Molecular Characterization of Genetic and Epigenetic Alterations in Gliomas

by

Christopher Gentry Duncan

Department of Pathology
Duke University

Date:_______________________

Approved:

___________________________
Hai Yan, Supervisor

___________________________
Soman N. Abraham

___________________________
Robin E. Bachelder

___________________________
Roger E. McLendon

___________________________
Xiao-Fan Wang

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Pathology in the Graduate School of Duke University

2012
ABSTRACT

Molecular Characterization of Genetic and Epigenetic Alterations in Gliomas

by

Christopher Gentry Duncan

Department of Pathology
Duke University

Date:_______________________
Approved:

Hai Yan, Supervisor
Soman N. Abraham
Robin E. Bachelder
Roger E. McLendon
Xiao-Fan Wang

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Pathology in the Graduate School of Duke University

2012
Abstract

Glioma development and progression are driven by complex genetic alterations, including point mutations and gain or loss of genomic copy number, as well as epigenetic aberrations, including DNA methylation and histone modifications. However, the molecular mechanisms underlying the causes and effects of these alterations are poorly understood, and improved treatments are greatly needed. Here, we report a comprehensive evaluation of the recurrent genomic alterations in gliomas and further dissect the molecular effects of the most frequently-occurring genomic events. First, we performed a multifaceted genomic analysis to identify genes targeted by copy number alteration in glioblastoma, the most aggressive malignant glioma. We identify EGFR negative regulator, ERRFI1, as a glioblastoma-targeted gene within the minimal region of deletion in 1p36.23. Furthermore, we demonstrate that Aurora-A kinase substrate, TACC3, displays gain of copy number on 4p16.3 and is overexpressed in a grade-specific pattern. Next, using a gene targeting approach, we knocked-in a single copy of the most frequently observed point mutation in gliomas, IDH1R132H/WT, into a human cancer cell line. We show that heterozygous expression of the IDH1R132H allele is sufficient to induce the genome-wide alterations in DNA methylation characteristic of these tumors. Together, these data provide insight on genetic and epigenetic alterations which drive human gliomas.
# Contents

Abstract .................................................................................................................................iv

List of Figures ........................................................................................................................ix

Acknowledgements ..............................................................................................................xi

1. Introduction to this dissertation .........................................................................................1
   1.1 Permissions and collaborative work.............................................................................1
   1.2 Genomic alterations and the pathogenesis of glioblastoma .........................................1

2. Integrated genomic analyses identify ERRFI1 and TACC3 as glioblastoma-targeted genes .........................................................................................................................................7
   2.1 Background: Genomic analysis of glioblastoma ............................................................7
   2.2 Methods ............................................................................................................................9
      2.2.1 Tumor samples .........................................................................................................9
      2.2.2 Digital karyotyping ..................................................................................................9
      2.2.3 High density SNP-Arrays .......................................................................................10
      2.2.4 TCGA data ..............................................................................................................10
      2.2.5 Bioinformatic analysis ...........................................................................................11
      2.2.6 Quantitative real-time PCR ..................................................................................11
      2.2.7 Migration assays ....................................................................................................12
   2.3 Results ............................................................................................................................13
      2.3.1 Detection of focal copy number alterations by DK and Illumina BeadChips...13
      2.3.2 ERRFI1 on 1p36 is a candidate tumor suppressor gene, whose products regulate glioblastoma cell migration .................................................................16
2.3.3 TACC3 on 4p16.3 is overexpressed and correlates with the expression of Aurora kinases in glioblastomas ................................................................. 21

2.4 Discussion of copy number alterations in glioblastoma ........................................ 25
  2.4.1 Deletions of 1p36 ....................................................................................... 25
  2.4.2 Copy number gains of 4p16 ....................................................................... 26

3. A heterozygous IDH1R132H/WT mutation induces genome-wide alterations in DNA methylation ...................................................................................... 28
  3.1 Background: IDH mutations and the epigenome .............................................. 28
  3.2 Methods ........................................................................................................... 31
    3.2.1 Cell culture and drug treatment ................................................................. 31
    3.2.2 Gene targeting of the human IDH1 locus ................................................ 32
    3.2.3 D-2-HG analysis ...................................................................................... 33
    3.2.4 Genome-wide CpG methylation profiling ................................................. 33
    3.2.5 Statistical analysis of DNA methylation .................................................... 34
    3.2.6 Genome-wide expression profiling ........................................................... 37
    3.2.7 Statistical analysis of gene expression ....................................................... 37
    3.2.8 Genomic bisulfite sequencing analysis .................................................... 39
    3.2.9 Histone extraction and western analysis .................................................. 39
    3.2.10 Chromatin immunoprecipitation ............................................................ 40
    3.2.11 Quantitative real-time reverse transcriptase PCR .................................... 41
    3.2.12 Analysis of primary glioblastomas and low grade gliomas ....................... 41
    3.2.13 Analysis of HCT116 ChIP-seq data ......................................................... 42
  3.3 Results .............................................................................................................. 43
3.3.1 Targeted knock-in of IDH1R132H/WT hotspot mutation in a human cancer cell line .................................................................43

3.3.2 IDH1R132H/WT induces alterations in DNA methylation ..........................48

3.3.3 Methylation alterations observed in HCT116 IDH1R132H/WT knock-in cell lines are similarly affected in a brain tumor cell line overexpressing IDH1R132H ..................54

3.3.4 Methylation alterations observed in HCT116 IDH1R132H/WT knock-in cell lines are similarly associated with IDH1 mutation in IDH1-mutant and G-CIMP+ primary gliomas .................................................................58

3.3.5 Effects of IDH1R132H/WT on gene expression ........................................65

3.3.6 Relationship between IDH1 mutation-induced alterations in DNA methylation and gene expression ............................................................................69

3.3.7 Global and gene-specific histone lysine methylation modifications correlate with DNA methylation and gene expression alterations in IDH1R132H/WT cells ..........73

3.3.8 Gene silencing at specific IDH1R132H/WT-targeted loci is reversed using a DNA hypomethylating agent ........................................................................79

3.4 Discussion of IDH mutations and their impact on the epigenome ..................82

3.4.1 Heterozygous IDH1 mutations induce alterations in DNA methylation .......82

3.4.2 Enzymatic activity of heterozygous IDH mutations ...................................83

3.4.3 TET proteins and 5-hydroxymethylcytosine .............................................84

3.4.4 Histone (de)methylation .............................................................................86

3.4.5 Relationship between DNA methylation and gene expression ..................88

3.4.6 Targets of mutant IDH1-mediated de novo epigenetic silencing ...............91

4. Summary and future directions ......................................................................94

4.1 Summary and conclusions .............................................................................94

4.2 Future studies .................................................................................................96
List of Figures

Figure 1: Schematic depiction of recurrent genetic alterations and major signaling pathways involved in the pathogenesis of glioblastoma.................................3

Figure 2: Focal high copy number gains and homozygous deletions in glioblastomas...15

Figure 3: High-resolution mapping of homozygous deletions detects DFFB, C1orf174, and LOC100133612 within secondary MDR on 1p36.32..................................................17

Figure 4: High-resolution mapping of homozygous deletions reveals ERRFI1 within the most frequent MDR on 1p36.23. .................................................................18

Figure 5: ERRFI1 is silenced in glioblastomas and reduces cell migration in H423 glioblastoma cells........................................................................................................20

Figure 6: High-resolution mapping identifies TACC3 as the glioblastoma-targeted gene on 4p16.3 ........................................................................................................22

Figure 7: TACC3 is the predominant gene upregulated in 4p16.3 and correlates with Aurora kinase expression. .................................................................23

Figure 8: Grade-specific TACC3 upregulation and TACC2 downregulation in gliomas.24

Figure 9: Targeted knock-in of IDH1R132H/WT hotspot mutation in a human cancer cell line. ........................................................................................................44

Figure 10: Diagnostic PCR screens for generation of knock-in clones. ......................46

Figure 11: IDH1R132H/WT-induced DNA methylation alterations in HCT116 cells. ....49

Figure 12: Consistent IDH1R132H/WT-induced DNA methylation changes between independent HCT116 knock-in clones.................................................................52

Figure 13: Human oligodendroglioma (HOG) cells overexpressing IDH1R132H recapitulate the changes in DNA methylation observed in HCT116 cells. .......................56

Figure 14: IDH1 mutant and G-CIMP+ gliomas recapitulate the DNA methylation alterations observed in cell line models .........................................................60
Figure 15: Comparison of HCT116 IDH1<sup>R132H/WT</sup> differentially methylated loci with primary tumors yields a significant overlap.................................................................62

Figure 16: Gene expression profiling of HCT116 IDH1<sup>R132H/WT</sup> cell lines. ........................................67

Figure 17: HCT116 DNA methylation and gene expression changes negatively correlate. .................................................................70

Figure 18: Distribution of HCT116 IDH1<sup>R132H/WT</sup> differentially methylated loci relative to transcription start sites (TSS) and CpG islands. .........................................................71

Figure 19: RNA polymerase II (Pol II) and Histone 3 Lysine 4 trimethylation (H3K4me3) ChIP-seq data indicates IDH1<sup>R132H/WT</sup> differentially methylated loci have reduced Pol II binding in HCT116 wild-type cells.................................................................72

Figure 20: Bisulfite sequence analysis of candidate CpG loci validates IDH1<sup>R132H/WT</sup>-induced DNA methylation changes.................................................................75

Figure 21: Bisulfite sequencing at candidate loci of HCT116 non-targeted clones indicates no change in methylation is resultant of the gene targeting procedure........................................76

Figure 22: Global and gene-specific histone lysine methylation coincides with IDH1<sup>R132H/WT</sup>-induced DNA methylation.................................................................78

Figure 23: Inhibition of DNA methylation results in restoration of gene expression for IDH1<sup>R132H/WT</sup>-repressed loci.................................................................80

Figure 24: Schematic depiction of a hypothesized link between IDH1 mutations and epigenetic alterations in gliomas. .................................................................98
Acknowledgements

was reproduced according to the journal/publisher policies from Landes Bioscience Journals (Duncan and Yan 2011), Impact Journals (Duncan et al. 2010), and Cold Spring Harbor Laboratory Press (Duncan et al. 2012) as described in Appendix A. The work described here was supported by American Cancer Society Research Scholar Award RSG-10-126-01-CCE, The Pediatric Brain Tumor Foundation Institute Grant, The Southeastern Brain Tumor Foundation Grant and NCI Grants R01-CA118822, 5R01-CA140316, 2RO1-CA077337 and 5RO1-CA132065, as well as by the following NIH grants: NINDS Grant 5P50 NS20023, NCI SPORE Grant 5P50 CA108786, and NCI Merit Award R37 CA 011898.
1. Introduction to this dissertation

1.1 Permissions and collaborative work

This dissertation contains figures and excerpts of the text of published articles. The excerpts and figures from published works were reproduced with permission as described in Appendix A. Where the publisher does not require explicit permission, excerpts and figures were reproduced in accordance with the policy of the journal or publisher as described in Appendix A. The work detailed in this dissertation was performed primarily by the candidate, although in many cases the work was aided by collaborators, for instance by performing analyses or by providing biological samples as described in the Acknowledgements.

1.2 Genomic alterations and the pathogenesis of glioblastoma

While years of research have contributed to understanding the molecular mechanisms underlying initiation and progression of glioblastoma, the most common malignant brain tumor, prognosis remains dismal. Recent efforts have focused on the identification of glioblastoma-targeted genes through comprehensive genomic studies, revealing patterns of genetic alterations that include coding sequence mutations, gain or loss of genomic DNA copy number, and alterations of mRNA expression signature (Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008; Ohgaki and Kleihues 2009; Purow and Schiff 2009).
The genomic landscape of glioblastoma is highly heterogeneous, as genetic alterations vary highly from tumor to tumor, but common themes have been elucidated. Together, major genes altered in glioblastomas have been identified and shown to contribute to core disease pathways including receptor tyrosine kinase (RTK), TP53, and RB signaling (Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008) (Figure 1). Among the most frequent alterations of RTK signaling are mutations or amplifications of the growth factor receptors EGFR, ERBB2, PDGFRA, and MET, mutation of phosphatidylinositol 3-kinases PIK3CA and PIK3R1, and mutation or homozygous deletion of the PTEN tumor suppressor gene. Inactivation of the RB tumor suppressor pathway was primarily realized as deletions of the CDKN2A/CDKN2B locus, amplification of CDK4, CDK6, and CCND2, and mutation or deletion of RB1. Somatic alterations of TP53 signaling comprised mutations and deletions of TP53 and amplifications of MDM2 and MDM4. Collectively, these core glioblastoma pathways promote cellular proliferation and survival, inhibit cell death, and advance cell cycle progression. An additional gene alteration that has recently been identified is point mutations in the isocitrate dehydrogenases, IDH1 and IDH2, at high frequency in secondary glioblastomas and progressive gliomas (Parsons et al. 2008; Yan et al. 2009b). IDH1 and IDH2 mutations confer gain of a novel enzymatic activity to catalyze the production of D-2-hydroxyglutarate while impairing normal conversion of isocitrate to α-ketoglutarate. While specific mechanistic details are uncertain, recent insights
Figure 1: Schematic depiction of recurrent genetic alterations and major signaling pathways involved in the pathogenesis of glioblastoma.

Frequent gene activation (shown in red) or inactivation (shown in blue) contributes to core pathways in glioblastoma which enhance cellular proliferation, survival, and translation, inhibit apoptotic signaling, promote cell cycle progression, and alter cellular metabolism. Additionally, central to the pathogenesis of glioblastoma is the maintenance of genome integrity and cell division. αKG, α-ketoglutarate; D2HG, D-2-hydroxyglutarate; IDH1*, mutant IDH1.
emphasize the importance of metabolic enzymes and cellular metabolism to gliomagenesis (Dang et al. 2009; Reitman et al. 2011). Based on gene expression profile and characterized by abnormalities in PDGFRα, IDH1, EGFR, and NF1, further analyses have classified glioblastomas into four distinct molecular subtypes: proneural, neural, classical, and mesenchymal (Verhaak et al. 2010).

The collective body of glioblastoma genomic knowledge comprises a sizeable list of altered genes (Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008), only a fraction of which are known to contribute to the core signaling pathways. The genes previously unlinked to glioblastoma can potentially contribute to currently unrecognized pathways or promote signaling of core pathways. Particularly challenging is the characterization of low frequency (1%-5%) events and genomic events which affect numerous genes, such as large amplifications or deletions. Integrated genomic analyses can now provide clues for recognition of critical events and expedite the process of understanding gene function in relation to tumorigenesis. Two particularly interesting genomic alterations recently cited in glioblastoma occur at 1p36 and 4p16.

Focal homozygous deletions of 1p36 are of the most common recurrent copy number alterations in glioblastoma, occurring in over 8% of patients (Duncan et al. 2010). Despite the frequency of 1p36 copy number loss, the potential tumor suppressor at that genomic locus remained elusive, as the bounds of observed 1p36 deletions are
highly heterogeneous and typically disrupt multiple genes. In attempt to pinpoint the 1p36 target genes, fine mapping analysis of focal 1p36 homozygous deletions was performed across multiple tumors and platforms, identifying two minimal regions of deletion at 1p36.23 and 1p36.32 (Duncan et al. 2010). Specifically, an EGFR negative regulator, ERRFI1, was identified within the most common minimal region of deletion in 1p36.23, which provides evidence that the majority of 1p36 deletions contribute to sustaining RTK signaling in glioblastomas.

Likewise, genomic alterations of 4p16 had been reported in several cancers, indicating the presence of one or more oncogenes. Genome-wide copy number analysis revealed recurrent focal duplications at 4p16.3 containing Aurora-A kinase substrate TACC3 (Duncan et al. 2010). Grade-specific overexpression of TACC3 in gliomas, as well as a reported TACC3 somatic sequence mutation (p.E680K) in one of 22 glioblastomas (Parsons et al. 2008), provides evidence that overactive TACC3 function is relevant to glioblastoma biology. Mitotic kinases, such as the Aurora kinase family, and their substrates direct chromosome assembly and segregation during mitosis, and their dysfunction can lead to abnormal mitotic processes and aneuploidy (Marumoto et al. 2005). Together, these data support the role of compromised genome integrity as a central component of the pathogenesis of glioblastoma.

As understanding of the glioblastoma functional network matures, a universal question remains: How do we exploit genomic knowledge for the development of
effective targeted therapeutics? To answer this question, future therapeutic targeting strategies should take into account the complex and heterogeneous genetic patterns observed in glioblastomas, including activation of feedback loops, molecular cross-talk pathways, and drug-resistant mutations (Cloughesy and Mischel 2011). Continued detailed characterization of genomic alterations coupled with functional analyses will permit a greater understanding of the forces driving this deadly disease and will likely provide the basis for improved therapies.
2. Integrated genomic analyses identify *ERFFI1* and *TACC3* as glioblastoma-targeted genes

2.1 Background: Genomic analysis of glioblastoma

Cancer cells undergo continuous acquisition of heritable genetic variation, manifested as mutations of single base pairs, large or small deletions or insertions, chromosomal translocations, and gain or loss of entire chromosomal regions (Vogelstein and Kinzler 2004; Stratton et al. 2009). The accumulation of these somatic genomic changes results in a highly heterogeneous and complex cancer genome. Glioblastoma is the most aggressive and frequently occurring type of brain tumor (Reardon et al. 2006; Furnari et al. 2007). Recent comprehensive analyses have opened the door to understanding the genetic and molecular alterations that characterize the glioblastoma genome. These studies identified sequence mutations and alterations of DNA copy number, gene expression, and methylation status that may contribute to a distinct network of oncogenic signaling pathways, including receptor tyrosine kinase (RTK), growth factor and phosphatidylinositol-3-OH kinase (PI3K), p53 and retinoblastoma (RB1) pathways (Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008).

Identification of a coding sequence alteration, amplification, or homozygous deletion can help define unequivocal driver genetic alterations in cancers (The Cancer Genome Atlas Research Network 2008). Several genes in the central glioblastoma pathways were identified through dramatic copy number alterations, such as amplification of the oncogenes, *EGFR*, *c-MYC*, *CDK4*, *PDGFRA*, *MDM2*, and *MDM4*, and
deletion of the tumor suppressor genes, CDKN2A, CDKN2B, and PTEN (Jen et al. 1994; Eley et al. 2002; Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008; Verhaak et al. 2010). Recent studies have validated the significance of well-known copy number alterations and have proposed additional candidates which may contribute to the development of glioblastomas (Lo et al. 2007; Freire et al. 2008; Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008; de Tayrac et al. 2009). Conversely, several core glioblastoma genes have been found to be exclusively altered by sequence mutation, such as PI(3)K, RAS, ERBB2, and IDH1/IDH2 (Parsons et al. 2008; Yan et al. 2009b). However, a large proportion of recognized glioblastoma driver genes, including EGFR, TP53, CDKN2A, PTEN, NF1 and RB1, are targeted by both sequence and copy number alterations. In addition, gene expression signatures, which are an essential component of global genomic studies, have been used to exploit disease-specific signaling pathways and as clinical prognostic factors in many cancers (Bild et al. 2006a; Bild et al. 2006b; West et al. 2006). However, despite major advancements, the current understanding of glioblastoma genetics is still inadequate, and additional molecular targets are urgently needed to be used in the development of clinically proven therapies.

In this study, we used Digital Karyotyping (DK) and Illumina BeadChip assays (Illumina, Inc., San Diego, CA) to identify genomic loci that are recurrently targeted by focal copy number alterations in 111 glioblastomas. Using these data and data from TCGA, we identified frequent gene copy number changes in 1p36.23 and 4p16.3. We
further demonstrated that ERRFI1 and TACC3 in 1p36.23 and 4p16.3, respectively, are potential glioblastoma-targeted genes.

2.2 Methods

2.2.1 Tumor samples

DNA samples were obtained from brain tumor cell lines, xenografts and primary brain tumors. Brain tumor tissue samples were obtained from the Preston Robert Tisch Brain Tumor Center Biorepository at Duke University Medical Center by an IRB-approved protocol. All samples were obtained in accordance with the Health Insurance Portability and Accountability Act. Frozen sections were made from each tumor sample and examined by light microscopy by a board-certified neuropathologist to ensure that more than 95% of each section consisted of tumor cells. Normal patient DNA samples were obtained from peripheral blood. Normal adult and fetal brain genomic DNA was obtained from BioChain Institute, Inc. (Hayward, CA). Samples were defined as pediatric if taken from a patient between the ages of 0 and 19.

2.2.2 Digital karyotyping

We generated 18 glioblastoma DK libraries. Additionally, we included 1 primary and 8 glioblastoma cell line DK libraries from the Cancer Genome Anatomy Project (Rao et al. 2010) (http://cgap.nci.nih.gov/SAGE/DKViewHome). Protocols for performing DK and software for the extraction and analysis of genomic tags are available at www.digitalkaryotyping.org (Leary et al. 2007). Experimental tag sequences were
compared to predicted human genome virtual tags and were visualized by using SageGenie DKView (http://cgap.nci.nih.gov/SAGE/DKViewHome). Homozygous deletions were screened by using a sliding window size of 150 virtual tags (~600 kb in size). Putative homozygous deletions were defined as events with a tag density ratio (observed tags/expected tags in window) of < 0.05. Amplifications were identified using a sliding window size of 50 virtual tags (~200 kb in size). Putative amplifications were defined as events with a tag density ratio of > 6.

2.2.3 High density SNP-Arrays

Eighty-four glioblastoma samples were analyzed for copy number variation by utilizing HumanHap550-Duo and Human610-Quad BeadChips with the Illumina Infinium Whole Genome Genotyping Assay. Data was pre-processed to generate logR intensity ratios using Illumina BeadStudio. Data was converted into copy number calls and visualized by using Nexus Copy Number software (BioDiscovery, Inc., El Segundo, CA).

2.2.4 TCGA data

Genome-wide Level 3 TCGA copy number data was downloaded with the TCGA Data Portal Data Access Matrix (https://tcga-data.nci.nih.gov/tcga/dataAccessMatrix.htm) from Hudson-Alpha Cancer Genome Characterization Center, Memorial Sloan Kettering Cancer Center, and Harvard Medical School-Dana Farber Cancer Institute. Candidate deletions were defined as events with
seg mean ≤ -1. Candidate gains were defined as events with seg mean ≥ 0.4. Events less than 30kb and events occurring in two or more tumors with identical start and stop positions were removed, because they were likely artifact or copy number polymorphisms. Candidate gene TCGA expression data from the University of North Carolina and Broad Institute was downloaded with the TCGA Data Portal Data Browser. Gene expression values from the TCGA database represent the ratio of tumor expression to normal expression. The expression value for a given gene for a given patient is the log2 ratio of the tumor expression of the gene in the patient to a synthetic normal sample.

2.2.5 Bioinformatic analysis

Nexus Copy Number software (BioDiscovery, Inc) was used to visualize copy number data. The following values were used for screening of genetic events of homozygous loss and amplification: Significance Threshold, 1×10⁻⁸; Min probes per segment, 2; Max contiguous probe spacing (Kbp), 1000; Gain, 0.5; Loss, -0.7. For identification of areas of duplication, the Nexus default settings were as follows: Significance threshold, 1×10⁻⁶, Min probes per segment, 5; Max contiguous probe spacing (Kbp), 1000; Gain, 0.2.

2.2.6 Quantitative real-time PCR

The genomic DNA content and mRNA expression levels of genes of interest within tumor and normal cells were quantified by quantitative real-time polymerase
chain reaction (Q-PCR). Genomic DNA from normal blood cells served as controls, and genomic DNA content was normalized to that of Line-1. For mRNA expression measurement, cDNA from normal human adult cortex was used as the control and cDNA content was normalized to that of GAPDH.

2.2.7 Migration assays

H423 glioblastoma cells were maintained in ZO⁺ media (Gibco, Carlsbad, CA) with 10% fetal calf serum (FCS). For transwell migration assays, H423 cells were transfected with pCMV6-entry or pCMV6-ERRFI1, washed, and resuspended in serum-free media, then seeded into transwell membranes at 10,000 cells per well with 5% FCS as a chemoattractant. After 24 hours, cells were removed from the upper chamber then fixed in 100% methanol on the lower surface of the membrane and stained with crystal violet (2% solution in ethanol) and counted (experiment performed in triplicate wells, 12 fields counted per well). Statistical analyses were conducted with PASW (SPSS) 18.0 (IBM, Armonk, North Castle, NY). For wound healing assays, cells were transfected with pCMV6-entry or pCMV6-ERRFI1, and cells were plated at 100% confluence in 6-well plates. The cell monolayer was wounded with a p1000 pipette tip and wound healing was documented at 0, 6 and 24 hours by photographs.
2.3 Results

2.3.1 Detection of focal copy number alterations by DK and Illumina BeadChips

DK is a highly quantitative copy number analysis platform that has previously been used to identify copy number events in human cancers, including glioblastomas (Di et al. 2005; Lu et al. 2009; Rao et al. 2010). Analysis of 27 glioblastoma samples revealed 52 high-copy amplification events ranging from 98kb to 6.8Mb with 12 to 205 copies per nucleus. The targeted genes within those regions include gain of \textit{EGFR}, \textit{CDK4}, \textit{PDGFRA}, \textit{MDM2}, and \textit{MDM4}. In addition, we identified 120 regions of homozygous deletion, ranging from 100kb to 5.1Mb. The most common loss is on chromosome 9p21, where tumor suppressor genes \textit{CDKN2A} and \textit{CDKN2B} are located.

Illumina BeadChips in conjunction with the Infinium assay have also been effectively used to examine copy number variations in human cancers at high-throughput levels (Parsons et al. 2008; The Cancer Genome Atlas Research Network 2008; Lu et al. 2009). First, using the highly quantitative DK data, we optimized the criteria for defining focal high-copy amplifications and homozygous deletions for Illumina high-density SNP arrays. Two glioblastoma samples, xenograft H456 and primary tumor TB2607, were analyzed by both DK and Illumina high-density SNP arrays. Using DK as a standard, we identified the values and filtering criteria (stated in Materials and Methods) that faithfully reveal amplifications and deletions from the data produced by the Illumina BeadChips. We generated Illumina high-density SNP array
profiles from 84 glioblastomas samples. As controls, we also analyzed genomic DNA from normal adult and fetal brain and 3 matching blood specimens from glioblastoma patients. In total, we identified 474 focal gain events and 1540 focal loss events.

Of the genomic copy number profiles from a total of 111 glioblastoma samples assessed by DK or Illumina BeadChips, we identified a high degree of heterogeneity and copy number instability across the glioblastoma genomes (Figure 2A). Despite heterogeneity, we detected several regions which are recurrently gained or lost in glioblastomas (Figure 2B, C). The two most prevalent focal amplifications were the $\text{EGFR}$ and $\text{CDK4}$ loci, occurring in 42% and 12% of all cases, respectively. The two most prevalent focal deletions were the $\text{CDKN2A/B}$ and 1p36 loci, occurring in 40% and 9% of all samples, respectively. Additionally, we detected multiple intragenic homozygous deletions within large genes, including $\text{LRP1B}$, $\text{WWOX}$, and $\text{A2BP1}$. 
Figure 2: Focal high copy number gains and homozygous deletions in glioblastomas.

Copy number analysis using DK and Illumina BeadChips revealed gains (indicated by green) and losses (indicated by red) spanning the genome of glioblastomas. (A) Glioblastoma copy number karyotype generated utilizing Nexus Copy Number Professional Software (BioDiscovery Inc.). The most common regions for (B) focal deletion and (C) focal gain of copy number in 111 glioblastomas. Loci are identified with reported tumor suppressor genes and oncogenes if available.
2.3.2 ERRFI1 on 1p36 is a candidate tumor suppressor gene, whose products regulate glioblastoma cell migration

In addition to the well-characterized glioblastoma genes, the most striking genomic changes identified in our analysis were recurrent focal copy number changes on 1p36 and 4p16. Focal deletions of 1p36 have been previously reported to occur in numerous cancer types, including glioblastoma, neuroblastoma, oligodendroglioma, and colorectal, lung, and breast cancer (Di Vinci et al. 1998; Bieche et al. 1999; Ragnarsson et al. 1999). We detected focal homozygous deletions on 1p36 in 9/111 of our glioblastoma samples. With the addition of samples from the TCGA glioblastoma data set, we mapped 1p36 deletions in a comprehensive data set of a total 430 glioblastomas and found they occurred in two distinct minimal deleted regions (MDRs) on 1p36.32 (Figure 3) and on 1p36.23 (Figure 4).

Glioblastoma cell line H542 presented a homozygous deletion at 1p36.32 (Figure 3A). We mapped the 1p36.32 deletion in H542 by quantitative real-time PCR (Q-PCR) and found that this region contained TP73, KIAA0495, CCDC27, LOC388588, LRRC47, KIAA0562, DFFB, C1orf174, LOC100133612, and LOC284661 (Figure 3B). Further analysis of additional samples with 1p36.32 deletions revealed an overlapped MDR containing DFFB, C1orf174, and LOC100133612 (Figure 3B).
Figure 3: High-resolution mapping of homozygous deletions detects DFFB, C1orf174, and LOC100133612 within secondary MDR on 1p36.32.

(A) Chromosome 1 DK analysis of glioblastoma cell line H542 indicates a single homozygous deletion occurring at the 1p36.32 cytoband. (B) High-resolution mapping depicts overlapping homozygous deletions from 3.0Mb to 5.5Mb coordinates on chromosome 1 (indicated by red bars). Copy number data includes tumor samples from our independent analysis plus the TCGA glioblastoma copy number data set. RefSeq gene positions are indicated. Genomic Q-PCR validation analysis (*) of H542, H561, and glioblastoma xenograft X397 confirms homozygous deletion. Q-PCR primer location depicted by black boxes.
Figure 4: High-resolution mapping of homozygous deletions reveals \textit{ERRFI1} within the most frequent MDR on 1p36.23.
Figure 4 (continued): High-resolution mapping of homozygous deletions reveals ERRFI1 within the most frequent MDR on 1p36.23.

(A) DK analysis of chromosome 1 in glioblastoma cell lines H423 and H502 reveals a single homozygous deletion at 1p36.23. (B) High resolution mapping depicts overlapping homozygous deletions from 6.0Mb to 10.0Mb coordinates on chromosome 1 (indicated by red bars). Copy number data include tumor samples from our independent analysis and the TCGA glioblastoma copy number data set. RefSeq gene positions are indicated. Validation by genomic Q-PCR copy number analysis (*) of H423, H502, and glioblastoma xenograft X298 confirms homozygous deletion. Q-PCR primer location depicted by black boxes.

Cytoband 1p36.23 harbors the most frequently deleted region (Figure 4) on 1p36 in glioblastomas. Two glioblastoma cell lines, H423 and H502, each contained a single focal deletion in 1p36.23 (Figure 4A). We further mapped the 1p36.23 deletions in the two cell lines by Q-PCR. The deleted region contains multiple genes, including CAMTA1, VAMP3, PER3, UTS2, TNFRSF9, PARK7, ERRFI1, SLC45A1, RERE, ENO1, CA6, SLC2A7 and SLC2A5 (Figure 4B). With the addition of TCGA glioblastoma samples, we found that 15/430 glioblastomas contain homozygous deletions in 1p36.23. Further analysis revealed the most commonly deleted region contains a single gene, ERRFI1 (Figure 4B).

ERRFI1 is a candidate tumor suppressor which functions in normal cells as a negative regulator of EGFR and the ErbB family (Anastasi et al. 2003; Ballaro et al. 2005; Zhang et al. 2007a). We measured ERRFI1 mRNA levels in a panel of 62 glioblastoma samples and detected downregulation of ERRFI1 expression in 34% of the samples tested (Figure 5A).
Figure 5: ERRFI1 is silenced in glioblastomas and reduces cell migration in H423 glioblastoma cells.

(A) Quantification of ERRFI1 mRNA levels in glioblastomas (32 cell lines, 15 xenografts, and 15 primary tumors) by Q-PCR. Results presented as fold expression relative to normal cortex control (indicated by red line). (B) Restoration of ERRFI1 expression in H423 cells results in decreased cell migration as measured by trans-well migration assay (T-test, p<0.001).

We further investigated the possible pathogenic function of ERRFI1 in glioblastoma cells. Glioblastoma cell line H423 harbors a homozygous deletion of ERRFI1 and did not express endogenous ERRFI1 (Figure 5A). We transfected H423 cells with pCMV-6-entry-ERRFI1 and performed wound-healing experiments on the
transfected cells. Compared to cells transfected with a control vector, H423 cells with ERRFI1 expression had impaired wound healing (data not shown). Furthermore, we conducted trans-well assays and demonstrated that H423 cells with ERRFI1 expression had a lower rate of trans-well migration than control cells (Figure 5B).

2.3.3 **TACC3 on 4p16.3 is overexpressed and correlates with the expression of Aurora kinases in glioblastomas**

Focal gains on 4p16.3 were also frequently identified in a subset of glioblastomas (Figure 6A). We mapped the 4p16 duplications in a data set of 430 glioblastomas. The gains clustered in a minimal gained region centered on 4p16.3 from 1.5 to 2.0Mb (Figure 6B), containing **SLBP, TMEM129, TACC3, FGFR3** and **LETM1**, of which **TACC3** most commonly displayed gain of copy number. A Q-PCR analysis of an independent panel of glioblastoma samples detected genomic duplications of **TACC3** in 5 out of 101 samples. In addition, compared with its adjacent genes, **SLBP, TMEM129** and **FGFR3** at 4p16.3, we found that **TACC3** displayed a predominant overexpression pattern in glioblastomas (Figure 7A).

Dysregulation of the human Transforming Acidic Coiled Coil (TACC) genes is thought to be important in the development of several cancers. The three human TACC proteins, TACC1, TACC2, and TACC3, are core components of the centrosome and have non-overlapping functions in the normal cell and the cancer cell (Gergely et al. 2000; Peset and Vernos 2008). **TACC1** overexpression promotes cellular transformation *in vitro* (Still et al. 1999) and *in vivo* (Cully et al. 2005).
Figure 6: High resolution mapping identifies TACC3 as the glioblastoma-targeted gene on 4p16.3.

(A) Illumina BeadChip analysis of chromosome 4 in glioblastoma primary tumor TB3837 reveals a single focal gain at 4p16.3. (B) High resolution mapping depicts overlapping gain of copy number from 1.5Mb to 2.1Mb coordinates on chromosome 4 (indicated by green bars). Copy number data includes tumor samples from our independent analysis and the TCGA glioblastoma copy number data set. RefSeq gene positions are indicated.
Figure 7: TACC3 is the predominant gene upregulated in 4p16.3 and correlates with Aurora kinase expression.

SAGE of (A) TACC3 and FGFR3 and (B) Aurora kinases, AURKA, AURKB, and AURKC in normal brain (n=2) and glioblastoma (n=16) samples (data adapted from Parsons et al. 2008). (C) Gene expression of TACC and Aurora kinase family members using the TCGA glioblastoma expression data set (n=266) indicates correlation between TACC3 and AURKA (correlation coefficient = 0.76) and between TACC3 and AURKB (correlation coefficient = 0.70). Red indicates high gene expression level and green indicates low gene expression level.

TACC2 (AZU-1) is a putative tumor suppressor in breast cancer (Chen et al. 2000).

TACC3 expression is upregulated and associated with shorter median survival in patients with non-small-cell lung cancer (Jung et al. 2006). However, TACC3 expression is reduced in ovarian and thyroid cancer (Lauffart et al. 2005; Ulisse et al. 2007). We therefore assessed expression of the TACC genes in a panel of gliomas (WHO Grades I-
IV) by Q-PCR (Figure 8). We found significant overexpression of TACC3 in Grade IV gliomas. Conversely, TACC2 transcript levels were significantly downregulated in Grade IV gliomas.

![Figure 8: Grade-specific TACC3 upregulation and TACC2 downregulation in gliomas.](image)

(A) TACC3 and (B) TACC2 mRNA levels in a panel of 93 glioma samples were assessed by Q-PCR. Results presented as fold expression relative to normal cortex control. I (WHO classification glioma Grade I), II (WHO classification glioma Grade II), III (WHO classification glioma Grade III), IV (WHO classification glioma Grade IV).

TACC3 can be phosphorylated by Aurora-A kinase and plays an important role in mitosis (Barros et al. 2005; Kinoshita et al. 2005; LeRoy et al. 2007). Therefore, to determine the correlation of TACC3 and Aurora kinase expression, we used Serial Analysis of Gene Expression (SAGE) (Parsons et al. 2008) and TCGA expression data sets to evaluate TACC3 and Aurora kinase gene expression in glioblastomas (Figure 7B, 6C). We found a significant expression correlation between TACC3 and Aurora A.
(correlation coefficient = 0.76), as well as between TACC3 and Aurora B (correlation coefficient = 0.70) (Figure 7C).

2.4 Discussion of copy number alterations in glioblastoma

2.4.1 Deletions of 1p36

While many of the most common copy number gains and deletions have been characterized in glioblastomas, the significance of many other genomic events remains unknown. Utilizing an independent set of glioblastomas and publicly available data, in this study, we focused on the identification and characterization of genes on two frequently altered regions, 1p36 and 4p16. We found that ERRFI1 is a potential glioblastoma-targeted tumor suppressor gene and TACC3 is a potential oncogene.

Deletions in 1p36 have been previously reported in glioblastoma genomic studies, indicating that multiple tumor suppressors exist in this locus (Barbashina et al. 2005; Law et al. 2005; Ichimura et al. 2008; Yin et al. 2009; Rao et al. 2010; Ying et al. 2010). In our analysis, we observed several regions of deletions across 1p36. However, the most striking and frequent deletions occur between the 5 and 10Mb coordinates on 1p36.23, centering on the candidate tumor suppressor ERRFI1. Deletion of ERRFI1 has been reported to activate EGFR and sustain MAPK signaling, resulting in tumor phenotypes in numerous tissues in ERRFI1 knockout mice (Ferby et al. 2006; Jin et al. 2007; Zhang et al. 2007b). ERRFI1 is also frequently deleted, mutated, or down-regulated in breast and lung cancers, as well as in glioblastomas (Anastasi et al. 2005;
Ichimura et al. 2008; Ying et al. 2010). In a recent study, overexpression of \textit{ERRFI1} was shown to decrease proliferation in glioblastoma cells, binding EGFR with STX8, and driving internalized EGFR to late endosomes for degradation, whereas knockdown of \textit{ERRFI1} expression resulted in increased tumor invasion (Ying et al. 2010). Our biological data are consistent with these previous findings, showing that restoring \textit{ERRFI1} expression in an \textit{ERRFI1}-deficient glioblastoma cell line decreases glioblastoma cell migration. Our genetic and biological data, as well as data from other studies, suggests that \textit{ERRFI1} is another key component in the EGFR signaling pathway involved in glioblastoma development.

\textbf{2.4.2 Copy number gains of 4p16}

Genomic alterations of 4p16.3 have been reported in several cancers, indicating the presence of one or more oncogenes (Chesi et al. 1997; Stewart et al. 2004; Castro et al. 2009; Kiemeney et al. 2010). Using an integrative genomic strategy, we have analyzed genes altered in the 4p16.3 region in glioblastomas and revealed that \textit{TACC3} is the primary glioblastoma-targeted gene in this region. Furthermore, a \textit{TACC3} somatic mutation (p.E680K) was reported in one of 22 glioblastomas (Parsons et al. 2008). \textit{TACC3} has a conserved function to promote centrosomal microtubule assembly, a process which is often altered in cancer cells (Gergely et al. 2003; Peset and Vernos 2008). Consequently, genetic silencing of \textit{TACC3} results in destabilized microtubules, defects
in chromosome alignment, and mitotic defects (Piekorz et al. 2002; Gergely et al. 2003; Schneider et al. 2007).

The Aurora family of serine-threonine kinases is comprised of three members (A, B, and C) that cooperate with many other proteins, including the TACC family, to direct chromosome assembly and segregation during mitosis (Meraldi et al. 2004; Giet et al. 2005; Marumoto et al. 2005; Peset and Vernos 2008). Dysfunction of Aurora kinases can disrupt genomic integrity and lead to aneuploidy, mitotic arrest, and cell death. Aurora A and B are of particular interest since they have been shown to be overexpressed in a broad range of human tumors and are often associated with poor outcome. Thus, regulators of the mitotic spindle apparatus, including Aurora kinases and its substrates, are attractive targets for small-molecule therapeutics (Warner et al. 2003; LeRoy et al. 2007; Gautschi et al. 2008; Schneider et al. 2008). In our study, we found a strong correlation between Aurora kinase and TACC3 expression, indicating that in the Aurora kinase/TACC pathway, the dysregulated kinase and its substrate, may contribute synergetically to glioblastoma pathogenesis and could serve as targets for molecular-based intervention.
3. A heterozygous $IDH1^{R132H/WT}$ mutation induces genome-wide alterations in DNA methylation

3.1 Background: IDH mutations and the epigenome

Mutations of the NADP+-dependent isocitrate dehydrogenase ($IDH$) genes $IDH1$ and $IDH2$ occur in more than 70% of Grade II-III gliomas and secondary glioblastomas (sGBM) (Balss et al. 2008; Parsons et al. 2008; Bleeker et al. 2009; Hartmann et al. 2009; Yan et al. 2009a; Yan et al. 2009b; Gravendeel et al. 2010), 15%-30% of acute myeloid leukemias (AMLs) (Mardis et al. 2009; Marcucci et al. 2010; Paschka et al. 2010; Wagner et al. 2010; Ward et al. 2010), 56% of chondrosarcomas (Amary et al. 2011), 87% of enchondromas, 70% of spindle cell hemangiomas (Pansuriya et al. 2011), 22%-28% of cholangiocarcinomas of intrahepatic origin (Borger et al. 2012; Kipp et al. 2012), and at lower frequencies in other malignancies, including colorectal cancer (Sjoblom et al. 2006), prostate carcinoma, and B-acute lymphoblastic leukemia (B-ALL) (Kang et al. 2009). $IDH$ mutations occur early in tumor development and may either cause or predispose cells to become malignant (Ichimura et al. 2009; Watanabe et al. 2009). In human tumors, $IDH$ mutations give rise to single amino acid substitutions at specific conserved residues, arginine 132 (R132) of IDH1 and arginine 140 (R140) or arginine 172 (R172) of IDH2. These $IDH$ point mutations primarily occur as somatically-acquired heterozygous events, with tumor cells showing one mutant allele and retention of the second wild-type allele, suggesting that the ratio of mutant to wild-type enzyme may be critical to its oncogenic activity. The $IDH$ mutation leads to an attenuation of the normal
catalytic activity, the oxidative decarboxylation of isocitrate to alpha-ketoglutarate (α-KG) (Yan et al. 2009b; Zhao et al. 2009), while concurrently imparting a gain of novel enzymatic function, wherein α-KG is reduced, leading to the aberrant accumulation of the onco-metabolite D-2-hydroxyglutarate (D-2-HG) (Dang et al. 2009). Both reduction of cellular α-KG levels and accumulation of D-2-HG have the potential to contribute to altered cellular phenotypes through the inhibition of multiple Fe(II)/2-oxoglutarate-dependent dioxygenases (Xu et al. 2011a), a superfamily of enzymes involved in a wide range of biological functions, including DNA repair and chromatin modification, such as the AlkB family of oxidative demethylases, the Jumonji-C domain family of histone demethylases (JHDMs), and the TET family of methylcytosine hydroxylases (Loenarz and Schofield 2008; Tahiliani et al. 2009; Figueroa et al. 2010a; Chowdhury et al. 2011).

Recent large-scale studies of DNA methylation distribution in primary tumors have identified biologically distinct subgroups of glioblastomas (GBMs) and AMLs associated with aberrant DNA methylation (Figueroa et al. 2010b; Noushmehr et al. 2010). A subset of primary GBMs exhibit the ‘CpG island methylator phenotype’ (CIMP) and show concordant hypermethylation of a large number of CpG islands (Toyota et al. 1999; Toyota and Issa 1999; Noushmehr et al. 2010). Interestingly, in gliomas, the CIMP phenotype (termed G-CIMP) has a striking association with IDH1-mutation (Noushmehr et al. 2010; Christensen et al. 2011; Laffaire et al. 2011; Turcan et al. 2012). Similarly, IDH1 and IDH2 mutations robustly associate with specific global
DNA hypermethylation phenotypes in AMLs (Figueroa et al. 2010a), enchondromas (Pansuriya et al. 2011), and low grade gliomas (LGGs) (Turcan et al. 2012). The evidence linking IDH1 and IDH2 mutations with distinct DNA methylation phenotypes in primary human tumors raises the question of whether these mutations can drive oncogenesis through epigenetic reprogramming of cancer cells. Such epigenetic changes, including DNA hypermethylation and hypomethylation, can play fundamental roles in the initiation and progression of human cancer through regulation of gene expression (Jones and Baylin 2002; Jones and Baylin 2007). In addition to epigenetic classifications, gene expression-based molecular classification systems have been developed in GBM to distinguish clinically relevant molecular subclasses (Phillips et al. 2006; Li et al. 2009; Verhaak et al. 2010). GBMs with IDH1 mutations were shown to primarily exhibit a proneural gene expression signature (Verhaak et al. 2010), although specific gene expression alterations resulting from mutant IDH1 proteins have not been elucidated.

Despite the apparent correlation between IDH mutations and CpG island hypermethylation, the contribution of heterozygous expression of IDH1 point mutations to this phenotype has not been addressed. One limitation of the field has been the lack of model systems that recapitulate naturally occurring monoallelic point mutations observed in human tumors. Previous studies have relied on ectopic overexpression in human and mouse cell lines to study the effects of mutant IDH proteins (Dang et al.
To address this issue, and to faithfully recapitulate the naturally-occurring genetic alterations, we utilized gene targeting to introduce heterozygous \( IDH1_{R132H/WT} \) substitutions in the human colorectal cancer cell line, HCT116. We determined the impact of this mutation on the genome-wide distribution of CpG methylation in the isogenic \( IDH1_{R132H/WT} \) cells and identified differentially methylated CpGs induced by the \( IDH1_{R132H/WT} \) mutation, which include both hyper- and hypomethylation events. Our data indicate that \( IDH1_{R132H/WT} \) mutations play a causal role in the widespread alteration of DNA and histone methylation observed in human cancers and can impact gene expression.

### 3.2 Methods

#### 3.2.1 Cell culture and drug treatment

HCT116 cell lines were cultured in McCoy's 5A Medium supplemented with 10% FBS. HCT116 parental cells were a generous gift from Dr. Bert Vogelstein (The Johns Hopkins University). Human oligodendroglioma (HOG) cell lines expressing mutant IDH1 or empty vector were generated as described previously (Reitman et al. 2011). All cells were maintained in 5% CO\(_2\) at 37°C. For reactivation studies, parental or \( IDH1_{R132H/WT} \) knock-in cells (1x10\(^6\)) were plated in a 25-cm\(^2\) flask 24 hours before treatment with 5 \( \mu \)M 5-aza-2'-deoxycytidine (DAC) (Sigma) or control (equal volume PBS). Medium containing fresh DAC was applied every 24 hours for a total of 48 hours.
Following treatment, cells were harvested, and RNA and genomic DNA were isolated for analysis.

### 3.2.2 Gene targeting of the human IDH1 locus

The approach for generating genetic knock-ins in human somatic cells was performed as previously described (Rago et al. 2007). Targeting constructs were designed to introduce the IDH1^{R132H} allele in HCT116 utilizing the pSEPT rAAV shuttle vector (Topaloglu et al. 2005). Homology arms for the targeting vector were PCR-amplified from HCT116 genomic DNA using Platinum Taq HiFi polymerase (Invitrogen). The R132H hotspot mutation was introduced in the targeting construct by using the Quickchange II site-directed mutagenesis kit (Stratagene). An infectious rAAV stock harboring the targeting sequence was generated and applied to the parental cell line as previously described (Kohli et al. 2004), and clones were selected in 0.5 mg/ml Geneticin (Invitrogen). Next, excision of the selectable element was induced with Cre recombinase (Vector Biolabs). Targeted homologous recombination and Cre-mediated excision were verified by using PCR-based assays. Genomic DNA and total RNA were isolated from cells with a QIAmp DNA Blood Kit and RNeasy Kit (Qiagen). First-strand cDNA was synthesized by using an iScript cDNA Synthesis Kit (BioRad). Genotypes of identified clones were validated and determined to have equal expression of mutant and wild-type IDH1 by PCR and direct sequencing of genomic DNA and cDNA.
3.2.3 D-2-HG analysis

Quantification of D-2-HG in cell culture media was performed by LC-negative electrospray ionization-MS/MS at the Duke University Cancer Center Clinical Pharmacology Lab as described previously (Struys et al. 2004; Jin et al. 2011a).

Intracellular D-2-HG levels were normalized to total protein levels. Media above cells was collected 48 hours after plating for analysis of D-2-HG concentration.

3.2.4 Genome-wide CpG methylation profiling

Genomic DNA was extracted by using a DNeasy Blood & Tissue Kit (Qiagen), and DNA was bisulfite converted by using the EZ DNA Methylation Kit (Zymo). DNA methylation was profiled using the Illumina HumanMethylation27 BeadChip platform which interrogates 27,578 CpG sites from 14,475 protein-coding genes and 110 microRNAs. This platform measures the methylation status of single CpG sites after bisulfite modification of DNA. For each interrogated CpG there are two bead (oligonucleotide) types: one corresponding to the methylated (C) sequence and the other to the unmethylated (T) state. Annealing of the bisulfite converted sample DNA is followed by single-base primer extension using DNP- and Biotin-labeled ddNTPs, followed by staining and scanning. Methylation levels are returned as a beta (β) value that ranges from 0 to 1 (where 1 approximates 100% methylation) and is calculated as the ratio of the fluorescence signal for the methylated (M) probe to the total signal for both probes (U+M) at each CpG site. Two independent clonal variants from each of two
unique recombinant knock-in clones were assayed in duplicate in addition to four biological replicates of the parental cells. Methylation arrays were processed at the Duke University Center for Human Genetics according to the manufacturer’s instructions and scanned on an Illumina iScan. Data were interpreted in the Methylation Module of GenomeStudio (v2010.3) software. Only those probes with a detection $P$-value of less than 0.01 in all samples were considered in subsequent analyses.

3.2.5 Statistical analysis of DNA methylation

All differential and statistical analyses were performed in R / Bioconductor (http://www.r-project.org/). Differential analyses were conducted with fixed-effects linear models implemented by the function ‘lm’ of the ‘stats’ package. Clonal analysis used a defined variable for clone to explain differences in DNA methylation level ($\beta$) for each CpG locus $i$ and clone $j$ where clone $j \epsilon$ (HCT116 parent, $IDH1^{R132H/WT}$ knock-in 1, $IDH1^{R132H/WT}$ knock-in 2), such that $\beta_i \sim$ clone $j$. This produced a $t$-statistic and $P$-value for each CpG locus $i$ and each knock-in compared to the wild-type HCT116 parent. $P$-values were corrected for multiple hypothesis testing by using the Benjamini-Hochberg False Discovery Rate (FDR) implemented by the ‘p.adjust’ function also of the ‘stats’ package in R where an FDR < 0.01 was considered significant. $IDH1^{R132H/WT}$ differential analysis used the same methodology as the clonal analysis except instead of a variable to define clone, a variable that defined $IDH1^{R132H/WT}$ genotype was used to explain the methylation
level ($\beta$) for each CpG locus $i$, thus combining the clones. Similarly, this produced a $t$-statistic and a $P$-value ($P$) for the effect of the $IDH1^{R132H}$ mutation on DNA methylation level for each locus $i$. Benjamini-Hochberg FDR was used for multiple hypothesis testing correction. The Q-Q plot was generated using the function ‘qqnorm’ of the ‘stats’ package. Hierarchical clustering for DNA methylation data was performed using the ‘heatmap.2’ function of the ‘gplots’ package, where clustering was performed using an average agglomerative algorithm and a Euclidean distance dissimilarity metric. Data were not normalized for DNA methylation hierarchical clustering or display. The significance of the 2,852 identified HCT116 differentially methylated loci in discriminating $IDH1$ mutant and wild-type TCGA GBM samples was tested by permutation analysis. The average Euclidean distance of $\beta$ between the $IDH1$ wild-type and mutant (or G-CIMP+ and G-CIMP−) samples was calculated for the 2,852 HCT116 differential loci and then compared to that achieved using 2,852 randomly chosen loci. This process was repeated 1,000 times. In no permutation was the distance between the two groups greater than the distance achieved using the actual differential loci, suggesting that the HCT116 differential loci have a greater ability to segregate $IDH1$ wild-type and mutant TCGA GBMs (Noushmehr et al. 2010) than a set of arbitrary loci. This same methodology was applied to the LGG set from Turcan et al. (Turcan et al. 2012) except that the G-CIMP positive and negative groups were compared only at those HCT116 differential loci common to both platforms ($n = 2,722$). Relative DNA
methylation distribution plots were generated using the ‘density’ function of the ‘stats’ package, which computes the probability density function of the \( \beta \) distribution (i.e. describing the relative likelihood of \( \beta \) to occur at a given value such that the area under the curve is equal to 1). Box-and-whisker plots were created using the ‘boxplot’ function of the ‘stats’ package using the default settings. Boxes represent the first to third quartiles of the data distribution and whiskers were drawn to the maximum data value no more than 1.5 times the interquartile distance (i.e. difference between the first and third quartiles). \( P \)-values for the box-and-whisker plots were calculated using the Mann-Whitney U-test implemented by the ‘wilcox.test’ of the ‘stats’ package and are two-sided when comparing the methylation level of HCT116 \( IDH1^{R132H/WT} \) hypermethylated and hypomethylated loci and one-sided when comparing either the hypermethylated or hypomethylated loci with the relevant \( IDH1 \) genotype comparison (e.g. \( IDH1^{WT} \) or \( IDH1^{mut} \)). Stripcharts were created with the ‘stripchart’ function of the ‘graphics’ package, and the horizontal solid black lines represent the median of each comparison. \( P \)-values were calculated by using Welch’s two-sided \( t \)-test. The spatial distribution of methylation probes relative to TSS and CpG island were calculated by mapping the location of each probe relative to the closest TSS (hg18 UCSC RefSeq genes; http://hgdownload.cse.ucsc.edu/goldenPath/hg18/database/RefFlat.txt.gz) or CpG island (hg18 Takai-Jones criteria (Takai and Jones 2002)). Distribution was then plotted using the ‘density’ function in R. Gene ontology analysis of the HCT116 hypermethylated and
hypomethylated loci was conducted using the GOstats package (Falcon and Gentleman 2007), where overrepresented biological processes were determined using a significance threshold of $P < 0.01$.

**3.2.6 Genome-wide expression profiling**

To profile gene expression levels, we utilized Affymetrix Human Genome U133A 2.0 Arrays. For expression analyses, $2 \times 10^6$ cells were plated to a 25-cm$^2$ flask and harvested after 48 hours for RNA isolation by using the RNeasy Kit (Qiagen), with array measurements performed on biological duplicates. The RNA samples were labeled using the Ambion MessageAmp Premier Package and interrogated with the U133A 2.0 Array at the Duke Microarray Facility.

**3.2.7 Statistical analysis of gene expression**

Data were loaded in Expression Console, RMA normalized, and exported for analysis in R / Bioconductor (http://www.r-project.org/). Data were filtered to remove non-expressed probes (i.e. probes where no sample had a MAS5 detection $P$-value < 0.01). In total, of the 22,215 probes on the Affymetrix platform, 10,796 were detected in one or more samples, and this set was used for further analysis. Differential analysis of the gene expression data used the same strategy and similar methods as employed for the DNA methylation analysis. Here, $IDH1^{R132H/WT}$ genotype was used to explain the gene expression level ($GX_i$) for each probe $i$, such that $GX_i \sim IDH1^{R132H/WT}$ genotype. This resulted in a $t$-statistic and $P$-value for each probe $i$ being differentially expressed in
HCT116 parental and \( IDH1^{R132H/WT} \) knock-in clones. Additionally, a fold-change criterion was imposed on the differential analysis, where \( P \)-values for probes with a fold-change greater than 1.5 were corrected for multiple hypothesis testing using Benjamini-Hochberg FDR. This resulted in 21 probes from 18 genes significantly altered by \( IDH1^{R132H/WT} \) knock-in (FDR < 0.05, fold-change > 1.5). Hierarchical clustering was done using the ‘heatmap.2’ function of the ‘gplots’ package, where data were normalized by probe and an average clustering algorithm was used with a Euclidean distance dissimilarity metric applied to the normalized data. Correlation of gene expression with DNA methylation data was evaluated using the average fold-change between \( IDH1^{WT} \) and \( IDH1^{R132H/WT} \) knock-in expression compared to the average change in DNA methylation (\( \beta \)) for the same gene. For this analysis, those genes with multiple gene expression or DNA methylation probes were summarized to one value by averaging the multiple probes and then computing the fold-change or DNA methylation difference between \( IDH1^{WT} \) and \( IDH1^{R132H/WT} \) knock-in groups. Correlation was computed using Spearman’s rank correlation (\( \rho \)), and significance of the correlation was assessed using permutation testing where the DNA methylation changes and gene expression fold-changes were randomly permuted 1,000 times and Spearman’s rank correlation was calculated for each permutation. The \( P \)-value shown in Figure 17 reflects the number of times the permuted correlation (\( \rho \)) was larger than the actual correlation. Gene set enrichment analysis (GSEA) was performed with GSEA software (v2.0) and the
Molecular Signatures Database (MsigDB v3.0) (Subramanian et al. 2005). The HCT116 gene expression data was analyzed with a signal-to-noise ratio and 1,000 gene set permutations against the curated MsigDB.

3.2.8 Genomic bisulfite sequencing analysis

To validate genome-wide methylation data, we employed bisulfite sequencing techniques to analyze CpG-rich regions of gene promoter DNA. Primers for sequencing bisulfite-modified DNA were designed to avoid CpG residues using MethPrimer software such that unmethylated and methylated sequences are equally amplified (Li and Dahiya 2002). PCR products were TA cloned (Invitrogen) and transformed into chemically competent Escherichia coli, and plasmid DNA isolated from 7-10 individual colonies was sequenced.

3.2.9 Histone extraction and western analysis

Histones were acid-extracted as outlined in the Abcam (Cambridge, MA) histone extraction protocol. Briefly, HCT116, KI-1 or KI-2 cells were harvested, washed twice with cold PBS, resuspended in cold Triton Extraction Buffer (TEB: 0.5% Triton X-100, 2 mM PMSF, 0.02% NaN₃ in PBS) and lysed on ice for 10 minutes. Nuclei were pelleted by centrifugation at 6500 g for 10 minutes at 4°C, washed once in 0.5 volumes of TEB and incubated overnight in 0.2 N HCl at 4°C to extract histones. Samples were centrifuged at 6500g, and protein concentration in the supernatant determined using the Bradford assay. For western blot analysis, 5 μg of the histone lysates were electrophoresed on 15%
SDS-PAGE gel, transferred to a 0.2 mm nitrocellulose membrane, and incubated with antibodies against histone H3 (Abcam; 1791), histone H4 (Millipore; 05-858), H3K27me3 (Cell Signaling; C36B11), H3K4me3 (Millipore; 07-473), H3K9me3 (Active Motif; 39161), and H4K20me3 (Abcam; 9053), followed by HRP-conjugated secondary and chemiluminescence detection.

### 3.2.10 Chromatin immunoprecipitation

ChIP with antibodies against histone H3 and H3 modifications were carried out essentially as described in the acetyl-histone H3 immunoprecipitation assay kit from Millipore (17-229) and as previously described (Kapoor-Vazirani et al. 2008), with the exception that 20 µg of sonicated chromatin in Lysis Buffer (1% SDS, 10mM EDTA, 50mM Tris, pH 8, 1x Protease inhibitor) was diluted 10-fold in ChIP dilution buffer (0.01% SDS, 1.0% Triton X-100, 2mM EDTA, 20mM Tris pH 8.1, 150mM NaCl, 1× Protease inhibitor) prior to preclearing and immunoprecipitation. A portion of each sonicated chromatin sample was retained and processed as input DNA and to determine sonication efficiency. Immunoprecipitated DNA was analyzed by real-time PCR using primers specific to the *RBP1, UBB, SERPINB5* and *PDLIM2* loci as previously described (Kapoor-Vazirani et al. 2008). Starting quantities of immunoprecipitated and input DNA were determined relative to a standard curve generated with MCF7 genomic DNA. Antibodies used were: pan-H3 (Abcam, ab1791), H3K9me3 (Abcam, ab8898), H3K27me3 (Cell Signaling, 9733s), and IgG (Santa Cruz, sc-2027).
3.2.11 Quantitative real-time reverse transcriptase PCR

The mRNA expression levels for genes of interest were determined by quantitative real-time polymerase chain reaction (Q-PCR). For mRNA expression measurement, cDNA content was normalized to that of GAPDH as an internal standard, and fold change was calculated relative to unmodified parental cell controls.

3.2.12 Analysis of primary glioblastomas and low grade gliomas

TCGA genome-wide level 2 DNA methylation data as well as tumor G-CIMP classification were downloaded from the TCGA Data Portal (http://tcga-data.nci.nih.gov/tcga/). Affymetrix gene expression data were downloaded from the TCGA publications site (http://tcga-data.nci.nih.gov/docs/publications/gbm_exp/ - see Broad202.txt file) (Verhaak et al. 2010). Gene expression values from the TCGA study were log2 transformed to make them comparable to other data. IDH1 status and gene expression subtype were obtained from Table S7 of Verhaak et al. (Verhaak et al. 2010) and were further supplemented with data available from the TCGA web site. Analyses of changes in gene expression and DNA methylation for specific genes in IDH1WT and IDH1mut tumors utilized only those samples where a definitive IDH1 genotype was available in the TCGA data (n = 61 for Infinium DNA methylation data; n = 117 for Affymetrix gene expression). Data for LGGs were obtained from Turcan et al. (Turcan et al. 2012). Gene expression (Affymetrix HG 133 2) and DNA methylation (Illumina HumanMethylation450) data were downloaded from gene expression omnibus (GEO:
GSE30339). These data sets include DNA methylation for 81 glioma samples (49 G-CIMP+ and 32 G-CIMP−) and gene expression for a subset of 52 glioma samples (36 G-CIMP+ and 16 G-CIMP−).

### 3.2.13 Analysis of HCT116 ChIP-seq data

ChIP-seq data for histone 3 lysine 4 trimethylation (H3K4me3) and RNA polymerase II (Pol II) in wild-type HCT116 cells were downloaded from the UCSC ENCODE Project website (http://genome.ucsc.edu/ENCODE/downloads.html). Specifically, the H3K4me3 data was generated by the University of Washington and the RNA Pol II data was generated by Yale University. Reads were mapped to hg18 using Bowtie (v0.12.7) (Langmead et al. 2009) with the following settings: -t -p 6 -m 1. This resulted in 11,346,686 of 18,222,198 (62.2%) of the Pol II and 21,340,297 of 25,929,878 (82.3%) of the H3K4me3 reads uniquely mapping to the hg18 genome. Data was analyzed for enrichment around the Infinium HumanMethylation27 CpG interrogated loci using Hypergeometric Optimization of Motif Enrichment (HOMER v3.10; http://biowhat.ucsd.edu/homer/index.html) (Heinz et al. 2010). Data were displayed using the ‘image’ function in R.
3.3 Results

3.3.1 Targeted knock-in of \( IDH1^{R132H/WT} \) hotspot mutation in a human cancer cell line

To analyze the epigenetic effects of heterozygous \( IDH1^{R132H/WT} \) point mutations, we established isogenic cell lines in which one wild-type allele of this gene was replaced with the \( IDH1^{R132H} \) allele. HCT116 was selected for \( IDH1^{R132H/WT} \) knock-in because the cell line is diploid at the \( IDH1 \) locus and is susceptible to gene targeting through homologous recombination (Shirasawa et al. 1993; Waldman et al. 1995). Additionally, HCT116 cells have the intact machinery to exhibit physiologically relevant methylation patterns and have been utilized to examine DNA methylation phenotypes resulting from genetic ablation of DNA methyltransferases (DNMTs) (Rhee et al. 2000; Rhee et al. 2002). We utilized gene targeting by homologous recombination with recombinant adeno-associated virus (rAAV) constructs (Kohli et al. 2004; Topaloglu et al. 2005; Rago et al. 2007) to replace one \( IDH1 \) allele with an \( IDH1^{R132H} \) allele in the HCT116 cell line (Figure 9A). Clones that underwent homologous recombination at the \( IDH1 \) locus were verified using a PCR-based approach (Figure 10A). Cre-mediated excision of the selection cassette was confirmed by both PCR-based and functional assays (Figure 10B). Two independent clones were obtained for further detailed analysis, \( IDH1^{R132H/WT} \) knock-in 1 (KI-1) and \( IDH1^{R132H/WT} \) knock-in 2 (KI-2). Resulting clones were sequenced to confirm presence of the heterozygous G>A \( IDH1 \) mutation (Figure 9B). We also sequenced the cDNA from each knock-in clone to confirm equal expression of
Figure 9: Targeted knock-in of $IDH1^{R132H/WT}$ hotspot mutation in a human cancer cell line.
Figure 9 (continued): Targeted knock-in of $IDH1^{R132H/WT}$ hotspot mutation in a human cancer cell line.

(A) To faithfully recapitulate expression of heterozygous $IDH1^{R132H/WT}$ mutations as observed in human tumors, a targeting vector was designed to introduce the $IDH1^{R132H}$ mutation in one endogenous allele of $IDH1$ in HCT116. Relative genomic positions of exons are indicated, including 5’ UTR (white boxes) and coding sequences (black boxes). Homology arms (HAs) were cloned from HCT116 parental cells and are shown in red. The left HA (LHA) was altered by site-directed mutagenesis to create the $IDH1^{R132H}$ mutation (indicated by yellow star). The homology arms flank a synthetic exon promoter trap (SEPT) cassette. The promoterless SEPT element contains a splice acceptor (SA), internal ribosomal entry sequence (IRES), neomycin selectable marker (neo), and polyadenylation site (pA), which are flanked by loxP sites (green triangles). Inverted terminal repeats (ITR) of the adeno-associated virus (AAV) vector flank the homology arms. Correctly targeted alleles result in incorporation of the SEPT cassette along with the R132H mutation. Targeted clones were infected with Cre adenovirus to excise the selectable element, generating a clone which differs from the parental cell line by the single base pair mutation in exon 4 and 34 base pair loxP scar in the adjacent intron. (B) Sequencing validation of $IDH1^{R132H/WT}$ knock-in clones. Representative sequencing chromatograms for $IDH1$ codons 131-133 in genomic DNA (left) and cDNA (right) of HCT116 parental cells (top) and $IDH1$ knock-in cells (bottom). Knock-in clones contain a heterozygous G>AG mutation at chr2:208,938,618 and are heterozygous for wild-type allele (CGT) and mutant allele (CAT) coding for an arginine (R) to histidine (H) change at amino acid 132. Measurement of D-2-hydroxyglutarate (D-2-HG) in (C) clarified cell lysate and (D) cell culture medium over cells, collected after 48 hour incubation. Shown is the mean ± standard deviation of triplicate measurements.
Figure 10: Diagnostic PCR screens for generation of knock-in clones.
Figure 10 (Continued): Diagnostic PCR screens for generation of knock-in clones.

(A) Diagnostic PCR screen for homologous integration. The screening approach employs one primer that anneals within the SEPT element and a second primer that is outside the homology region of the targeting vector. Diagnostic PCR primer pairs are indicated as P1-P5. The five primer sets confirm homologous integration for clones KI-1 and KI-2. (B) Diagnostic PCR screen for excision of selectable element. Recombinant clones were infected with Cre adenovirus, diluted to single cell, and analyzed by diagnostic PCR. Diagnostic PCR primer pairs are indicated as P1 and P2. P1 employs one primer that anneals within the SEPT element and a second primer that is outside the homology region of the targeting vector, detecting alleles with intact selectable elements. P2 employs one primer that anneals outside the homology region and a second that anneals beyond the second loxP site, amplifying both WT and targeted alleles, differing by the 34bp loxP sequence. Representative excised and non-excised clones are depicted. In addition, complete excision was confirmed through analysis of sensitivity to G418.
the wild-type (R\textsuperscript{132}) and mutant (H\textsuperscript{132}) alleles (Figure 9B). Quantification of intracellular D-2-HG showed that the levels of D-2-HG were more than 100 fold higher in clarified cell lysates from the knock-in clones KI-1 ($P = 0.023$) and KI-2 ($P = 0.020$) as compared to parental cells, confirming neomorphic IDH1\textsuperscript{R132H} enzymatic activity (Figure 9C).

Additionally, IDH1\textsuperscript{R132H/WT} knock-in cells readily excrete D-2-HG to their surrounding environment, as concentrations of D-2-HG in the cell culture media from both KI-1 ($P < 0.01$) and KI-2 ($P < 0.01$) were more than 100 fold higher than that in culture media from parental cells (Figure 9D).

### 3.3.2 IDH1\textsuperscript{R132H/WT} induces alterations in DNA methylation

We next examined the effects of IDH1\textsuperscript{R132H/WT} on DNA methylation patterns using Infinium Methylation27 profiling (Illumina, Inc., San Diego, CA). This platform uses a single nucleotide extension approach on bisulfite modified DNA to interrogate the methylation status of 27,578 CpG sites spanning 14,475 genes. The relative methylation levels are returned as a “$\beta$-value” (0 to 1) calculated from the ratio of signal intensity from the methylated probe to the total signal from both unmethylated and methylated probes for each CpG site, approximating the percent methylation. Two replicate samples of two clonal variants derived from both the independent knock-in clones KI-1 and KI-2 were analyzed as compared to four wild-type IDH1 parental cell line replicates. Comparison of the relative methylation ($\beta$) distribution of all assayed CpG sites in KI-1, KI-2, and parental cells revealed a consistent, significant shift in the relative DNA
Figure 11: $\text{IDH}^1_{\text{R}132\text{H/WT}}$-induced DNA methylation alterations in HCT116 cells.
Figure 11 (continued): \(IDH1^{R132H/WT}\)-induced DNA methylation alterations in HCT116 cells.

HCT116-\(IDH1^{R132H/WT}\) knock-in clones and parental cells were analyzed using the Illumina HumanMethylation27 assay. (A) Relative DNA methylation (\(\beta\)) distribution for \(IDH1^{WT}\) HCT116 parental cells (blue) and \(IDH1^{R132H/WT}\) knock-in clones (light and dark gray). Frequency (y-axis) is plotted by \(\beta\) such that the total probability (area under the curve) is equal to one. Both knock-in clones show an increase in methylated loci as compared to the wild-type parental cells. (B) Hierarchical clustering of HCT116 samples using \(IDH1^{R132H/WT}\) differential loci. Samples are represented by columns and differential CpG loci by rows. Samples are annotated by \(IDH1\) genotype where the wild-type parental cells are in blue and the \(IDH1^{R132H/WT}\) knock-ins (KI-1 and KI-2) are in light and dark gray. CpG loci are annotated by their differential methylation, where red is hypermethylated (FDR < 0.01, \(n = 2,010\)) and green is hypomethylated (\(n = 842\)) in \(IDH1^{R132H/WT}\) knock-ins as compared to the wild-type parental line. The color of the heat map represents \(\beta\), where unmethylated is white (\(\beta = 0\)), partially methylated is burgundy (\(\beta = 0.5\)), and fully methylated is black (\(\beta = 1\)). Clustering is performed with an average clustering algorithm and Euclidean distance dissimilarity metric. (C) Relative DNA methylation distribution for HCT116 \(IDH1^{R132H/WT}\) differentially methylated loci. Frequency of the differentially methylated loci are plotted relative to \(\beta\) for hypomethylated (green) and hypermethylated loci (red) in wild-type parental cells (WT: solid line) and \(IDH1^{R132H/WT}\) knock-in cells (\(IDH1^{R132H}\); dashed line), such that the total probability (area under any given curve) is equal to one. (D) Box-and-whisker plot of DNA methylation levels for hypomethylated and hypermethylated loci in wild-type (WT) and \(IDH1^{R132H/WT}\) knock-in cells (R132H). Loci that were hypermethylated have a higher methylation level in parental cells than loci that were hypomethylated (\(P < 10^{-300}\); Mann-Whitney U test).
methylation (β) density distribution upon IDH1\textsuperscript{R132H/WT} knock-in (Figure 11A). Most notably, IDH1\textsuperscript{R132H/WT} knock-in cells showed an increased frequency of CpG sites with high level of methylation (β ~ 0.8-1.0). Unsupervised hierarchical cluster analysis showed that the wild-type and IDH1\textsuperscript{R132H/WT} knock-in cells separate by genotype based on their overall methylation patterns (Figure 12A). Differential methylation analyses were performed comparing each of the two independent knock-in clones to the wild-type parental cells. These analyses yielded concordant results wherein the majority of differentially methylated loci associated with IDH1\textsuperscript{R132H/WT} in either clone was common to the other, demonstrating a largely similar effect on methylation resulting from two independent recombination events (Figure 12B,C). We combined the data from all replicates of both knock-in clones and wild-type parental cells in a cumulative analysis and identified 2,852 loci that were significantly differentially methylated between knock-in and wild-type cells (Figure 11B). Of these, 2,010 and 842 loci were significantly hyper- or hypomethylated, respectively, in the knock-in clones relative to wild-type cells (FDR < 0.01; Figure 12D). Gene ontology analysis of unique genes represented by the 2,010 hypermethylated CpG sites showed enrichment of functional categories surrounding intracellular signaling and response to extracellular stimuli. Similar ontologies were also found to be significantly enriched among CpG loci previously published as hypermethylated in G-CIMP positive (G-CIMP+) GBMs (Noushmehr et al. 2010) and
Figure 12: Consistent $IDH1^{R132H/WT}$-induced DNA methylation changes between independent HCT116 knock-in clones.
Figure 12 (continued): Consistent $IDH1^{R132H/WT}$-induced DNA methylation changes between independent HCT116 knock-in clones.

(A) Unsupervised hierarchical clustering of HCT116 samples using all loci interrogated. Samples are represented by columns and CpG loci by rows. Samples are annotated by $IDH1$ genotype where the wild-type parental cells are in blue and the $IDH1^{R132H/WT}$ knock-in clones are in light and dark gray. The color of the heat map represents the methylation level ($\beta$) where unmethylated is white ($\beta = 0$), partially methylated is burgundy ($\beta = 0.5$), and fully methylated is black ($\beta = 1$). Clustering is performed with an average clustering algorithm and Euclidean distance dissimilarity metric. (B) Scatterplot of CpG loci $t$-statistics of differential methylation for $IDH1^{R132H/WT}$ knock-in clones 1 (KI-1: x-axis) and 2 (KI-2: y-axis) as compared to the wild-type parent. CpG loci significant in both clones are colored red and green for hyper- and hypo-methylated loci, respectively. Dashed blue lines represent the significance thresholds (FDR < 0.01). (C) Venn diagram showing the overlap of $IDH1^{R132H/WT}$-associated differential loci in each clone as compared to the parent. (D) Quantile-quantile (Q-Q) plot of $t$-statistics for the cumulative differential DNA methylation analysis between $IDH1^{R132H/WT}$ knock-in and parental cells show a larger distribution than that expected by the null hypothesis. Each point represents the $t$-statistic of one CpG locus where those that are significantly hypermethylated ($n = 2,010$) or hypomethylated ($n = 842$) in $IDH1^{R132H/WT}$ knock-ins are colored red and green, respectively. The solid blue line represents the distribution expected under the null hypothesis, and the dashed lines represent point-wise significant thresholds (FDR < 0.01).
LGGs (Turcan et al. 2012). Conversely, hypomethylated loci showed overrepresentation of nucleotide metabolism and differentiation. Interestingly, an examination of the relative methylation distribution of the 2,010 hypermethylated and 842 hypomethylated loci showed a depletion of moderately methylated loci (i.e. probes with $\beta \sim 0.2-0.7$ in wild-type cells) and a corresponding shift toward greater or lesser methylation levels in the knock-in cells (Figure 11C). Indeed, loci that became hypermethylated in the $\text{IDH1}^{R132H/WT}$ knock-ins generally arose from those CpG sites harboring a considerable degree of pre-existing methylation in the wild-type parental cells. These hypermethylated loci had on average a greater starting methylation level in the parental HCT116 cells than those that became hypomethylated, which tended to arise from CpG sites with lower starting methylation levels (Figure 11D). Together, these data support that endogenous heterozygous expression of the $\text{IDH1}^{R132H/WT}$ mutation systematically impacts global DNA methylation levels, and furthermore that the direction of change is dependent upon the initial methylation levels in the parental cells.

### 3.3.3 Methylation alterations observed in HCT116 $\text{IDH1}^{R132H/WT}$ knock-in cell lines are similarly affected in a brain tumor cell line overexpressing $\text{IDH1}^{R132H}$

To determine the general applicability of the above results, we also assayed genome-wide DNA methylation in a human oligodendroglioma (HOG) cell line stably overexpressing $\text{IDH1}^{R132H}$ (Reitman et al. 2011). Comparison of the relative methylation ($\beta$) distribution of all assayed CpG sites revealed a shift in DNA methylation upon
IDH1<sup>R132H</sup> expression, as compared to HOG cells expressing empty vector, resembling the shift in overall methylation observed in the HCT116 IDH1<sup>R132H/WT</sup> cells and resulting in increased frequency of CpG sites with high levels of methylation (β ~ 0.8-1.0) (Figure 13A). To determine whether the expression of IDH1<sup>R132H</sup> in HOG cells had a similar effect on DNA methylation at specific loci, we examined the influence of IDH1<sup>R132H</sup> overexpression on methylation at loci shown to be differentially methylated in the HCT116 knock-in cells. There was a largely concordant influence of IDH mutation on the behavior of the individual CpG loci in both models (Figure 13B). Considering the different cell backgrounds, it was not surprising that loci that undergo hypermethylation in HCT116 IDH1<sup>R132H/WT</sup> knock-in clones, which started with relatively high methylation levels in parental HCT116 cells, showed a broader methylation distribution in HOG cells, as did the loci that became hypomethylated in HCT116 IDH1<sup>R132H/WT</sup> knock-in clones (compare Figure 11C and Figure 13C). Nevertheless, there was a similar influence of the IDH1 mutation on methylation distribution in both sets, in that there was a general depletion of sites with intermediate methylation levels (β ~ 0.3-0.7), and an accumulation of sites with methylation levels at the extremes (β > 0.8 or < 0.2; Figure 13C). Moreover, loci that were significantly hypermethylated in HCT116 IDH1<sup>R132H/WT</sup> cells showed an overall increase in methylation in the HOG cells overexpressing the IDH1<sup>R132H</sup> mutant (P ≤ 0.0098), and loci that were significantly
Figure 13: Human oligodendroglioma (HOG) cells overexpressing IDH1$^{R132H}$ recapitulate the changes in DNA methylation observed in HCT116 cells.
Figure 13 (continued): Human oligodendroglioma (HOG) cells overexpressing IDH1<sup>R132H</sup> recapitulate the changes in DNA methylation observed in HCT116 cells.

(A) Relative DNA methylation (β) distribution for HOG cells overexpressing IDH1<sup>R132H</sup> (IDH1<sup>R132H+</sup>: gray) and HOG cells infected with an empty vector (WT: blue). Frequency (y-axis) is plotted by β such that the total probability (area under the curve) is equal to one. The HOG cell line overexpressing IDH1<sup>R132H</sup> shows an increase in methylated loci as compared to the wild-type parental cells. (B) Hierarchical clustering of HOG samples using the HCT116 IDH1<sup>R132H+/WT</sup> differential loci. Samples are represented by columns and differential CpG loci by rows and denoted by blue (WT) and gray (IDH1<sup>R132H+</sup>). CpG loci are annotated by their differential methylation in HCT116 cells, where red is hypermethylated and green is hypomethylated in HCT116 IDH1<sup>R132H+/WT</sup> knock-ins as compared to the wild-type parental line. The color of the heat map represents β, where unmethylated is white (β = 0), partially methylated is burgundy (β = 0.5), and fully methylated is black (β = 1). Clustering is performed with an average clustering algorithm and Euclidean distance dissimilarity metric. (C) Relative DNA methylation distribution for the HOG samples using the HCT116 IDH1<sup>R132H+/WT</sup> differentially methylated loci. Frequency of the differentially methylated loci are plotted relative to β for hypomethylated (green) and hypermethylated loci (red) in HOG cells infected with an IDH1<sup>R132H</sup> overexpression vector (IDH1<sup>R132H+</sup>; dashed line) or an empty vector (WT: solid line). (D) Box-and-whisker plot of DNA methylation levels for HCT116 IDH1<sup>R132H+/WT</sup> hypomethylated (green outline) and hypermethylated loci (red outline) in the HOG cells overexpressing IDH1<sup>R132H</sup> (R132H) or an empty vector (WT). Loci that were hypermethylated in HCT116 IDH1<sup>R132H+/WT</sup> knock-ins have a higher methylation level in HOG IDH1<sup>R132H+</sup> cells as compared to HOG cells overexpressing an empty vector (WT) (P = 9.5 × 10<sup>-6</sup>; Mann-Whitney U test). Additionally, those loci which underwent hypomethylation in HCT116 IDH1<sup>R132H+/WT</sup> knock-ins have a lower methylation level in HOG cells overexpressing IDH1<sup>R132H</sup> as compared to the empty vector (WT) control (P = 0.0098).
hypomethylated in the HCT116 \( IDH1^{R132H/WT} \) cells showed an overall decrease in methylation levels in the HOG cell line expressing the \( IDH1^{R132H} \) mutant (\( P \leq 0.084 \); Figure 13D). Like the HCT116 cell line, those loci that became hypermethylated in the \( IDH1^{R132H} \)-expressing HOG cells had a higher initial DNA methylation level in HOG parental cells than those that became hypomethylated (Figure 13D). Taken together, these data suggest that loci differentially methylated in the HCT116 \( IDH1^{R132H/WT} \) knock-in model are similarly affected by overexpression of mutant IDH1 in oligodendroglioma cells.

### 3.3.4 Methylation alterations observed in HCT116 \( IDH1^{R132H/WT} \) knock-in cell lines are similarly associated with IDH1 mutation in IDH1-mutant and G-CIMP+ primary gliomas

Previously, the Cancer Genome Atlas (TCGA) consortium identified a subset of GBMs exhibiting a high degree of concordant hypermethylation events, dubbed the glioma-specific “CpG island methylator phenotype” (G-CIMP), and further showed that this phenotype was significantly associated with \( IDH1 \) mutation in primary tumors (Noushmehr et al. 2010). A more recent study also characterized a cohort of primarily G-CIMP+ LGGs using DNA methylation and expression platforms (Turcan et al. 2012). We selected 61 primary GBMs characterized as part of the TCGA project for which definitive \( IDH1 \) mutational status and Illumina HumanMethylation27 data were available (Noushmehr et al. 2010) as well as 81 LGGs for which G-CIMP status and Illumina HumanMethylation450 data were available (Turcan et al. 2012) for further
analysis. Although the overall relative methylation ($\beta$) distribution for the two cohorts differed somewhat due to the use of the two different DNA methylation platforms, there was an overall shift toward higher methylation levels in tumors with $IDH1$ mutations and/or G-CIMP+ classification, particularly in the frequency of CpG sites with a high level of methylation ($\beta > 0.7$), and a reduced frequency of loci with low levels of methylation ($\beta < 0.2$; Figure 14A,B).

To determine whether the alterations in DNA methylation induced by $IDH1^{R132H/WT}$ are reflective of $IDH1$ mutation-associated methylation changes observed in primary tumors, we performed hierarchical clustering on the 61 TCGA GBMs and the 81 LGGs using the subset of CpG sites altered in the HCT116 $IDH1^{R132H/WT}$ model. We found that the methylation status of the HCT116 $IDH1^{R132H/WT}$ differentially methylated loci segregated wild-type from $IDH1$ mutated GBMs (Figure 14C) as well as G-CIMP+ from G-CIMP negative (G-CIMP−) LGGs (Figure 14D). Permutation analyses confirmed that the segregation achieved for the 2,852 HCT116 differential CpG loci was significantly greater than that achieved using the same number of randomly selected probes ($P < 0.001$). Although HCT116 cells have different underlying methylation patterns than those of the brain, the loci hypermethylated in HCT116 $IDH1^{R132H/WT}$ cells show the same trend toward hypermethylation in the $IDH1$-mutant and G-CIMP+ gliomas relative to $IDH1$-wild-type or G-CIMP– tumors ($P \leq 0.018$; Figure 14E,F; Figure 15A,C).
Figure 14: IDH1 mutant and G-CIMP+ gliomas recapitulate the DNA methylation alterations observed in cell line models.
Figure 14 (continued): IDH1 mutant and G-CIMP+ gliomas recapitulate the DNA methylation alterations observed in cell line models.

(A) Relative DNA methylation (β) distribution for 61 TCGA GBMs with definitive IDH1 mutational status and HumanMethylation27 data available (Noushmehr et al. 2010). Tumors that have wild-type (IDH1 WT: blue) and mutated (IDH1 mut: gray) IDH1 are drawn separately. (B) Relative β distribution for 81 LGGs from Turcan et al. classified as G-CIMP negative (G-CIMP-: blue) or positive (G-CIMP+: gray) profiled on the HumanMethylation450 array (Turcan et al. 2012). Hierarchical clustering of the (C) TCGA GBM and (D) Turcan et al. LGG cohorts using the 2,852 HCT116 IDH1R132H/WT differentially methylated loci separates IDH1 WT from IDH1 mut and G-CIMP+ from G-CIMP− tumors (P < 0.001). Samples are represented by columns and CpG loci by rows. Samples are annotated by IDH1 genotype for wild-type (blue) and mutated (gray) tumors. CpG loci are annotated by their differential methylation status, where red is hypermethylated and green is hypomethylated in HCT116 IDH1R132H/WT. The color of the heat map represents β where unmethylated is white (β = 0), partially methylated is burgundy (β = 0.5), and fully methylated is black (β = 1). Clustering was performed with an average clustering agglomerative algorithm and Euclidean distance dissimilarity metric. Relative β distribution of HCT116 IDH1R132H/WT differentially methylated loci in the (E) TCGA GBM and (F) Turcan et al. LGG cohorts. Frequency of the differentially methylated loci are plotted relative to β for hypomethylated (green) and hypermethylated loci (red) in IDH wild-type or G-CIMP− tumors (solid line) and IDH mutant or G-CIMP+ tumors (dashed line).
Figure 15: Comparison of HCT116 \( IDH1^{R132H/WT} \) differentially methylated loci with primary tumors yields a significant overlap.
Figure 15 (continued): Comparison of HCT116 IDH1$^{R132H/WT}$ differentially methylated loci with primary tumors yields a significant overlap.

(A) Box-and-whisker plot of DNA methylation levels for HCT116 IDH1$^{R132H/WT}$ hypomethylated (green outline) and hypermethylated (red outline) loci in TCGA primary GBMs that are wild-type for IDH1 (WT) or harbor a mutation in IDH1 (mut) (Noushmehr et al. 2010). Loci that were hypermethylated in HCT116 IDH1$^{R132H/WT}$ knock-ins have a higher methylation level in TCGA primary GBMs that harbor IDH1 mutations (mut) compared to those that do not (WT; $P = 1.5 \times 10^{-5}$; Mann-Whitney U test). (B) Venn diagram overlap of HCT116 differentially methylated genes (dashed line) with those reported by the TCGA (solid line) in G-CIMP+ versus G-CIMP− GBMs of the Proneural subtype (note that genes not loci are shown here to be consistent with the results reported by Noushmehr et al.) (Noushmehr et al. 2010). A significant overlap exists for the hypermethylated genes ($P = 3.1 \times 10^{-11}$; Fisher’s exact test). (C) Box-and-whisker plot of DNA methylation levels as in (A) except for LGGs that are G-CIMP+ or G-CIMP− published by Turcan et al. (Turcan et al. 2012). The HCT116 IDH1$^{R132H/WT}$ hypermethylated loci had on average a greater methylation level in G-CIMP+ LGGs than G-CIMP− LGGs ($P = 0.018$; Mann-Whitney U test). (D) Venn diagram of HCT116 IDH1$^{R132H/WT}$ differentially methylated loci (long dashed lines) compared to those reported for LGGs and an astrocyte model overexpressing IDH1$^{R132H}$ (Turcan et al. 2012). The HCT116 IDH1$^{R132H/WT}$ hypermethylated loci had a significant overlap with the LGG hypermethylated loci ($P = 7.8 \times 10^{-70}$) as well as with both hypermethylated and hypomethylated loci in the astrocyte model ($P \leq 0.12$; Fisher’s exact test). (E) Box-and-whisker plot of DNA methylation levels in G-CIMP+ and G-CIMP− LGGs for the hypomethylated (green outline) and hypermethylated (red outline) loci reported as significantly different between G-CIMP groups by Turcan et al. (Turcan et al. 2012). These data show that hypermethylation in primary tumors occurs primarily at loci that already have a large degree of methylation (average $\beta = 0.62$). Conversely, hypomethylation tends to occur at CpG loci that are less methylated than those at which hypermethylation occurs ($P < 10^{-300}$; Mann-Whitney U test). (F) Relative DNA methylation distribution for the LGG tumors across the LGG G-CIMP differentially methylated loci reported by Turcan et al. (Turcan et al. 2012). Frequency of the differentially methylated loci are plotted relative to $\beta$ for hypomethylated (green) and hypermethylated loci (red) for LGG tumors that are categorized as G-CIMP+ (dashed line) or G-CIMP− (solid line).
A further comparison of the genes hypermethylated in HCT116 $IDH1^{R132H/WT}$ knock-in cells with those reported to be hypermethylated in the G-CIMP+ versus G-CIMP− GBMs (Noushmehr et al. 2010) and those reported to be differentially methylated in G-CIMP+ versus G-CIMP− LGGs (Turcan et al. 2012) revealed a significant overlap (Figure 15B,D; $P \leq 1.8 \times 10^{-22}$, Fisher’s exact). Approximately 17% of the sites that became hypermethylated in HCT116 $IDH1^{R132H/WT}$ were also hypermethylated in $IDH1$ mutated GBMs and more than 40% were also hypermethylated in G-CIMP+ LGGs. Interestingly, analysis of data from primary human astrocytes overexpressing $IDH1^{R132H}$ (Turcan et al. 2012) showed that there was not only a similar frequency of IDH mutation-induced differential methylation events, and fraction of hyper- versus hypomethylation, as detected in the HCT116 $IDH1^{R132H/WT}$ knock-in lines (in both cases 9% of CpG sites interrogated were differentially methylated overall, 70% of which were hypermethylated), but also a similar degree of concordance between hypermethylation induced in these two models and that associated with the G-CIMP+ phenotype in primary LGGs (e.g. 40% of HCT116 $IDH1^{R132H/WT}$ versus 36% of $IDH1^{R132H}$ astrocyte hypermethylated sites were concordantly hypermethylated in G-CIMP+ LGGs; $P \leq 7.9 \times 10^{-41}$, Fisher’s exact; Figure 15D). Together, these analyses show that loci hypermethylated in HCT116 $IDH1^{R132H/WT}$ knock-in lines are also subject to hypermethylation in $IDH1$-mutant or G-CIMP+ primary gliomas, and further that the impact of $IDH1$ mutation on the methylome is similar regardless of the cell type of
origin or starting methylation level. However, unlike the cell culture models, the HCT116 IDH1\textsuperscript{R132H/WT} hypomethylated loci did not show evidence of reduced methylation in mutant versus wild-type IDH1 GBMs or G-CIMP+ versus G-CIMP− LGGs (Figure 14E,F; Figure 15A,C). Nevertheless, CpG sites that became hypomethylated in response to IDH1\textsuperscript{R132H/WT} in HCT116 cells generally exhibited lower methylation levels in the gliomas than those prone to hypermethylation (Figure 14E,F).

An examination of the loci identified by Turcan et al. (Turcan et al. 2012) as differentially methylated in G-CIMP+ versus G-CIMP− LGGs revealed a similar relationship to methylation level in that the hypomethylated loci tended to exhibit a lower average $\beta$ across all tumors (mean $\beta = 0.32$) than those that showed G-CIMP-associated hypermethylation (Figure 15E,F). Thus, the same relative depletion of moderately methylated loci observed in the HCT116 knock-in model was also observed in G-CIMP+ and IDH-mutant primary gliomas.

### 3.3.5 Effects of $\textit{IDH1}^{R132H/WT}$ on gene expression

In addition to $\textit{IDH1}$-mutant GBMs being classified almost exclusively as G-CIMP positive, they also robustly associate with distinct gene expression signatures, in particular, the proneural GBM subclassification (Verhaak et al. 2010). We took advantage of our isogenic system and sought to examine the contribution of a one-base-pair allelic substitution at $\textit{IDH1}$ on gene expression patterns. Genome-wide expression data were generated for the HCT116 parent and $\textit{IDH1}^{R132H/WT}$ knock-in clones using
Affymetrix Human Genome U133A 2.0 Arrays. Analysis of parental versus KI-1 and KI-2 $IDH1^{R132H/WT}$ knock-in clones yielded 21 differentially expressed probes from 19 unique genes, 18 of which were downregulated in the $IDH1^{R132H/WT}$ clones (FDR < 0.05; fold-change > 1.5; Figure 16A). Gene Set Enrichment Analysis (GSEA) identified gene sets involved in protein degradation and Wnt signaling as enriched among the genes downregulated in $IDH1^{R132H/WT}$ cells. Interestingly, 6 of 18 and 7 of 18 downregulated genes overlapped with those reported to be downregulated in the G-CIMP+ as compared to G-CIMP− proneural GBMs and LGGs, respectively (Noushmehr et al. 2010; Turcan et al. 2012), which is significantly more than would be expected based on random chance ($P \leq 0.023$, Fisher’s exact test; Figure 16A). We selected four of the 18 genes to verify their expression level in the cell lines. Direct analysis of the expression of these $IDH1^{R132H/WT}$–mediated downregulated genes by quantitative reverse transcriptase PCR validated the microarray findings and showed that the mRNA expression was significantly reduced for all four genes in the $IDH1^{R132H/WT}$ knock-in clones relative to parental HCT116 cells (Figure 16B). Furthermore, a comparison of the expression levels of the same genes in 117 GBMs from the TCGA project (Verhaak et al. 2010) and 52 LGGs from the Turcan et al. (Turcan et al. 2012) cohort, for which gene expression data and $IDH1$ mutational status or G-CIMP classification were available, showed a corresponding downregulation of these same genes in $IDH1$-mutant and G-CIMP+ gliomas relative to $IDH1$-wild type and G-CIMP− tumors (Figure 16C).
Figure 16: Gene expression profiling of HCT116 IDH1<sup>R132H/WT</sup> cell lines.
Figure 16 (continued): Gene expression profiling of HCT116 IDH1^{R132H/WT} cell lines.

HCT116 IDH1^{R132H/WT} knock-in clones and parental cells were analyzed using Affymetrix Human Genome 2.0 Arrays. (A) Hierarchical clustering of probes differentially expressed in HCT116 parent versus knock-in cells. Samples are represented by columns and differential probes by rows. Samples are annotated by IDH1 genotype for wild-type HCT116 (blue) and IDH1^{R132H/WT} knock-in cells (gray). Each probe is normalized (Z-score), and the color of the heat map represents the relative expression of each sample (red: overexpressed; green: underexpressed). Probes are annotated for overlap with genes found differentially expressed in TCGA GBMs (black) (Noushmehr et al. 2010) and LGGs (gray) (Turcan et al. 2012). Clustering is performed using an average clustering algorithm and a Euclidean distance dissimilarity metric of the normalized expression. (B) Quantitative real-time PCR (Q-PCR) validation of candidate genes UBB, RBP1, VIM, and GJA1 for IDH1^{R132H/WT}–mediated transcriptional repression. Gene expression fold-changes were quantified for each candidate gene by using three independent mRNA samples from each clone and calculated relative to parental cell line. Shown is the mean ± standard deviation of the triplicate determinations relative to HCT116 cells. (C) Stripcharts of gene expression values for validated genes in HCT116 parental (WT: blue) and IDH1^{R132H/WT} (R132H: gray) cells as well as the same probes from 117 TCGA primary GBMs that are IDH1 wild-type (WT: blue, n = 98) or mutated (mut: gray, n = 19) and 52 LGGs that have gene expression data and G-CIMP negative (CIMP−: blue; n = 16) or positive (CIMP+: gray; n = 36) classification. Lines represent the median values for each group. P-values were calculated using Welch’s two-sided t-test.
3.3.6 Relationship between \textit{IDH1} mutation-induced alterations in DNA methylation and gene expression

We next examined the relationship between DNA methylation and gene expression changes induced by \textit{IDH1}^{R132H/WT} mutation in the HCT116 knock-ins. Analysis of all genes interrogated on both the DNA methylation and gene expression platforms indicated a subtle yet significant negative correlation between the average change in gene expression and average change in DNA methylation for a given gene (Spearman’s $\rho = -0.021$, $P = 0.027$; Figure 17). To further explore the relationship between changes in gene expression and DNA methylation, we analyzed the distribution of \textit{IDH1}^{R132H/WT} hyper- and hypomethylated loci relative to the canonical transcription start site (TSS) of the closest gene or CpG island. Interestingly, hypomethylated loci were depleted around the canonical TSS and tended to occur towards the edges of the CpG island, whereas hypermethylated loci were more broadly distributed, with most occurring within CpG islands or CpG island ‘shores’, but many without (Figure 18), consistent with these loci having significant average pre-existing methylation. The analysis of chromatin immunoprecipitation and massively parallel sequencing (ChIP-seq) data from wild-type HCT116 cells (Birney et al. 2007) showed that CpG sites that became hypermethylated in the HCT116 \textit{IDH1}^{R132H/WT} cells are depleted of RNA polymerase II (Pol II) binding and histone H3 lysine 4 trimethylation (H3K4me3) relative to sites that become hypomethylated or remain unchanged (Figure 19A-E). These data suggest that sites that undergo hypermethylation in response
Figure 17: HCT116 DNA methylation and gene expression changes negatively correlate.

Average change in DNA methylation ($\beta$) between HCT116 parental and $IDHI^{R132H/WT}$ knock-in is plotted (x-axis) as compared to change in gene expression (log2 scale; y-axis). Each point represents one common gene between the Illumina Infinium HumanMethylation27 and Affymetrix HG133_2 platforms. Genes for which multiple probes exist have the average change plotted. Genes that have a significant increase in DNA methylation at one probe and a significant decrease in gene expression are colored green. A significant negative correlation exists (Spearman’s rank correlation $\rho = -0.021$; $P = 0.027$).
Figure 18: Distribution of HCT116 IDH1\textsuperscript{R132H/WT} differentially methylated loci relative to transcription start sites (TSS) and CpG islands.

Distribution of hypermethylated (red) and hypomethylated (green) loci in HCT116 IDH1\textsuperscript{R132H/WT} knock-in cells relative to TSS (A) and CpG island (B). All assay probes are shown in black. Infinium probes were mapped to the closest UCSC RefSeq hg18 TSS and CpG island. CpG islands were defined by the criteria of Takai and Jones (Takai and Jones 2002). Plots are oriented in the direction of transcription (A) and the direction of transcription of the closest gene (B). (C) Proportion of HCT116 IDH1\textsuperscript{R132H/WT} differentially methylated CpG sites in CpG islands (blue), CpG island ‘shores’ (2000 bp from either CpG island boundary) (light blue) and other regions (tan).
Figure 19: RNA polymerase II (Pol II) and Histone 3 Lysine 4 trimethylation (H3K4me3) ChIP-seq data indicates $IDH1^{R132H/WT}$ differentially methylated loci have reduced Pol II binding in HCT116 wild-type cells.
Figure 19 (continued): RNA polymerase II (Pol II) and Histone 3 Lysine 4 trimethylation (H3K4me3) ChIP-seq data indicates $IDH1^{R132H/WT}$ differentially methylated loci have reduced Pol II binding in HCT116 wild-type cells.

(A) H3K4me3 and (B) Pol II average enrichment is plotted relative to the HumanMethylation27 interrogated CpG loci oriented in the direction of transcription of the closest gene. Lines represent the average normalized (10^8 reads / genome) reads at all Infinium interrogated loci (black), HCT116 $IDH1^{R132H/WT}$ hypermethylated loci (red), and hypomethylated loci (green). (C) Plot of the HCT116 $IDH1^{R132H/WT}$ differential t-statistics sorted from most hypomethylated (top) to most hypermethylated (bottom). CpG loci that met statistical significance are shown in color. (D) Heatmap representation of H3K4me3 data summarized in (A) where each row represents the sequence surrounding a CpG dinucleotide on the HumanMethylation27 assay and each column represents a 10bp region. The probes are sorted as in (C) and significantly hypomethylated (green) and hypermethylated (red) loci are denoted as above and below the green or red lines, respectively. Normalized ChIP-seq read depth is represented by color (key below). (E) Heatmap as described in (D) for Pol II binding.

to heterozygous expression of $IDH1^{R132H/WT}$ reside in domains that are less transcriptionally active and/or permissive in the parental HCT116 cells prior to $IDH1^{R132H/WT}$ knock-in. Although relatively enriched in Pol II, there was a decisive dip in Pol II binding surrounding the hypomethylated loci (Figure 19B). The finding that many sites that become hypermethylated may already be less active in the parental HCT116 cells may at least in part explain the apparent lack of correlation between DNA methylation changes and altered gene expression overall.

3.3.7 Global and gene-specific histone lysine methylation modifications correlate with DNA methylation and gene expression alterations in $IDH1^{R132H/WT}$ cells

To validate and further examine the relationship between epigenetic alterations and gene expression changes induced by $IDH1$ mutations, we analyzed in more detail three
hypermethylated loci that exhibited a significant reduction in gene expression in HCT116 \textit{IDH1}^{R132H/WT} knock-in cells (\textit{UBB}, \textit{RBP1}, and \textit{SERPINB5}) and one that did not (\textit{PDLIM2}) using bisulfite sequencing analysis (Figure 20). Primers were designed to amplify a region of several hundred base pairs and covered 1-3 of the CpG probes interrogated on the Illumina methylation array in each case. The methylation density, as estimated from the total number of methylated CpG sites among the total CpGs and alleles analyzed, showed a high correspondence with the methylation levels of underlying probes as assessed on the Illumina platform (compare Figure 20A with Figure 20B), and confirmed the increase in methylation observed in the \textit{IDH1}^{R132H/WT} knock-in clones relative to the parental HCT116 cell line (Figure 20B). In contrast, there was no significant accumulation of DNA methylation in 3 clonal control lines that were subject to the same infection and selection protocol as the \textit{IDH1}^{R132H/WT} clones, but did not exhibit homologous recombination at the \textit{IDH1} locus (Figure 21), excluding the possibility that the hypermethylation was the result of a non-targeted effect or clonal selection. Comparison of the patterns of methylation across individual alleles in a population of parental and \textit{IDH1} knock-in cells suggested that hypermethylation resulted both from new methylation arising on relatively unmethylated alleles and from an increase in the density across alleles, suggesting a filling-in or spreading of pre-existing methylation, depending on the gene analyzed (Figure 20B).
Figure 20: Bisulfite sequence analysis of candidate CpG loci validates $IDH1^{R132H/WT}$-induced DNA methylation changes.
Figure 20 (continued): Bisulfite sequence analysis of candidate CpG loci validates \( IDH1^{R132H/WT} \)-induced DNA methylation changes.

(A) Stripcharts of DNA methylation values (β) in HCT116 parental (WT: black) and \( IDH1^{R132H/WT} \) (R132H: gray) cells are shown next to data for the same loci in TCGA GBMs that are \( IDH1 \) wild-type (WT: black) or mutated (mut: gray) and LGGs that are G-CIMP negative (CIMP−: black) or positive (CIMP+: gray). \( P \)-values are from Welch’s two-sided \( t \)-test. (B) Bisulfite sequence analysis of the loci shown in (A). (Top) The interrogated region is depicted with a schematic of the gene with CpG dinucleotides represented by vertical tick marks on the x-axis. Browser tracks denote CpG islands, the region analyzed by bisulfite sequencing (Bis seq) and CpG loci interrogated by the HumanMethylation27 platform (Inf27). The CpG locus plotted in (A) is denoted by a black arrowhead above the Inf27 track. (Bottom) Bisulfite sequencing results for the regions denoted in the schematic above. Each row represents a sequenced allele and each dot represents a CpG. Black dots represent methylated CpGs, and white dots represent unmethylated CpGs. The CpG shown in (A) is denoted with a black arrowhead, and other CpGs interrogated by the assay are shown in gray.

Figure 21: Bisulfite sequencing at candidate loci of HCT116 non-targeted clones indicates no change in methylation is resultant of the gene targeting procedure.

Each row represents a sequenced allele and each dot represents a CpG. Black dots represent methylated CpGs, and white dots represent unmethylated CpGs. Compare to Figure 20B HCT116 \( SERPINB5 \) results.
Next, we examined levels of histone lysine methylation in IDH1R132H/WT knock-in and parental cells. Western blot analysis showed that the global levels of H3K4me3, H3K9me3, H3K27me3, and H4K20me3 were increased in KI-1 and KI-2 clones as compared to parental cells (Figure 22A). These results are consistent with previous models overexpressing IDH mutants (Xu et al. 2011a; Lu et al. 2012; Turcan et al. 2012). Chromatin immunoprecipitation (ChIP) followed by quantitative PCR was employed to examine the levels of these histone marks at the promoters of the same four hypermethylated candidate loci. All four loci showed an accumulation of H3K9me3 that accompanied the hypermethylation of DNA (compare Figure 20B and Figure 22B). H3K27me3 was much less abundant at these loci in the parental HCT116 cells and showed only modest accumulation only at the SERPINB5 locus in the knock-in cells. Taken together, these data suggest that both global and gene-specific alterations in histone methylation occur in conjunction with DNA methylation in response to heterozygous mutation of IDH1R132H/WT and that the two can work together to reshape the epigenome.
Figure 22: Global and gene-specific histone lysine methylation coincides with IDH1\textsuperscript{R132H/WT}-induced DNA methylation.

(A) Western blots for total H3K4me3, H3K9me3, H3K27me3 and H4K20me3 histone modifications for parental (HCT116) and IDH1\textsuperscript{R132H/WT} knock-in (KI-1 and KI-2) cell lines. Also shown are total H3 and H4 controls. (B) ChIP was performed using antibodies against H3K9me3, H3K27me3, pan-H3, or IgG and immunoprecipitated DNA quantified by Q-PCR using primers specific to the promoter regions of the indicated genes. Bar graphs depict the mean enrichment of the specific histone mark relative to that of total histone H3 for the same genomic region. Error bars represent the standard deviation of two independent experiments except for PDLIM2.
3.3.8 Gene silencing at specific $IDH1^{R132H/WT}$-targeted loci is reversed using a DNA hypomethylating agent

The relationship between DNA methylation and gene silencing at specific loci affected by the $IDH1^{R132H/WT}$ mutation, and the potential reversibility of these alterations, was examined by determining the effects of exposure to a DNA methyltransferase inhibitor on DNA methylation and gene expression at the $RBP1$, $UBB$, and $SERPINB5$ genes. Parental HCT116 cells and knock-in clones were treated with 5-aza-2'-deoxycytidine (decitabine, Dacogen, DAC) or control, and gene expression levels were determined by RT-PCR (Figure 23A), while corresponding DNA methylation levels were determined by bisulfite sequencing (Figure 23B). The $UBB$ promoter, which was partially methylated (~80-89%) and exhibited low levels of gene expression in the parental line, approached complete methylation (90-98%; Figure 20B, 23B) commensurate with a complete loss of $UBB$ mRNA expression in the knock-in clones (Figure 23A; see also Figure 16B). Upon DAC treatment, DNA was hypomethylated, and a robust restoration of $UBB$ gene expression was observed in both the parental and knock-in clones (Figure 23A,B). Similarly, the promoter regions of $SERPINB5$ and $RBP1$, which were virtually unmethylated or partially methylated (28-35%), respectively, and expressed in the parental HCT116 cell line, underwent an increase in methylation in the HCT116 knock-in clones (from 30 to 71% and 35 to 89%, respectively) (Figure 20B, 23B), and showed a corresponding ~2-fold decrease in gene expression (Figure 23A; see also Figure 16B).
Figure 23: Inhibition of DNA methylation results in restoration of gene expression for $IDH1^{R132H/WT}$-repressed loci.
Figure 23 (continued): Inhibition of DNA methylation results in restoration of gene expression for IDH1\textsuperscript{R132H/WT}-repressed loci.

(A) Candidate gene expression reactivation in 5-aza-2’-deoxycytidine-treated IDH1-mutant cell lines. Parental or IDH1\textsuperscript{R132H/WT} knock-in (KI-1, KI-2) cells were treated with 5 μM of 5-aza-2’-deoxycytidine (DAC) or control (PBS) for 48 hours. Following treatment, relative mRNA levels of UBB, RBP1, and SERPINB5 were measured by Q-PCR. Bar graph represents expression fold-change relative to untreated parental cell line. Error bars represent standard deviation of three independent experiments. (B) Bisulfite sequencing analysis of candidate genes in (A) for untreated and DAC-treated cell lines. Each row represents a sequenced allele and each dot represents a CpG. Black dots represent methylated CpGs, and white dots represent unmethylated CpGs.

DAC treatment resulted in hypomethylation of DNA and a restoration of RBP1 and SERPINB5 gene expression to control levels and induced a similar relative increase in both parental and knock-in clones (Figure 23A,B). Taken together, our data support a model in which heterozygous IDH1 mutations can promote widespread alterations in DNA methylation, at least some of which contribute to altered gene expression patterns and can be reversed by treatment with a DNA hypomethylating agent.
3.4 Discussion of IDH mutations and their impact on the epigenome

3.4.1 Heterozygous IDH1 mutations induce alterations in DNA methylation

Heterozygous IDH1 mutations occur frequently in human tumors and display robust association with specific DNA methylation signatures. Using a somatic knock-in cell line system, we induced physiologic expression of the most frequently observed IDH1 mutation, IDH1\(^{R132H/WT}\), in a human cancer cell line. In this report, we establish that genome-wide alterations in DNA methylation result from induction of the IDH1\(^{R132H/WT}\) genotype. We found that the expression of the single IDH1\(^{R132H}\) allele was sufficient to induce widespread alterations in DNA methylation, including both hypermethylation and hypomethylation events. Notably, the IDH1\(^{R132H/WT}\)-mediated direction of methylation change was associated with the initial methylation state of the parental cells. Hypermethylation events tended to arise at CpG sites that were already moderately methylated (average 78%), whereas hypomethylation events occurred at CpG sites that were methylated to a lesser degree (average 17%), resulting in an accentuation of the bimodal distribution of CpG methylation. We further demonstrate that these alterations are similarly associated with IDH1 mutation and G-CIMP+ classification in gliomas and that the methylation signature identified can distinguish tumors with these features in two independent primary tumor cohorts. Together, these data provide evidence that heterozygous IDH1\(^{R132H/WT}\) mutations drive epigenetic
instability and initiate the methylation phenotypes observed in patients. Comparison of our data to that recently published by Turcan et al. (Turcan et al. 2012) showed a remarkably similar influence of the IDH1R132H mutation on the methylome, independent of the cell type of origin (HCT116 colon cancer cells versus primary astrocytes) or method of introduction, with an estimated 9% of CpG sites significantly affected overall, 70% of which underwent hypermethylation in both studies. Interestingly, whereas there was considerable concordance between sites that undergo IDH1 mutation-driven hypermethylation in these cell culture models and sites preferentially hypermethylated in CIMP+ gliomas, there was little relationship between hypomethylated sites in the different models and primary tumors, suggesting that the mechanisms driving IDH1 mutation-associated hypermethylation may differ from those driving hypomethylation.

3.4.2 Enzymatic activity of heterozygous IDH mutations

Tumor-associated IDH mutations are overwhelmingly observed to occur in the heterozygous state in primary tumors, and current theories have speculated that both the wild-type and mutant alleles are necessary for oncogenic function. IDH1 functions as a dimer, and previous studies have shown that presence of wild-type IDH1 can stimulate the reductive enzymatic activity of mutant IDH1, potentially through formation of a highly active IDH1R132H-IDH1WT heterodimer (Bralten et al. 2011). Therefore, gene dosage of mutant and wild-type alleles has the potential to influence IDH1 enzyme activity and D-2-HG production, the relative ratios of α-KG depletion and
D-2-HG accumulation, and ultimately the inhibition of Fe(II)/2-oxoglutarate-dependent dioxygenases. As D-2-HG is a weak antagonist of α-KG, competitive inhibition of α-KG-dependent dioxygenases requires a large fold excess of D-2-HG (Xu et al. 2011a). Therefore, a relatively small change in α-KG production resulting from replacement of one IDH1 allele can potentially have a large impact on α-KG-dependent enzymatic reactions or alternatively may poise the cell for D-2-HG-mediated effects.

3.4.3 TET proteins and 5-hydroxymethylcytosine

Several non-exclusive hypotheses have been proposed as potential mechanistic links between IDH mutants, reduction of cellular α-KG levels and/or accumulation of D-2-HG, and epigenetic alterations in cancer cells. One potential mechanism involves inhibition of TET proteins on the flux between 5-methylcytosine (5-mC), 5-hydroxymethylcytosine (5-hmC), and cytosine. The TET family proteins catalyze the conversion of 5-mC to 5-hmC (Tahiliani et al. 2009; Ito et al. 2010) and higher oxidation states, including 5-formylcytosine (5-fC) and 5-carboxylcytosine (5-caC) (He et al. 2011; Ito et al. 2011). These modified methylcytosine bases are likely intermediates in active DNA demethylation arising from removal by the base excision repair machinery and replacement with cytosine (Wu and Zhang 2010; Cortellino et al. 2011; Guo et al. 2011; He et al. 2011). Recent advances have provided insight into the roles of TET enzymes on the regulation of 5-hmC, the cellular context and genomic distribution of 5-hmC, and the importance of 5-hmC in regulation of transcription and cellular differentiation (Ito et al.
Bisulfite-based technologies, including the Illumina methylation platform used here, do not distinguish between 5-mC and 5-hmC (Nestor et al. 2010). Thus, the methylation values reported at any one CpG site represents the sum of the two, and there may be some CpG sites that are predominately one or the other. This could be one explanation for the hyper- and/or hypomethylation observed in response to IDH mutation in ours and other (Turcan et al. 2012) studies. Hypermethylation events could occur through inhibition of TET proteins and accumulation of 5-mC that would ensue due to the decreased conversion of 5mC to 5-hmC and decreased DNA demethylation (Wu and Zhang 2010; Cortellino et al. 2011; Dahl et al. 2011; Guo et al. 2011). Apparent hypomethylation could result from the conversion of all pre-existing 5-hmC to cytosine by base excision repair (e.g. thymidine glycosylase or MBD4) and/or the inability to maintain this mark during replication (Valinluck and Sowers 2007; Hashimoto et al. 2012). In the absence of TET function, the inability to generate new 5-hmC would lead to a depletion of 5-hmC and an increase in unmodified cytosine. Consistent with this, whereas IDH mutations in gliomas (Noushmehr et al. 2010) and in AMLs (Figueroa et al. 2010a) are associated with a CpG island hypermethylation phenotype, TET2 loss of function in myeloid tumors has been associated with both the CIMP phenotype (Figueroa et al. 2010a) as well as global
hypomethylation (Ko et al. 2010). The potential contribution of TET inhibition to the
IDH-mutation driven epigenetic alterations is further confounded by the findings that
whereas 5-hmC is depleted in a number of cancers and is reduced upon overexpression
of IDH1R132H in astrocytes (Turcan et al. 2012), total 5-hmC levels do not appear to
correlate with IDH mutations in primary gliomas (Jin et al. 2011b). Any influence that
IDH mutations and/or TET mutations might have on the site-specific patterns of 5-hmC
across the genome in cancer cells remains to be determined.

3.4.4 Histone (de)methylation

Another mechanism involves effects on histone methylation patterns. Of
particular interest are members of the Jumonji-C domain family of histone demethylases
(JHDMs), which demethylate histone lysine residues in an α-KG-dependent manner. D-
2-HG competitively inhibits JHDM activity in vitro (Chowdhury et al. 2011; Xu et al.
2011a). Overexpression of mutant IDH leads to alterations in histone methylation levels
in cell lines (Xu et al. 2011a; Lu et al. 2012; Turcan et al. 2012), and IDH mutations are
associated with increased histone methylation in primary human tumors (Xu et al.
2011a; Lu et al. 2012). Our data indicate that heterozygous expression of IDH1R132HWT
results in both global and site-specific alterations in histone lysine methylation patterns.
We find that whereas the global levels of multiple histone H3 and H4 methylation
modifications were increased, consistent with a broad inhibition of histone
demethylation, hypermethylation at H3K9 in particular accompanied site-specific
hypermethylation of DNA at several genes that were downregulated in IDH1<sup>R132H/WT</sup> knock-in cells. This is consistent with recent work by Turcan et al. (Turcan et al. 2012) who similarly showed a concomitant accumulation of H3K9me3 and DNA methylation at several genes that undergo epigenetic silencing in response to ectopic expression of mutant IDH1 in human astrocytes. Although it is currently unknown whether DNA methylation or H3K9me3 is the primary mediator of gene silencing in these cases, a recent study by Lu et al. (Lu et al. 2012) showed that increased H3K9me3 could occur independently of DNA methylation changes at the Cebpa and Adipoq loci in IDH2<sup>R172K</sup>-expressing mouse fibroblasts, and further that the global accumulation of H3K9me3 preceded that of DNA methylation in human astrocytes ectopically expressing IDH1<sup>R132H</sup>. Although the extent to which mutant IDH-driven alterations in DNA methylation are dependent upon changes in H3K9me3 (or vice versa) remains to be determined, these data suggest that altered H3K9me3, through the D-2-HG-mediated inhibition of an H3K9 demethylase, may underlie DNA hypermethylation-associated gene silencing at least at some loci. Close mechanistic ties are thought to exist between these two modifications, particularly in the maintenance of constitutive heterochromatin and the silencing of repetitive elements (Lehnertz et al. 2003; Cedar and Bergman 2009; Hashimoto et al. 2010). Targeted deposition of H3K9me2/3 can direct de novo DNA methylation in an integrated transgene system (Schultz et al. 2002) and precedes de novo methylation of stem cell genes during lineage specification of embryonic stem cells.
(Cedar and Bergman 2009). At present, the extent to which D-2-HG-mediated inhibition of one or more α-KG-dependent enzymes is responsible for the pathogenesis of IDH mutation-associated tumors remains to be determined. Future studies aimed at the integrated analyses of histone methylation and DNA (hydroxy)methylation at a genome-wide scale will be necessary to resolve the relative contribution to IDH mutation-associated cancers.

3.4.5 Relationship between DNA methylation and gene expression

We examined gene expression alterations resulting from induction of IDH1R132H/WT in HCT116 and revealed a subset of 19 genes that demonstrated significant changes in gene expression. Nearly all of the dysregulated genes (18 of 19) were downregulated and exhibited concomitant hypermethylation. In contrast, there was no significant association between hypomethylation and gene expression in our model. This is perhaps not surprising considering that most of the mutant IDH-driven hypomethylation events affected sites that had only low levels of methylation to begin with (average $\beta = 0.17$) and occurred more frequently at sites away from the canonical TSSs and toward the edges of CpG islands. Whereas the inverse correlation between the methylation status of CpG sites within CpG island-containing promoters and gene expression is well-described, recent genome-wide studies indicate that gene body methylation is positively correlated with gene expression in normal cells (Lister et al. 2009; Rauch et al. 2009; Maunakea et al. 2010) and furthermore, that loss of methylation
in such regions in cancer cells correlates with gene repression (Hon et al. 2012). Even in DNMT triple knock-out mouse embryonic stem cells (DNMT TKO) (Karimi et al. 2011), where there is a dramatic demethylation of the genome, surprisingly few genes \((n = 239)\) are significantly upregulated.

Somewhat surprisingly, despite the widespread hypermethylation events observed, the number of significantly downregulated genes was relatively small. Nevertheless, our findings are consistent with previous reports showing that only a small fraction (17\%) of loci hypermethylated in G-CIMP+ primary GBMs were associated with a concomitant decrease in gene expression (Noushmehr et al. 2010). A growing number of studies indicate a low correlation between DNA hypermethylation and gene expression at a global level in cancer cells (Hahn et al. 2008; Pike et al. 2008; Takeshima et al. 2009). This appears to be in part because many of the genes that undergo DNA hypermethylation in cancer cells are already marked by repressive chromatin and exist in a relatively low expression state in the parental cells. Consistent with this idea, CpG sites that became hypermethylated in response to \(IDH1^{R132H/WT}\) were already on average partially methylated and depleted of marks of active transcription (H3K4me3) and Pol II occupancy in the parental HCT116 cells relative to sites that became hypomethylated or were unaffected. The finding that a significant proportion of CpG sites prone to hypermethylation in the HCT116 \(IDH1^{R132H/WT}\) knock-in cells are similarly hypermethylated in \(IDH\) mutant and/or G-CIMP+ gliomas and \(IDH1^{R132H}\)
overexpressing cells suggests that the metabolic disturbance imposed by IDH mutations may impact upon a common set of vulnerable genomic sites. One common feature may be pre-existing marking by repressive chromatin. In particular, genes that are targets of the Polycomb repressor complex 2 (PRC2) and marked by H3K27me3 in embryonic or adult stem cells are highly prone to CpG island hypermethylation across tumor types (Ohm et al. 2007; Schlesinger et al. 2007; Widschwendter et al. 2007; Gal-Yam et al. 2008; Easwaran et al. 2012), including those hypermethylation events associated with the CIMP phenotype (Fang et al. 2011; Easwaran et al. 2012; Turcan et al. 2012). Indeed, sites that underwent IDH-driven hypermethylation in our HCT116 IDH1R132H/WT knock-in cells are enriched in Polycomb target genes, and there was even greater enrichment among those commonly hypermethylated between the HCT116 IDH1R132H/WT knock-in cells and astrocytes overexpressing the IDH1R132H mutant or G-CIMP+ LGGs (Turcan et al. 2012), suggesting that this may be an common underlying feature. In stem cells, marking by the Polycomb complex is thought to keep key developmental regulators in a low but poised transcriptional state to prevent lineage commitment and differentiation and to maintain self-renewal capacity. It has been proposed that the replacement or superimposition of such reversible chromatin-mediated repression by the more stable silencing associated with promoter DNA methylation may contribute to tumor progression by reducing the epigenetic plasticity that would normally allow stem/progenitor cells to activate differentiation programs or respond to cellular stress.
(Gal-Yam et al. 2008; Easwaran et al. 2012). Consistent with this idea, we found that the IDH-driven hypermethylated sites seen here were enriched in genes that play a role in the cellular response to extracellular stimuli, and cell-type specific differentiation functions.

### 3.4.6 Targets of mutant IDH1-mediated de novo epigenetic silencing

Although small in number, we identified several candidate genes that are targets of mutant IDH1-mediated *de novo* epigenetic silencing, including *RBP1*, *UBB*, and *SERPINB5*. Interestingly, over 33% of the genes downregulated in HCT116 *IDH1<sup>R132H/WT</sup>* cells overlapped with those reported to be downregulated in the G-CIMP+ as compared to G-CIMP− proneural GBMs and LGGs (Noushmehr et al. 2010; Turcan et al. 2012). *RBP1* has been demonstrated to be one of the genes most frequently subject to epigenetic silencing in G-CIMP+ primary GBMs (Noushmehr et al. 2010; Laffaire et al. 2011) and was also among the most strikingly subject to concordant hypermethylation and downregulation in our model. In addition to gliomas, *RBP1* is also silenced in conjunction with promoter hypermethylation in several cancer types, including lymphomas, esophageal squamous cell carcinomas, and gastric carcinomas (Esteller et al. 2002; Mizuiri et al. 2005; Shutoh et al. 2005; Chu et al. 2006). Involved in the transport of retinol, RBP1 regulates intracellular retinoic acid (RA) homeostasis. RA influences many important and diverse functions, including cell proliferation and differentiation, and retinoid signaling is often compromised in early carcinogenesis (Tang and Gudas
2011). RBP1 downregulation in cancer promotes loss of cellular differentiation and tumor progression through inhibition of retinoic acid receptor (RAR) activity and derepression of PI3K/Akt signaling (Farias et al. 2005a; Farias et al. 2005b). Further, RBP1 has been proposed as a tumor suppressor in bladder cancer, as its silencing contributes to cell proliferation and migration (Toki et al. 2010). Additionally, markers for RA signaling, including RBP1 mRNA, have been shown to be significantly decreased in long-term GBM survivors (associated with IDH1 mutation) and may be one potential area to target therapeutically (Barbus et al. 2011). We also identified UBB, coding for a highly conserved polyubiquitin precursor, as a target of IDH1R132H/WT-driven epigenetic silencing. Ubiquitin-mediated protein degradation plays important roles in the control of numerous cellular processes, including signal transduction, cell-cycle progression and transcriptional regulation, and abnormalities in ubiquitin-mediated processes are involved in several pathological conditions, including malignant transformation (Hershko and Ciechanover 1998). At the UBB locus, a seemingly small increase in methylation led to drastic downregulation of the gene. Although we cannot completely explain this observation, it may be that the expression observed in the parental cells derives from a relatively few (< 10%) unmethylated alleles. Nevertheless, treatment with DAC led to an upregulation of gene expression that was proportional to the decrease in DNA methylation. Even in the parental line, which is already substantially methylated, treatment with DAC results in demethylation and a 6-fold upregulation of
the gene. These data are consistent with the idea that many of the sites that become hypermethylated in response to \textit{IDH1} mutation are already partially methylated and repressed in the parental cell line.
4. Summary and future directions

4.1 Summary and conclusions

The glioma genome displays remarkable chromosomal aberrations, including point mutations, deletions, and amplifications, which activate or inactivate critical glioma-specific genes contributing to several oncogenetic pathways. Here, we describe the identification and molecular characterization of novel recurrent genomic alterations which likely contribute to the development and progression of gliomas.

We first describe a multifaceted genome-wide analysis to characterize the most significant copy number aberrations occurring in glioblastomas. We performed copy number analysis of 111 glioblastomas by Digital Karyotyping and Illumina BeadChip assays and validated our findings using data from the TCGA (The Cancer Genome Atlas) glioblastoma project. From this study, we identified recurrent focal copy number alterations in 1p36.23 and 4p16.3. Expression analyses of genes located in the two regions revealed genes which are dysregulated in glioblastomas. Specifically, we identify EGFR negative regulator, ERRFI1, within the minimal region of deletion in 1p36.23. In glioblastoma cells with a focal deletion of the ERRFI1 locus, restoration of ERRFI1 expression slowed cell migration. Furthermore, we demonstrate that TACC3, an Aurora-A kinase substrate, on 4p16.3, displays gain of copy number, is overexpressed in a glioma-grade-specific pattern, and correlates with Aurora kinase overexpression in glioblastomas.
Next, we examine the cellular effects of monoallelic point mutations of the NADP+-dependent isocitrate dehydrogenase, IDH1. IDH mutations occur frequently in gliomas and display robust association with specific DNA hypermethylation signatures. Here we show that heterozygous expression of the IDH1<sup>R132H</sup> allele is sufficient to induce the genome-wide alterations in DNA methylation characteristic of these tumors. Using a gene targeting approach, we knocked-in a single copy of the most frequently observed IDH1 mutation, R132H, into a human cancer cell line and profiled changes in DNA methylation at over 27,000 CpG dinucleotides relative to wild-type parental cells. We find that IDH1<sup>R132H/WT</sup> mutation induces widespread alterations in DNA methylation, including hypermethylation of 2,010 and hypomethylation of 842 CpG loci. We demonstrate that many of these alterations are consistent with those observed in IDH1-mutant and G-CIMP+ primary gliomas and can segregate IDH wild-type and mutated tumors, as well as those exhibiting the G-CIMP phenotype in unsupervised analysis of two primary glioma cohorts. Further, we show that the direction of IDH1<sup>R132H/WT</sup>-mediated DNA methylation change is largely dependent upon pre-existing DNA methylation levels, resulting in depletion of moderately methylated loci. Additionally, whereas the levels of multiple histone H3 and H4 methylation modifications were globally increased, consistent with broad inhibition of histone demethylation, hypermethylation at H3K9 in particular accompanied locus-specific DNA hypermethylation at several genes downregulated in IDH1<sup>R132H/WT</sup> knock-in cells.
Together, our multifaceted genomic evaluation provides insight on three key genomic alterations involved in gliomas. We establish \textit{ERRFI1} as a potential candidate tumor suppressor gene targeted by deletions of 1p36. Our genetic and biological data, as well as data from other studies, suggest that \textit{ERRFI1} is a major component in the receptor tyrosine kinase signaling pathway, one of the core signaling pathways involved in glioblastoma development. Also, we identify \textit{TACC3} as a potential oncogene activated by gains of 4p16. In our study, we found a strong correlation between Aurora kinase and \textit{TACC3} expression, indicating that in the Aurora kinase/TACC pathway may contribute synergistically to glioblastoma pathogenesis. Most strikingly, our data provide insight on epigenetic alterations induced by \textit{IDH1} mutations and support a causal role for \textit{IDH1^{R132H/WT}} mutants in driving epigenetic instability in human cancer cells. Together with previous genetic and functional analyses of IDH mutants in cancer cells, our data provide additional evidence that mutations of \textit{IDH1} are likely one of the major driving events in glioma.

\subsection*{4.2 Future studies}

The high frequency and specificity of the glioma-targeted genomic alterations described here provide strong evidence for their importance in tumorigenesis. However, details of the precise oncogenic consequences and molecular mechanisms of these mutations remain unclear. Future studies of these glioma-targeted genes and their
affected cellular processes and signaling pathways will be critical to understanding the malignant properties of these deadly cancers.

Here we show that a single base-pair mutation in IDH1 can cause dramatic effects on the epigenome. However, mechanistic details of this phenotype remain to be determined. We hypothesize that the accumulation of D-2-HG resulting from monoallelic mutations of IDH1 contributes to altered epigenetic phenotypes through the inhibition of multiple Fe(II)/2-oxoglutarate-dependent dioxygenases (Figure 24). This superfamily of enzymes is involved in a wide range of biological functions, including DNA repair and chromatin modification and includes the Jumonji-C domain family of histone demethylases (JHDMs) and the TET family of methylcytosine hydroxylases. At present, the extent to which D-2-HG-mediated inhibition of one or more α-KG-dependent enzymes is responsible for the pathogenesis of IDH mutation-associated tumors remains to be determined. Future studies aimed at the integrated analyses of histone methylation and DNA (hydroxy)methylation at a genome-wide scale will be necessary to resolve the relative contribution to IDH mutation-associated cancers.

Additionally, further studies are needed to fully understand the pathways targeted by IDH-induced epigenetic alterations, as well as the cellular contexts in which they function to drive tumorigenesis.
Single base-pair mutations of IDH1 result in neomorphic enzymatic production of D-2-hydroxyglutarate (D-2-HG). Accumulation of D-2-HG has the potential to impact the enzymatic activity of multiple Fe(II)/2-oxoglutarate-dependent dioxygenases, including the Jumonji-C domain family of histone demethylases (JHDMs), and the TET family of methylcytosine hydroxylases. Alterations of these enzyme families are hypothesized to result in broad-scale epigenetic alterations in cancer cells.
In addition to understanding the mechanistic effects of genetic and epigenetic alterations in cancer cells, it is important to understand whether these molecules are effective targets for oncogenic pathway-based therapy. For instance, future studies should explore the feedback inhibition pathways involved in receptor tyrosine kinase signaling and further evaluate whether exploiting feedback inhibitors such as ERRFI1 may be able to impact tumor growth. Also, regulators of the mitotic spindle apparatus, including Aurora kinases and its substrates, including TACC3, are attractive targets for small-molecule therapeutics and their impact on genome integrity should be further explored. Another intriguing question remaining is whether the epigenetic alterations resulting from IDH mutants and/or other epigenetic modifiers mutated in cancer could serve as targets for molecular-based intervention. Ultimately, understanding how genetic and epigenetic alterations work together to drive the expression of critical cancer genes is a major remaining challenge, but will yield insights to benefit patients in the future.
Appendix A

Figure 1 was reproduced from (Duncan and Yan 2011), and Section 1.2 is an excerpt reproduced from (Duncan and Yan 2011). According to the Cell Cycle policy (http://www.landesbioscience.com/docs/JournalCopyright.pdf), use of all or part of the contribution, or reuse of figures, is permitted in any dissertation written by the authors without first requiring permission from Landes, provided that the authors first acknowledge and reference publication in the journal. Figures 2 through 8 were reproduced from (Duncan et al. 2010), and Sections 2.1 through 2.4 are excerpts reproduced from (Duncan et al. 2010). According to the Oncotarget policy, reproduction of these figures and excerpts is permitted under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/3.0/), provided that the original author and source are credited. Figures 9 through 23 were reproduced from (Duncan et al. 2012), and Sections 3.1 through 3.4 are excerpts reproduced from (Duncan et al. 2012). According to the Genome Research policy, six months after publication authors may reproduce or reuse their article under the terms of a Creative Commons License (http://creativecommons.org/licenses/by-nc/3.0/) provided its original publication in Genome Research is fully acknowledged.
References


relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17(1): 98-110.


Biography

Christopher Gentry Duncan was born on November 27, 1982 in Raleigh, North Carolina, to Kim G. Duncan and Anita H. Duncan. He attended North Carolina State University and received a B.S. in Chemical Engineering and a B.S. in Paper Science & Engineering in 2005. After attaining his Bachelor’s degrees, Christopher developed an interest in the biological sciences and joined the laboratory of Dr. Hai Yan at Duke University Medical Center as a Research Technician. Two years later, he joined the Department of Pathology Graduate Program, continuing his scientific training in the Yan lab. At Duke, Christopher worked to understand the genetic and epigenetic aberrations which are involved in the pathogenesis of brain tumors. Christopher is an Associate Member of the American Association for Cancer Research and has presented abstracts at four national scientific conferences. His publications are as follows:


