1/2 and akt pathways. *Oncogene*, 21(41):6255–6263, Sep 2002.
- [155] Hai Yan, D Williams Parsons, Genglin Jin, Roger McLendon, B Ahmed Rasheed, Weishi Yuan, Ivan Kos, Ines Batinic-Haberle, Siân Jones, Gregory J Riggins, Henry Friedman, Allan Friedman, David Reardon, James Herndon, Kenneth W Kinzler, Victor E Velculescu, Bert Vogelstein, and Darell D Bigner. Idh1 and idh2 mutations in gliomas. *N Engl J Med*, 360(8):765–73, Feb 2009.
- [156] S. Yano, E. Nakataki, S. Ohtsuka, M. Inayama, H. Tomimoto, N. Edakuni, S. Kakiuchi, N. Nishikubo, H. Muguruma, and S. Sone. Retreatment of lung adenocarcinoma patients with gefitinib who had experienced favorable results from their initial treatment with this selective epidermal growth factor receptor inhibitor: a report of three cases. *Oncol Res*, 15(2):107–11, 2005.
- [157] Christina Yau, Yixin Wang, Yi Zhang, John A Foekens, and Christopher C Benz. Young age, increased tumor proliferation and foxm1 expression predict early metastatic relapse only for endocrine-dependent breast cancers. *Breast Cancer Res Treat*, 126(3):803–810, Apr 2011.
- [158] Stephen Yip, Yaron S Butterfield, Olena Morozova, Suganthi Chittaranjan, Michael D Blough, Jianghong An, Inanc Birol, Charles Chesnelong, Readman

- Chiu, Eric Chuah, Richard Corbett, Rod Docking, Marlo Firme, Martin Hirst, Shaun Jackman, Aly Karsan, Haiyan Li, David N Louis, Alexandra Maslova, Richard Moore, Annie Moradian, Karen L Mungall, Marco Perizzolo, Jenny Qian, Gloria Roldan, Eric E Smith, Jessica Tamura-Wells, Nina Thiessen, Richard Varhol, Samuel Weiss, Wei Wu, Sean Young, Yongjun Zhao, Andrew J Mungall, Steven J M Jones, Gregg B Morin, Jennifer A Chan, J Gregory Cairncross, and Marco A Marra. Concurrent cic mutations, idh mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers. *J Pathol*, 226(1):7–16, Jan 2012.
- [159] M Yoshimoto, S R Caminada De Toledo, E M Monteiro Caran, M T de Seixas, M L de Martino Lee, S de Campos Vieira Abib, S M Vianna, S T Schettini, and J Anderson Duffles Andrade. Mycn gene amplification. identification of cell populations containing double minutes and homogeneously staining regions in neuroblastoma tumors. *Am J Pathol*, 155(5):1439–43, Nov 1999.
- [160] J. Yun, A. Pannuti, I. Espinoza, H. Zhu, C. Hicks, X. Zhu, M. Caskey, P. Rizzo, G. D’Souza, K. Backus, M. F. Denning, J. Coon, M. Sun, E. H. Bresnick, C. Osipo, J. Wu, P. R. Strack, D. A. Tonetti, and L. Miele. Crosstalk between pkc and notch-4 in endocrine-resistant breast cancer cells. *Oncogenesis*, 2:e60, 2013.
- [161] Siyuan Zheng, Jun Fu, Rahulsimham Vegesna, Yong Mao, Lindsey E Heathcock, Wandaliz Torres-Garcia, Ravesanker Ezhilarasan, Shuzhen Wang, Aaron McKenna, Lynda Chin, Cameron W Brennan, W K Alfred Yung, John N Weinstein, Kenneth D Aldape, Erik P Sulman, Ken Chen, Dimpy Koul, and Roel G W Verhaak. A survey of intragenic breakpoints in glioblastoma identifies a distinct subset associated with poor survival. *Genes Dev*, 27(13):1462–72, Jul 2013.
- [162] J. Zhu, J. Z. Sanborn, S. Benz, C. Szeto, F. Hsu, R. M. Kuhn, D. Karolchik, J. Archie, M. E. Lenburg, L. J. Esserman, W. J. Kent, D. Haussler, and T. Wang. The ucsc cancer genomics browser. *Nat Methods*, 6(4):239–40, 2009.