Genomic dynamics associated with malignant transformation in IDH1 mutated gliomas

Chul-Kee Park[1,*], Inho Park[2,*], Seungmook Lee[3,*], Choong-Hyun Sun[3], Youngil Koh[4], Sung-Hye Park[5], Ja Eun Kim[1], Hongseok Yun[3], Se-Hoon Lee[6]

[1]Department of Neurosurgery, Seoul National University Hospital, Seoul, Korea

[2]SD Genomics, Seoul, Korea

[3]Samsung SDS, Seoul, Korea

[4]Department of Internal Medicine, Seoul National University Hospital, Seoul, Korea

[5]Department of Pathology, Seoul National University Hospital, Seoul, Korea

[6]Division of Hematology-Oncology, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea

[*]These authors have contributed equally to this work

Correspondence to: Se-Hoon Lee, **e-mail:** sehoon.lee119@gmail.com
            Hongseok Yun, **e-mail:** hongseok.yun@samsung.com

ABSTRACT

The genomic mechanism responsible for malignant transformation remains an open question for glioma researchers, where differing conclusions have been drawn based on diverse study conditions. Therefore, it is essential to secure direct evidence using longitudinal samples from the same patient. Moreover, malignant transformation of IDH1-mutated gliomas is of potential interest, as its genomic mechanism under influence of oncometabolite remains unclear, and even higher rate of malignant transformation was reported in IDH1-mutated low grade gliomas than in wild-type IDH1 tumors. We have analyzed genomic data using next-generation sequencing technology for longitudinal samples from 3 patients with IDH1-mutated gliomas whose disease had progressed from a low grade to a high grade phenotype. Comprehensive analysis included chromosomal aberrations as well as whole exome and transcriptome sequencing, and the candidate driver genes for malignant transformation were validated with public database. Integrated analysis of genomic dynamics in clonal evolution during the malignant transformation revealed alterations in the machinery regulating gene expression, including the spliceosome complex (U2AF2), transcription factors (TCF12), and chromatin remodelers (ARID1A). Moreover, consequential expression changes implied the activation of genes associated with the restoration of the stemness of cancer cells. The alterations in genetic regulatory mechanisms may be the key factor for the major phenotypic changes in IDH1 mutated gliomas. Despite being limited to a small number of cases, this analysis provides a direct example of the genomic changes responsible for malignant transformation in gliomas.

INTRODUCTION

One of the classic concepts of cancer progression includes an evolutionary process that results from stepwise mutations with sequential subclonal selection [1]. This evolutionary process in cancer is still applicable despite traditional cancer treatment strategies that involve artificial alterations in cancer-clone dynamics [2]. However, this concept did not have direct evidence until recent advances in next-generation sequencing and bioinformatics techniques allowed direct observation of this concept [3, 4]. Clonal evolution in cancer occurs through the interaction of genomic changes of advantageous driver lesions, neutral passenger lesions, and disadvantageous lesions [2]. It is accepted that the identification of driver lesions is based on increased frequency in multiple tumors

compared to normal background. Furthermore, once a genetic alteration is identified as a driver in one tumor type, that event can be more reliably interpreted even if it is infrequent [2, 5]. However, the dynamics of somatic evolution in cancer is a complex process involving the interaction of the accumulation of mutations and clonal expansions. Moreover, mutagenesis in cancer cells varies by different types of genomic abnormality, ranging from small-scale aberrations to large-scale genome events. Therefore, comprehensive analysis using a polygonal approach to genetic data in longitudinal samples is essential for understanding cancer progression. Previous studies investigated clonal dynamics in cancer progression using longitudinal samples with modern sequencing technology [3, 6–14]. However, most of the studies are from hematological malignancies or systemic metastasis of solid tumors. Unexpectedly, direct evidence of genomic dynamics of malignant transformation in glioma using longitudinal samples is rare.

The literature reports that malignant transformation of low-grade gliomas (LGGs) occurs in 45% of oligodendrogliomas, 70% of oligoastrocytomas, and 74% of astrocytomas [15]. Strong evidence shows that somatic mutation of isocitrate dehydrogenase 1 (IDH1) is related to a better prognosis in LGGs [16]. On the other hand, notwithstanding a longer latent time before malignant transformation in IDH1-mutated LGGs, it has been proposed that a higher rate of malignant transformation occurred in IDH1-mutated LGGs than in wild-type IDH1 tumors [17]. However, IDH1 mutation itself may not be the driving force for malignant transformation [18]. A recent genetic study using 23 glioma patients with longitudinal samples showed that tumor progression is related to a broad spectrum of genetic changes that cannot be explained by a simple genetic event [19]. Here, we investigated 3 pairs of longitudinal samples of IDH1-mutated LGGs and malignant transformations from the 3 samples with minimal artificial treatment effect, to seek the driver genes responsible for malignant transformation. To do this, we employed a comprehensive interpretation of multiple genomic data, including analysis of chromosomal aberrations as well as whole exome and transcriptome sequencing.

RESULTS

Landscape view of genomic changes of longitudinal samples of LGGs

We studied 3 pairs of IDH1 mutated low grade gliomas and their high grade phenotype transformed after the lapse of time (Figure 1). The histological diagnosis were made after 2007 WHO classification criteria. Case 1 is histologically classified as astrocytoma with IDH1 mutation and intact 1p19q status, which progressed to anaplastic astrocytoma followed by glioblastoma.

Case 2 and 3 are oligodendrogliomas with IDH1 mutation and 1p19q co-deletion, which progressed to anaplastic oligodendroglioma. Preservation of original tumor cells, after initial histological diagnosis by biopsy only or partial resection, enabled us to observe their clonal evolution during the progression of tumor.

To profile genomic changes during malignant transformation, whole exome sequencing (WES) was performed with pairs of low grade and high grade tumor samples, as well as with normal DNA from white blood cells. Using a paired-end sequencing strategy, nonsynonymous somatic point mutations and small insertions/deletions (indels) that change the protein amino acid sequence were identified from WES results, and gene sets were built to enable comparison between low grade and high grade counterparts in patient samples (*Dataset 1*). The number of mutated genes identified from WES is summarized in Table 1. From the standpoint of changes in the number of nonsynonymous mutations, major genetic dynamics were significantly altered during the malignant transformation process in Cases 1 and 2, while genetic changes were relatively stable in Case 3 (Figure 2). We also detected a number of somatic copy number alterations (SCNAs) within the tumor samples by using single nucleotide polymorphism (SNP) array and WES data. The segments with heterozygous deletion are of primary interest because they can be further utilized for inferring clonal dynamics of tumors. We summarize the DNA copy number alterations with the estimated allele specific copy number status and cellular fraction in *Dataset 2*. We visualized the whole-genome alterations of each sample in Circos and Manhattan-style plots for an overall view of genomic changes related to malignant transformation (Figure S1).

Clonal evolution during the malignant transformation

To track clonal evolution within tumors, we developed a modeling system that included a stratified grouping of SCNAs and somatic point mutations based on the estimated fraction of cancer cells harboring each alteration in a sample (Table S1). Using this modeling system, clonal fractions in tumors were estimated from SCNAs in each sample from the 3 cases (Table 2). In addition, we analyzed clonal fractions in tumors by clustering somatic point mutations based on the estimated fraction of cancer cells harboring each point mutation in a sample, which are calculated from the mutation allele frequency and copy number status of the segment containing the mutation of interest (Figure 3). Combining SCNAs and mutation analysis, lists of genes were filtered to define the subclones involving malignant transformation. The filtered genes are summarized and categorized according to their changes in status during the tumor progression (Figure 4). Collectively, we built
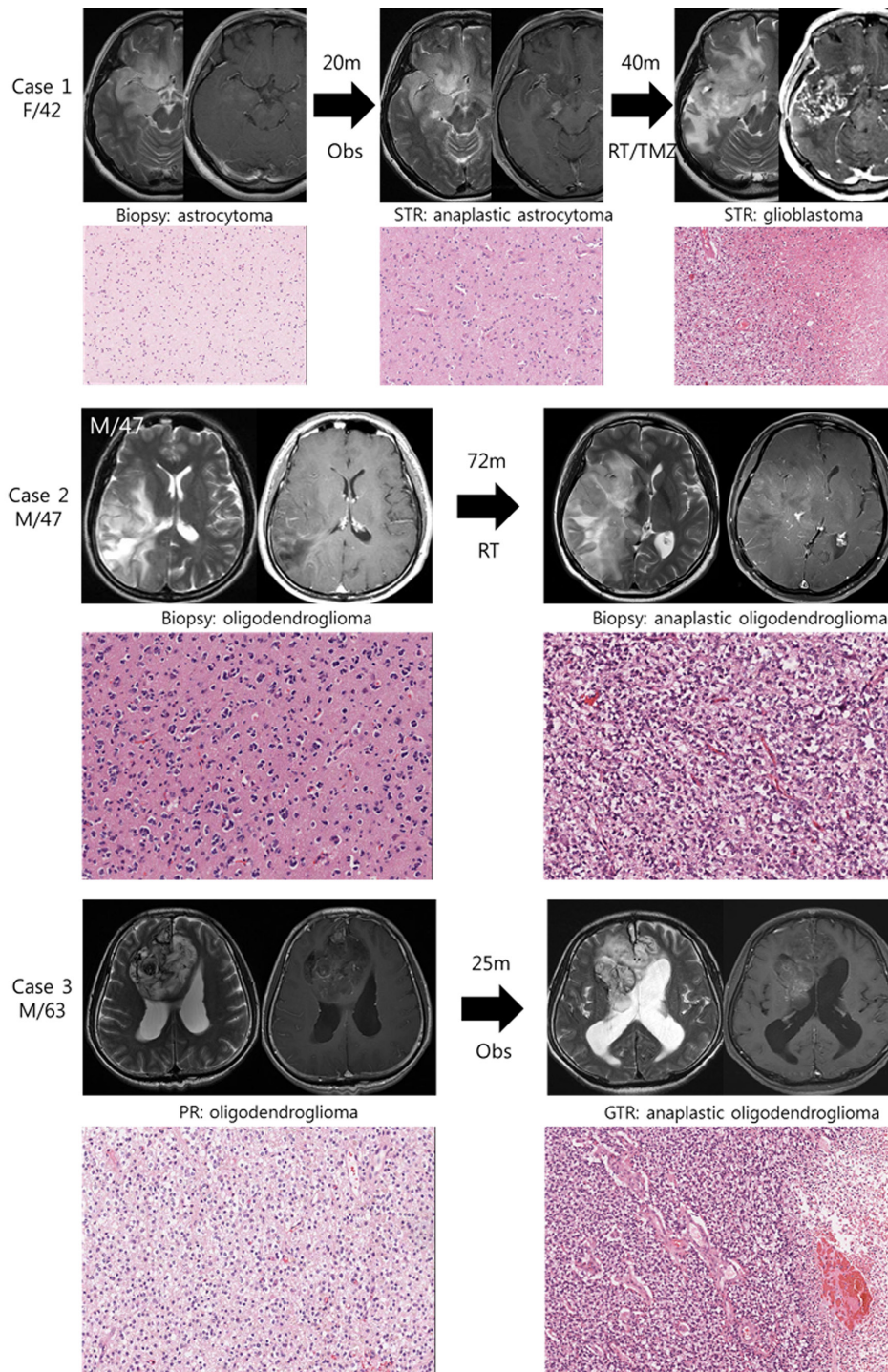
Figure 1: Illustrative history of patients. Case 1. A 42-year-old female patient who was initially diagnosed with diffuse astrocytoma in right fronto-temporal area after stereotactic biopsy. At the option of the patient, she was monitored without any treatment and was in stable state for 20 months. However, enhancing area in magnetic resonance images (MRI) was developed in temporal lobe lesion, and subtotal resection was performed. The diagnosis was anaplastic astrocytoma, and she was treated with radiotherapy followed by 6 cycles of temozolomide chemotherapy. Forty months after the second surgery, progression of the disease was observed and the diagnosis of glioblastoma was confirmed after the third surgery. A pair of diffuse astrocytoma and glioblastoma samples were used in this study. **Case 2.** A 47-year-old male patient with right temporo-parietal mass was diagnosed with oligodendroglioma after stereotactic biopsy and was treated with standard radiotherapy alone. The mass was aggravated to frontal area after 72 months of stable period, and the anaplastic oligodendroglioma was confirmed after stereotactic biopsy. **Case 3.** A 63-year-old male patient with heterogeneous bifrontal mass lesion was diagnosed with oligodendroglioma after partial resection, and the remaining mass was observed over period. After 25 months of stable period, the disease started to progress and the second surgery was done. The histological diagnosis was confirmed to be anaplastic oligodendroglioma after gross total resection.

Table 1: Sequencing coverage parameters and number of genes profiled and analyzed from whole exome sequencing

|	Case 1			Case 2			Case 3		
	Normal	Low grade	High grade	Normal	Low grade	High grade	Normal	Low grade	High grade
Mean depth	153	156	152	156	160	159	170	159	157
10X	91.3%	91.1%	90.5%	91.9%	91.8%	91.9%	93.1%	91.5%	91.6%
30X	83.3%	83.3%	80.8%	84.6%	83.8%	84.0%	86.3%	83.1%	83.3%
50X	75.5%	76.2%	71.8%	77.7%	76.1%	76.5%	80.1%	75.3%	75.5%
100X	56.2%	56.9%	51.5%	58.0%	57.4%	57.6%	62.5%	56.6%	56.6%
Number of nonsynonymous somatic mutations									
Missense	-	25	52	-	7	41	-	15	18
Start Lost	-	0	0	-	1	1		0	0
Stop Gain	-	4	4	-	0	0	-	0	0
Stop Loss	-	1	1	-	0	0	-	0	0
Codon InDel	-	0	2	-	0	0	-	0	0
Frame Shift	-	0	6	-	1	6	-	3	2
Number of nonsynonymous mutational changes during the tumor progression									
newly developed		49			41			9	
disappeared		14			2			7	
preserved		16			7			11	
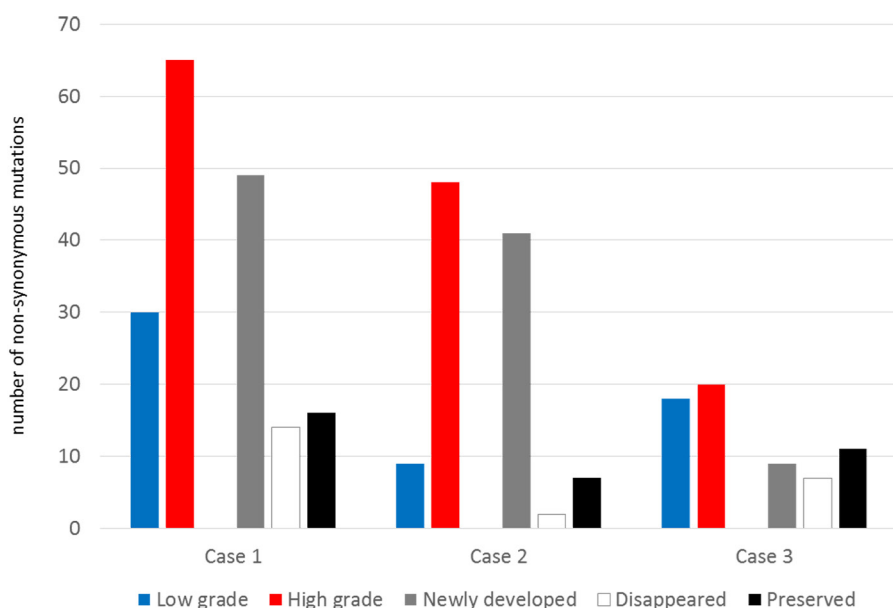
Figure 2: The number of non-synonymous mutations and their dynamics during the malignant transformation.

Table 2: Estimated clonal fractions of each tumor sample analyzed from somatic copy number alteration (SCNA)

Group	Clonal fraction		CNA segments
	Low grade	High grade	
Case 1			
S1.1	64%	84%	4q13–4q35; 11p15; 19q13
S3.1		84%	1p36; 5p15–5q35; 8p23–8p23; 8p22–8q11; 9p24–9p23; 10p12–10q26; 10p12–10q26; 11q13–11q25; 13p13–13q34; 16q11–16q24; 20q11–20q13; 22p13–22q13; Xp22-Xq28
S3.2		43%	1p36–1q44; 3p26–3p21; 6q13–6q27; 12p13–12q24; 21p13–21q22
Case 2			
S1.1	60%	89%	1p36–1p11; 19q11–19q13
S3.1		88%	4q12–4q24; 14q11–14q32; 17p13–17p11
S3.2		52%	18p11–18q23
Case 3			
S1.1	88%	90%	1p36–1p11; 19q11–19q13

a model for the genomic dynamics of clonal evolution during the malignant transformation in each case of IDH1-mutated glioma (Figure 5).

Validation of candidate genes for malignant transformation

To validate genes involved in malignant transformation that were chosen for the present analysis, we used datasets from The Cancer Genome Atlas (TCGA) (https://tcga-data.nci.nih.gov/tcga/) and the cBioPortal for Cancer Genomics (http://www.cbioportal.org/public-portal/). Case sets were built from the brain lower-grade glioma (provisional) and glioblastoma (provisional) datasets. Among 262 cases from lower-grade glioma and 235 cases from glioblastoma with complete sequencing and CNA data, a total of 216 cases with an IDH1 mutation and available information about histological grade were analyzed (Table S2). We tested all the genes that showed changes during the malignant transformation in our cases. Only genes with novel or additional mutations and/or CNAs observed repeatedly in the stratified subgroup of 1p19q chromosomal status and histological grade were highlighted (Figure 6). As verified recently, TP53 and ATRX mutations were the hallmark of 1p19q intact IDH1-mutated gliomas, while CIC and FUBP1 mutations were found in 1p19q co-deleted gliomas [20–22]. However, it is worth noting that TP53 and ATRX mutations were observed only in a small fraction of 1p19q co-deleted gliomas of high grade, which implies that these types of trans-lineage mutations can contribute to the malignant transformation that was also observed in Case 2. Recent observation of changes in TP53 expression in sequential samples of oligodendrogliomas supports that the de novo TP53 mutation or the proliferation of a subset of cells with nuclear expression of TP53 could lead to tumor progression in some IDH1-mutated oligodendroglial tumors [17]. Among the genes that showed differences in incidence of mutation or CNA among grades, U2AF2 (also known as U2AF65), TCF12 (also known as HEB, HTF4 and ALF1), and ARID1A (also known as BAF250a) were commonly observed to be altered progressively in both 1p19q intact and co-deleted tumors. These genes are components of the machinery that regulates gene expression, including the spliceosome complex (U2AF2), transcription factors (TCF12), and chromatin remodelers (ARID1A).

U2AF2 had a copy number loss in case 3 from the low-grade stage and developed additional missense mutations at the high-grade stage. And in TCGA samples, 5.9% of grade 3 gliomas with IDH1 mutation/1p19q co-deletion harbored mutations in U2AF2 while no mutations were found in grade 2 gliomas with the same molecular signature (Figure 6). Moreover, in IDH1 mutation/1p19q intact gliomas subjected to TCGA, there was increase tendency of copy number loss with WHO grade (4.3%, 10.7% and 25.0% in grade 2, 3, and 4, respectively). A novel frameshift deletion in TCF12 was found in the high grade sample of case 1, and mutations were observed in 4.7% of grade 2 and 5.9% of grade 3 of IDH1 mutation/1p19q co-deletion samples in TCGA (Figure 6). Again, we observed a stepwise increase of mutation incidence with WHO grade (1.4%, 3.6%, and 8.0% in grade 2, 3, and 4) (Figure 6C). All these mutations generate either frameshifts or occur at splice-sites which suggest loss-of-function mechanisms correlating with lower expression [23]. Although a mutation in low grade phase was already present, newly developed copy number
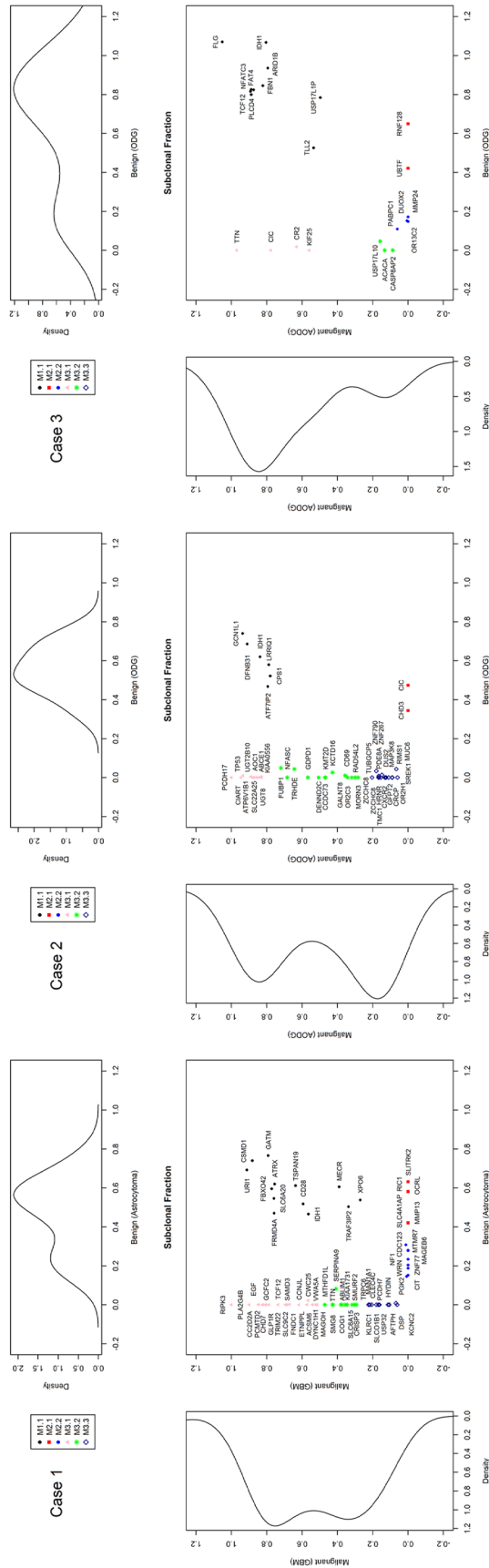
Figure 3: Estimated clonal fractions and their genetic signatures based on somatic point mutation allele frequency and copy number status.

loss of ARID1A was found in the high grade sample of case 1. TCGA data shows that increase in mutation rate in high grade phenotype was observed in both 1p19q co-deleted and intact gliomas with IDH1 mutation (Figure 6). Moreover, copy number loss was accompanied by 1p19q co-deleted tumors, but there was no mutation in grade 4 GBMs as observed previously [5].

Transcriptomal changes involved in the malignant transformation

Two pairs of tumor samples (Cases 2 and 3) were analyzed with RNA-sequencing (RNA-seq). The results of the RNA-seq analysis are summarized (Table S3), and there were minimal differences in read counts among the samples. After we quantified gene expression levels as RPKM (Reads Per Kilobase per Million mapped reads) from the RNA-seq data, we performed fold change analysis between the low grade

and high grade samples for each case. To define a group of differentially expressed genes (DEGs) in each case, we used a log2 fold change cut off of > 3 (*Dataset 3* and Figure S2). Among the DEGs between low grade and high grade phenotypes in each case, HBB and HBA1 were commonly overexpressed in high grade samples of both cases. The low number of common DEGs originates from the low number of DEGs in Case 3. This implies that, although the histological diagnosis distinguishes the tumor grades, Case 3 was on track for the early phase of malignant progression at the time of the first surgery, which is also suggested by the relatively short span of time before recurrence (25 months) and the stable status of mutation frequency (Figure 2). The hierarchical clustering analysis also supports the relatively similar genomic signatures between low grade and high grade samples in Case 3 compared with those of Case 2 (Figure S3). So, we focused on Case 2 to evaluate the expression changes that are responsible for the malignant
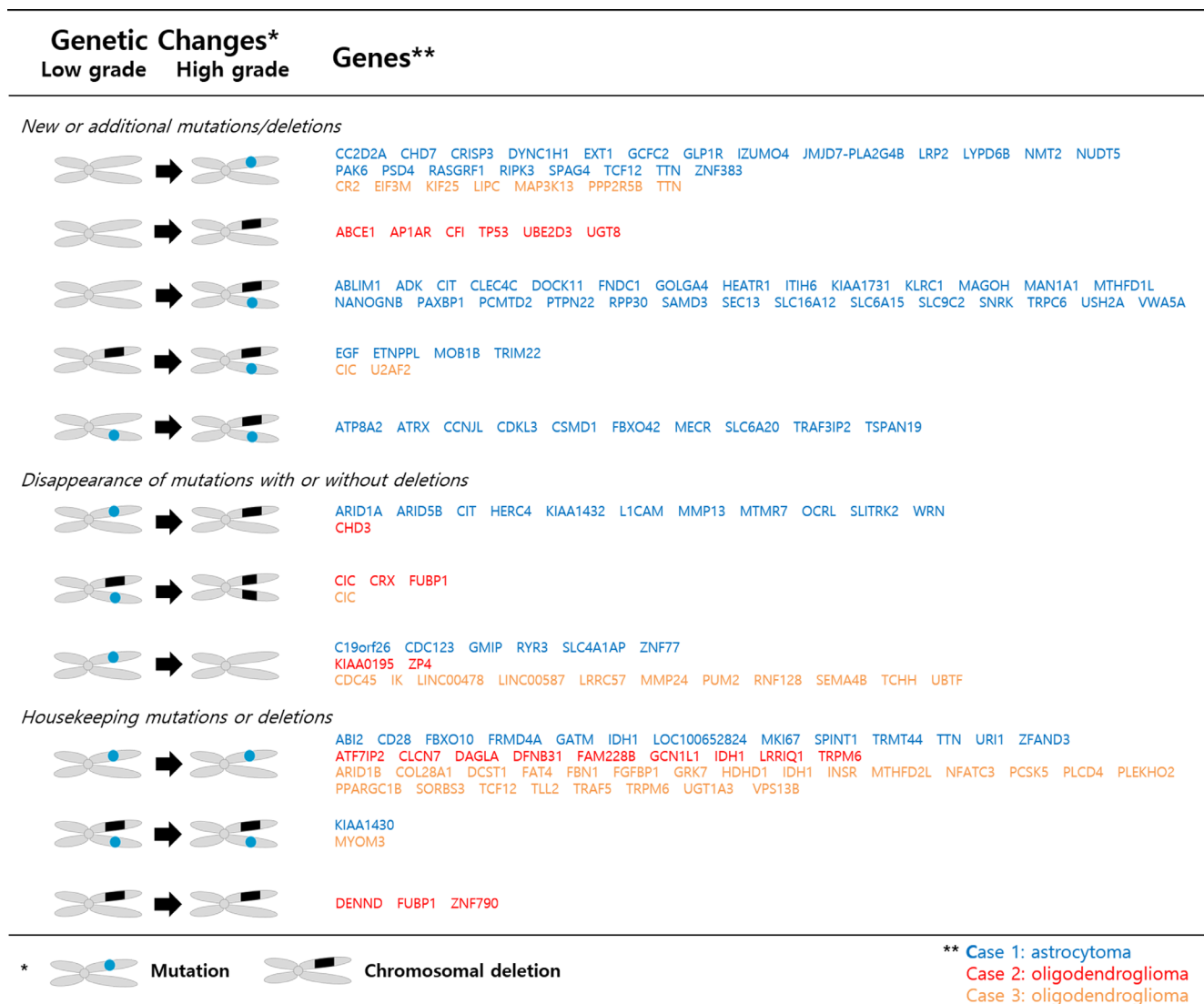

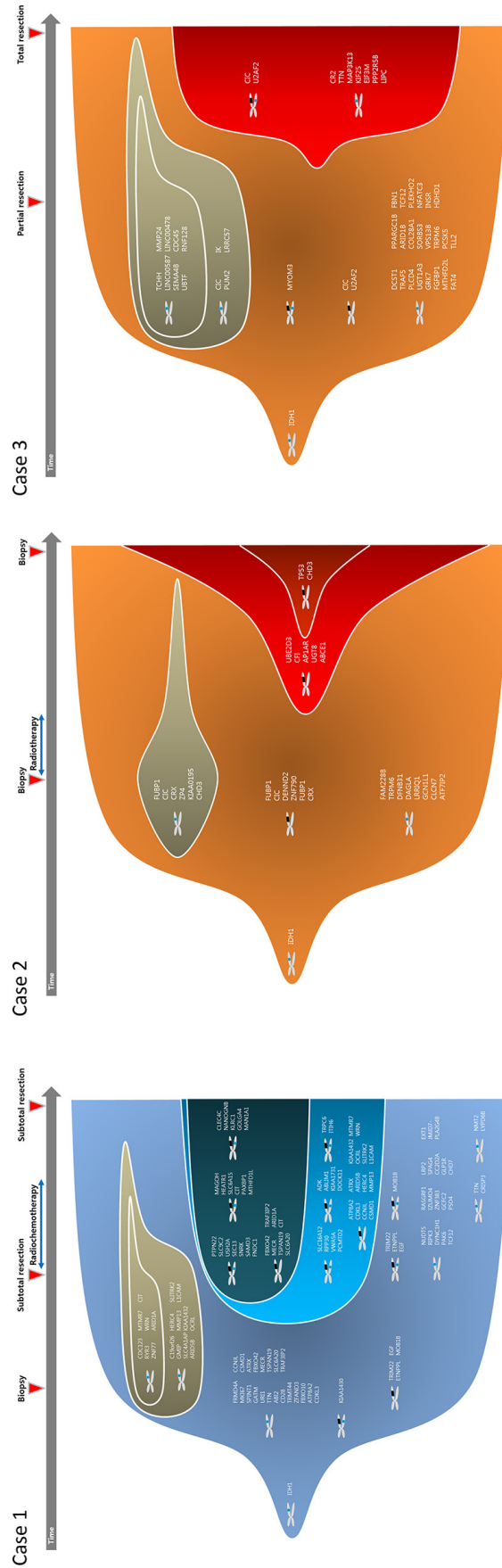**Figure 4: Subclone defining genes involved in genetic changes during malignant transformation.**

**Figure 5: Clonal evolution from low grade to high grade gliomas.**

transformation. Among the overexpressed genes in the high grade phenotype, the notable genes are OLIG1, OLIG2, VGF, SOX4, SOX8, MYT1, and PDGFRA, which are known to regulate oligodendrogenesis [24]. Recent studies suggest that overexpression of these genes could be used as a representative feature of a specific subtype of glioma [25]. We performed a gene set enrichment analysis (GSEA) with the DEGs in Case 2, using C2 category gene sets in MSigDB [26, 27]. The analysis identified 7 overexpressed and 2 down-regulated gene sets in the malignant phenotype, which are significantly enriched at the nominal value of FDR q < 1.0e-10 (Table 3). Interestingly, genes that are down-regulated during differentiation of the oligodendroglial precursor (Gene set: GOBERT_OLIGODENDROCYTE_ DIFFERENTIATION_DN) were reactivated during the malignant transformation, indicating that malignant transformation accompanies the restoration of stemness of the cancer cells (Figure S4) [28]. This is also supported by the overexpression of genes normally enriched in

embryonic stem cells (Gene Set: BENPORATH_SUZ12_ TARGETS) in the high grade phenotype (Figure S4) [29]. Identification of genes with high-CpG-density promoters bearing histone H3 dimethylation at K4 (H3K4me2) and trimethylation at K27 (H3K27me3) (Gene set: MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_ H3K27ME3) from embryonic stem-cell-derived neural precursor cells in the malignant phenotype provides further supporting evidence for the restoration of stemness (Figure S4) [30]. Interestingly, the DEG from the high grade phenotype share a common genetic signature with a proneural type of glioblastoma (Gene set: VERHAAK_GLIOBLASTOMA_PRONEURAL), which is distinguished from other glioblastomas by lower age, better prognosis, PDGFRA expression, and frequent IDH1 mutation (Figure S4) [31]. Using a recently suggested glioma classification module based on genes related to EGFR or PDGFRA expression, the PDGFRA signature became more evident with malignant transformation in Case 2 (Figure S5) [25].
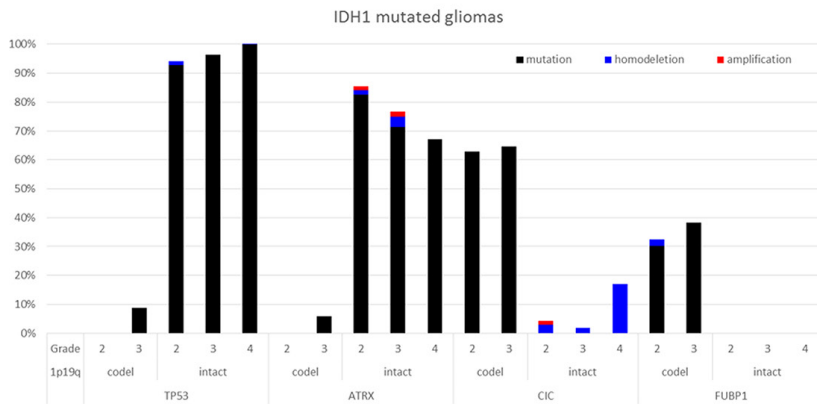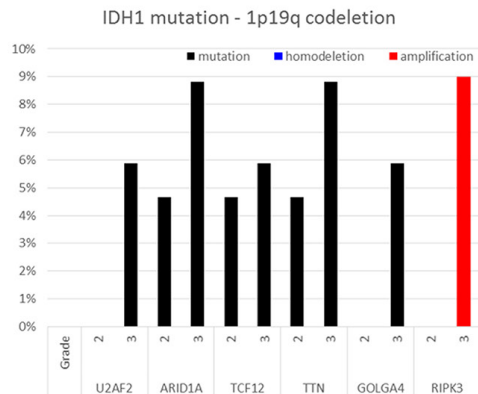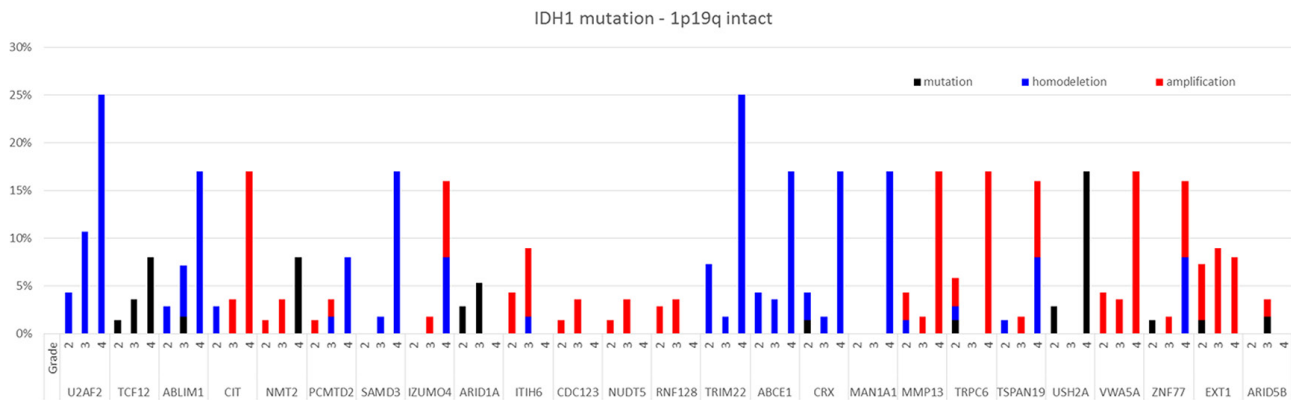

Figure 6: Frequency of genetic alterations of genes showing disparity according to the histological grade in the TCGA database among those with novel or additional mutations and/or copy number alterations observed in the present study.

Table 3: Gene set enrichment analysis for differentially expressed genes in low grade and high grade phenotype of case 2

	Gene Set Name	# Genes in Gene Set (K)	# Genes in Overlap (k)	k/K	Enrichment score (ES)	*p*-value	FDR *q*-value
A. Gene sets enriched with genes over-expressed in high grade phenotype							
(a)	MILI_PSEUDOPODIA_HAPTOTAXIS_DN	668	41	0.0614	0.2736	8.04E–22	3.80E–18
(b)	GOBERT_OLIGODENDROCYTE_DIFFERENTIATION_DN	1080	50	0.0463	0.1704	4.58E–21	1.08E–17
(c)	PATIL_LIVER_CANCER	747	31	0.0415	0.2521	3.19E–12	1.07E–9
(d)	MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	1069	47	0.0440	0.2492	5.88E–19	9.25–16
(e)	VERHAAK_GLIOBLASTOMA_PRONEURAL	210	21	0.1000	0.3899	6.36E–16	6.00E–13
(f)	REACTOME_BETA_DEFENSINS	42	11	0.2619	0.8475	9.75E–14	5.76E–11
(g)	BENPORATH_SUZ12_TARGETS	1038	43	0.0414	0.2278	1.70e–16	2.01E–13
B. Gene sets enriched with genes underexpressed in high grade phenotype							
(a)	BLALOCK_ALZHEIMERS_DISEASE_UP	1691	48	0.0284	–0.2388	3.69E–12	1.16E–9
(b)	LU_AGING_BRAIN_UP	262	17	0.0649	–0.3781	3.77E–10	8.09E–8

DISCUSSION

We have performed an integrated analysis of genomic dynamics in clonal evolution during the malignant transformation and identified alterations in the machinery regulating gene expression, including the spliceosome complex (U2AF2), transcription factors (TCF12), and chromatin remodelers (ARID1A). U2AF2 is a core member of the spliceosome machinery [32], so mutations in this gene can affect the normal function of spliceosomes resulting in the formation of aberrant mature mRNAs by misunderstanding of splice site recognition [33, 34]. This kind of abnormal processing may alter the expression of multiple genes. Mutations to spliceosome genes are related to hematological malignancy and its prognosis and can act as a driver of oncogenesis in colon cancer [34–38]. Since alternative splicing is observed in extensive numbers of genes from many different types of cancer, targeting spliceosome function may unlock a novel strategy for cancer therapy [32]. TCF12 encodes a transcription factor of the basic helix-loop-helix (bHLH) E-protein family that can directly bind to E-box motfis [39, 40]. TCF12 is associated with proliferation, survival, and fate decisions in the oligodendrocyte lineage [41]. Interestingly, significant differences in TCF12 expression between 1p19q co-deleted tumors and intact tumors (higher with 1q19q codeletion) as well as among WHO grades (highest in grade 2 and lowest in grade 4) were reported previously [42]. However, whether or not TCF12 plays a role in cancer development and progression has not been determined yet except for the contradictory evidence in colorectal cancer [40, 43]. ARID1A is recurrently mutated in various cancer types [44]. ARID1A is a member consisting of SWI/SNF chromatin remodeling complexes from which disordered chromatin regulation can induce a distinct mechanism contributing to tumor development [44]. Originally, ARID1A loss was known to be an early cancer promoting event in endometriosis leading to ovarian clear cell carcinoma [45]. Evidences of discordance between expression and heterozygous mutations or loss of heterozygosity in cancer samples imply that reduced levels of ARID1A may be the contributing factor in promoting cancer [44]. Our data suggest that not only mutations, but also copy number alterations and expression of ARID1A should be investigated to confirm its oncogenic role in gliomas.

A recent sequencing based analysis of paired gliomas of low grade and their relapse has revealed that a wide spectrum of genomic dynamic exists during the tumor progression from linear clonal evolution to branched clonal evolution [19]. And they found that IDH1 was the only shared mutation among longitudinal samples in every patients [19]. That means it is difficult to identify common target gene that drive malignant transformation or tumor progression in gliomas, which is retold in the present study. It is notable that the genes of interest, drawn from comprehensive genomic analysis of malignant transformation using longitudinal samples, are regulating the components of multiple genes with diverse mechanisms involving spliceosome machinery, transcription factors, and chromatin remodelers. This suggests that alterations in genetic regulatory mechanisms may be the key factor for the major phenotypic changes in gliomas. Moreover, expression changes resulting from genomic alterations appear to activate genes associated with the restoration of stemness in cancer cells. Whether restoration of stemness is really occurring will require further investigation incorporating a large collection of longitudinal samples will provide more detailed and definite answers on this issue. The limitation of this study harbors the small number of cases, which is compensated with the incorporation of TCGA dataset in search of common gene of interest. However, a sufficient number of appropriate paired samples are needed to confirm the solid conclusion in the future. The other thing is that initial samples of low grade status might have not represented the characteristics of tumor on the whole as they were histologically diagnosed with only a spot biopsy. This issue also exhibits problem about intratumoral heterogeneity and may have act as a confounding factor for the mutational analysis.

MATERIALS AND METHODS

Patients and samples

A total of 3 pairs of snap-frozen samples of sequential low grade and high grade histology from 3 patients with IDH1-mutated gliomas were used for DNA and RNA extraction. DNA from white blood cells from the same patients were used as a control. This study was approved by the Institutional Review Board of Seoul National University Hospital, Seoul, Korea. Genomic DNA was extracted using the QIAamp DNA mini kit (Qiagen, Cat. No. 51304), and total RNA was extracted using RNeasy Plus Universal Mini Kit (Qiagen, Valencia, CA, USA, Cat no. 73404) according to the manufacturer's recommendations. DNA content was quantitated using the Qubit DNA quantification kit (Invitrogen, Carlsbad, CA), and DNA integrity was assessed by gel electrophoresis. Samples with an RIN (RNA Integrity Number) > 5 were selected for the study.

Single nucleotide polymorphism array

We applied a genome-wide SNP array (Illumina HumanOmini5-Quad BeadChip, Illumina) to genomic DNA according to the manufacturer's instructions. The SNP array data was processed with GenomeStudio to generate B allele frequencies (BAF) and then applied to allele-specific copy number analysis of tumors (ASCAT) [46] to estimate tumor purity and ploidy of tumor samples.

Whole exome sequencing

Whole exome sequencing (WES) using Agilent SureSelect Human All Exon 50Mb (Agilent Technologies Inc., Santa Clara, CA) was performed on genomic DNA, followed by sequencing with 100-bp paired end reads on the Illumina HiSeq platforms. After we aligned the raw sequencing reads to hg19 with BWA-MEM and preprocessed the initially aligned bam files using the work flow for data pre-processing steps from GATK Best Practices, we obtained final bam files with more than 150 times depth of coverage on target for all samples (Table 1).

Single nucleotide variants (SNVs) and indels were detected with three different variant callers (UnifiedGenotyper, LoFreq, and SNVer), and the resulting call-sets were filtered using false positive filters and germline filters. Recurrent cancer mutations were investigated separately to rescue mutations that could be missed by the above variant callers. Somatic variant calling was done by Fisher's exact test with p-value < 0.02 and odd ratio > 5.0 with read counts supporting reference allele and alternate allele at the variant position between a tumor and normal sample. Details of the somatic variant analysis pipeline are described in Figure S6.

Whole transcriptome sequencing

RNA-seq libraries construction was performed using the manufacturer's instructions, and RNA sequencing was done on the Illumina HiSeq platform with 100-bp paired end reads. For RNA sequencing data analysis, alignment was performed by TopHat with hg19 and GENCODE version 10 (Ensembl 65), and expression profiles were analyzed using the Reads Per Kilobase per Million mapped reads (RPKM) values. Details of data analysis are described in Figure S7.

Tumor purity and ploidy

Tumor purity and the average ploidy of the tumor samples were estimated from SNP array data by applying ASCAT [46], whose results are summarized in Table S4. The average ploidy of the tumor samples was estimated to be 1.8–2.1, except for the high grade sample from Case 2, whose ploidy was estimated to be 3.80. However, we found that the ASCAT algorithm overestimated the ploidy of the high grade sample from Case 2 because

the algorithm prefers to assign integer copy numbers to segments with heterozygous deletion in chromosomes 2 and 18, even though they are only likely to be altered in just under half of the cancer cells from the sample (Figure S8). Tumor purity assessment revealed 62–91% purity, which implies the tumor samples are eligible for this study.

Subclonal analysis

With SNP array and WES data, we could infer clonal architectures of tumor samples for a patient using the following steps. 1) We identified segments with heterozygous deletions whose size is greater than 10Mb in each tumor sample. Then, we grouped all the segments identified from all the tumor samples for a patient. The identified segments for each case is shown in *Dataset* 2.2. For each tumor sample, we calculated the fraction of a subclone harboring a segment with heterozygous deletion by utilizing two types of information: a) the normalized read count ratio between the normal and tumor sample within the segment, and b) the altered allele frequencies of germline heterozygous SNVs in a tumor sample within the segments. We qualified germline heterozygous SNVs as having allele frequencies between 0.3 and 0.7 in the normal sample. 3) Then, we infer the fraction of cells that harbor the copy number alteration on the segment of interest by using the method described in Figure S9. 4) We grouped the segments based on the inferred clonal fractions in each tumor samples, and each group represents different subclones in the tumor samples. 5) For somatic point mutations, we estimated the fraction of a subclone harboring a somatic SNV by considering the copy number status of the locus. We only considered SNVs within diploid and segments with heterozygous deletions. The estimations of cellular fraction for each somatic SNV are shown in *Dataset 1* (B. Frac and M. Frac columns in the excel sheet). And then, we grouped somatic SNVs based on the cellular fractions in tumor samples. Of note, the estimated fraction of subclones based on heterozygous deletions and on somatic SNVs is quite similar, showing the robustness of estimation of clonal architecture.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

Data availability

Genomic data used in this study are available at http://www.ncbi.nlm.nih.gov/bioproject/PRJNA276922

Editorial note

This paper has been accepted based in part on peer-review conducted by another journal and the authors' response and revisions as well as expedited peer-review in Oncotarget.

REFERENCES

1. Nowell PC. The clonal evolution of tumor cell populations. Science. 1976; 194:23–28.

2. Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012; 481:306–313.

3. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, McLellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'Laughlin M, Fan X, et al. Clonal architecture of secondary acute myeloid leukemia. The New England journal of medicine. 2012; 366:1090–1098.

4. Shah SP, Roth A, Goya R, Oloumi A, Ha G, Zhao Y, Turashvili G, Ding J, Tse K, Haffari G, Bashashati A, Prentice LM, Khattra J, Burleigh A, Yap D, Bernard V, et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature. 2012; 486:395–399.

5. Jones S, Li M, Parsons DW, Zhang X, Wesseling J, Kristel P, Schmidt MK, Markowitz S, Yan H, Bigner D, Hruban RH, Eshleman JR, Iacobuzio-Donahue CA, Goggins M, Maitra A, Malek SN, et al. Somatic mutations in the chromatin remodeling gene ARID1A occur in several tumor types. Human mutation. 2012; 33:100–103.

6. Song Y, Zhang Q, Kutlu B, Difilippantonio S, Bash R, Gilbert D, Yin C, O'Sullivan TN, Yang C, Kozlov S, Bullitt E, McCarthy KD, Kafri T, Louis DN, Miller CR, Hood L, et al. Evolutionary etiology of high-grade astrocytomas. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:17933–17938.

7. Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I, Girsberger S, Lehmann T, Passweg J, Stern M, Beisel C, Kralovics R Skoda RC. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. Blood. 2014; 123:2220–2228.

8. Haffner MC, Mosbruger T, Esopi DM, Fedor H, Heaphy CM, Walker DA, Adejola N, Gurel M, Hicks J, Meeker AK, Halushka MK, Simons JW, Isaacs WB, De Marzo AM, Nelson WG, Yegnasubramanian S. Tracking the clonal origin of lethal prostate cancer. The Journal of clinical investigation. 2013; 123:4918–4922.

9. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, Kamiyama M, Hruban RH, Eshleman JR, Nowak MA,

Velculescu VE, Kinzler KW, Vogelstein B, Iacobuzio-Donahue CA. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010; 467:1114–1117.

10. Castellarin M, Milne K, Zeng T, Tse K, Mayo M, Zhao Y, Webb JR, Watson PH, Nelson BH, Holt RA. Clonal evolution of high-grade serous ovarian carcinoma from primary to recurrent disease. The Journal of pathology. 2013; 229:515–524.

11. Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, Morsberger LA, Latimer C, McLaren S, Lin ML, McBride DJ, Varela I, Nik-Zainal SA, Leroy C, Jia M, Menzies A, et al. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. Nature. 2010; 467:1109–1113.

12. Vermaat JS, Nijman IJ, Koudijs MJ, Gerritse FL, Scherer SJ, Mokry M, Roessingh WM, Lansu N, de Bruijn E, van Hillegersberg R, van Diest PJ, Cuppen E, Voest EE. Primary colorectal cancers and their subsequent hepatic metastases are genetically different: implications for selection of patients for targeted treatment. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012; 18:688–699.

13. Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J, Steidl C, Holt RA, Jones S, Sun M, Leung G, Moore R, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. Nature. 2009; 461:809–813.

14. Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJ, Witt H, Croul S, Bouffet E, Fults DW, Eberhart CG, Garzia L, Van Meter T, Zagzag D, Jabado N, Schwartzentruber J, Majewski J, et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. Nature. 2012; 482:529–533.

15. Jaeckle KA, Decker PA, Ballman KV, Flynn PJ, Giannini C, Scheithauer BW, Jenkins RB, Buckner JC. Transformation of low grade glioma and correlation with outcome: an NCCTG database analysis. Journal of neuro-oncology. 2011; 104:253–259.

16. Goze C, Blonski M, Le Maistre G, Bauchet L, Dezamis E, Page P, Varlet P, Capelle L, Devaux B, Taillandier L, Duffau H, Pallud J. Imaging growth and isocitrate dehydrogenase 1 mutation are independent predictors for diffuse low-grade gliomas. Neuro-oncology. 2014; 16:1100–1109.

17. Kanamori M, Kumabe T, Shibahara I, Saito R, Yamashita Y, Sonoda Y, Suzuki H, Watanabe M, Tominaga T. Clinical and histological characteristics of recurrent oligodendroglial tumors: comparison between primary and recurrent tumors in 18 cases. Brain tumor pathology. 2013; 30:151–159.

18. Juratli TA, Peitzsch M, Geiger K, Schackert G, Eisenhofer G, Krex D. Accumulation of 2-hydroxyglutarate is not a biomarker for malignant progression in IDH-mutated low-grade gliomas. Neuro-oncology. 2013; 15:682–690.

19. Johnson BE, Mazor T, Hong C, Barnes M, Aihara K, McLean CY, Fouse SD, Yamamoto S, Ueda H, Tatsuno K, Asthana S, Jalbert LE, Nelson SJ, Bollen AW, Gustafson WC, Charron E, et al. Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. Science (New York, NY). 2014; 343:189–193.

20. Killela PJ, Reitman ZJ, Jiao Y, Bettegowda C, Agrawal N, Diaz LA, Jr., Friedman AH, Friedman H, Gallia GL, Giovanella BC, Grollman AP, He TC, He Y, Hruban RH, Jallo GI, Mandahl N, et al. TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. Proceedings of the National Academy of Sciences of the United States of America. 2013; 110:6021–6026.

21. Killela PJ, Pirozzi CJ, Healy P, Reitman ZJ, Lipp E, Rasheed BA, Yang R, Diplas BH, Wang Z, Greer PK, Zhu H, Wang CY, Carpenter AB, Friedman H, Friedman AH, Keir ST, et al. Mutations in IDH1, IDH2, and in the TERT promoter define clinically distinct subgroups of adult malignant gliomas. Oncotarget. 2014; 5:1515–1525.

22. Bettegowda C, Agrawal N, Jiao Y, Sausen M, Wood LD, Hruban RH, Rodriguez FJ, Cahill DP, McLendon R, Riggins G, Velculescu VE, Oba-Shinjo SM, Marie SK, Vogelstein B, Bigner D, Yan H, et al. Mutations in CIC and FUBP1 contribute to human oligodendroglioma. Science. 2011; 333:1453–1455.

23. Sharma VP, Fenwick AL, Brockop MS, McGowan SJ, Goos JA, Hoogeboom AJ, Brady AF, Jeelani NO, Lynch SA, Mulliken JB, Murray DJ, Phipps JM, Sweeney E, Tomkins SE, Wilson LC, Bennett S, et al. Mutations in TCF12, encoding a basic helix-loop-helix partner of TWIST1, are a frequent cause of coronal craniosynostosis. Nature genetics. 2013; 45:304–307.

24. Nicolay DJ, Doucette JR, Nazarali AJ. Transcriptional control of oligodendrogenesis. Glia. 2007; 55:1287–1299.

25. Sun Y, Zhang W, Chen D, Lv Y, Zheng J, Lilljebjorn H, Ran L, Bao Z, Soneson C, Sjogren HO, Salford LG, Ji J, French PJ, Fioretos T, Jiang T, Fan X. A glioma classification scheme based on coexpression modules of EGFR and PDGFRA. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:3538–3543.

26. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102:15545–15550.

27. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nature genetics. 2003; 34:267–273.

28. Gobert RP, Joubert L, Curchod ML, Salvat C, Foucault I, Jorand-Lebrun C, Lamarine M, Peixoto H, Vignaud C,

Fremaux C, Jomotte T, Francon B, Alliod C, Bernasconi L, Abderrahim H, Perrin D, et al. Convergent functional genomics of oligodendrocyte differentiation identifies multiple autoinhibitory signaling circuits. Molecular and cellular biology. 2009; 29:1538–1553.

29. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, Weinberg RA. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nature genetics. 2008; 40:499–507.

30. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, Zhang X, Bernstein BE, Nusbaum C, Jaffe DB, Gnirke A, Jaenisch R, Lander ES. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008; 454:766–770.

31. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer cell. 2010; 17:98–110.

32. Matera AG, Wang Z. A day in the life of the spliceosome. Nature reviews Molecular cell biology. 2014; 15:108–121.

33. Singh RK, Cooper TA. Pre-mRNA splicing in disease and therapeutics. Trends in molecular medicine. 2012; 18:472–482.

34. Larsson CA, Cote G, Quintas-Cardama A. The changing mutational landscape of acute myeloid leukemia and myelodysplastic syndrome. Molecular cancer research: MCR. 2013; 11:815–827.

35. Herold T, Metzeler KH, Vosberg S, Hartmann L, Rollig C, Stolzel F, Schneider S, Hubmann M, Zellmeier E, Ksienzyk B, Jurinovic V, Pasalic Z, Kakadia PM, Dufour A, Graf A, Krebs S, et al. Isolated trisomy 13 defines a homogeneous AML subgroup with high frequency of mutations in spliceosome genes and poor prognosis. Blood. 2014; 124:1304–1311.

36. Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. The New England journal of medicine. 2011; 365:2497–2506.

37. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011; 478:64–69.

38. Adler AS, McCleland ML, Yee S, Yaylaoglu M, Hussain S, Cosino E, Quinones G, Modrusan Z, Seshagiri S, Torres E, Chopra VS, Haley B, Zhang Z, Blackwood EM, Singh M, Junttila M, et al. An integrative analysis of colon cancer identifies an essential function for PRPF6 in tumor growth. Genes & development. 2014; 28:1068–1084.

39. Hu JS, Olson EN, Kingston RE. HEB, a helix-loop-helix protein related to E2A and ITF2 that can modulate the DNA-binding ability of myogenic regulatory factors. Molecular and cellular biology. 1992; 12:1031–1042.

40. Lee CC, Chen WS, Chen CC, Chen LL, Lin YS, Fan CS, Huang TS. TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. The Journal of biological chemistry. 2012; 287:2798–2809.

41. Sussman CR, Davies JE, Miller RH. Extracellular and intracellular regulation of oligodendrocyte development: roles of Sonic hedgehog and expression of E proteins. Glia. 2002; 40:55–64.

42. Riemenschneider MJ, Koy TH, Reifenberger G. Expression of oligodendrocyte lineage genes in oligodendroglial and astrocytic gliomas. Acta neuropathologica. 2004; 107:277–282.

43. Thorsen K, Schepeler T, Oster B, Rasmussen MH, Vang S, Wang K, Hansen KQ, Lamy P, Pedersen JS, Eller A, Mansilla F, Laurila K, Wiuf C, Laurberg S, Dyrskjot L, Orntoft TF, et al. Tumor-specific usage of alternative transcription start sites in colorectal cancer identified by genome-wide exon array analysis. BMC genomics. 2011; 12:505.

44. Wu JN, Roberts CW. ARID1A mutations in cancer: another epigenetic tumor suppressor? Cancer discovery. 2013; 3:35–43.

45. Wiegand KC, Shah SP, Al-Agha OM, Zhao Y, Tse K, Zeng T, Senz J, McConechy MK, Anglesio MS, Kalloger SE, Yang W, Heravi-Moussavi A, Giuliany R, Chow C, Fee J, Zayed A, et al. ARID1A mutations in endometriosis-associated ovarian carcinomas. The New England journal of medicine. 2010; 363:1532–1543.

46. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, Weigman VJ, Marynen P, Zetterberg A, Naume B, Perou CM, Borresen-Dale AL, Kristensen VN. Allele-specific copy number analysis of tumors. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:16910–16915.