TOP2A substitution enhances topoisomerase activity and causes transcriptional dysfunction in glioblastoma patients

Bartłomiej Gielniewski[1]*, Katarzyna Poleszak[1]*, Adria-Jaume Roura[1], Paulina Szadkowska[1], Sylwia K. Król[1], Rafal Guzik[1], Paulina Wiechecka[1], Marta Maleszewska[1], Beata Kaza[1], Andrzej Marchel[2], Tomasz Czernicki[2], Andrzej Koziarski[3], Grzegorz Zielinski[3], Andrzej Styk[3], Maciej Kawecki[4,7], Cezary Szczylik[4], Ryszard Czepko[5], Mariusz Banach[5], Wojciech Kaspera[6], Wojciech Szopa[6], Mateusz Bujko[7], Bartosz Czapski[8], Miroslaw Zabek[8,13], Ewa Iżycka-Świeszewska[9], Wojciech Kloc[10,11], Pawel Nauman[12] , Bartosz Wojtas[#1], Bozena Kaminska[#1]

[1]Laboratory of Molecular Neurobiology, Nencki Institute of Experimental Biology of the Polish Academy of Sciences, Warsaw, Poland;

[2]Department of Neurosurgery, Medical University of Warsaw, Warsaw, Poland;

[3]Department of Neurosurgery, Military Institute of Medicine, Warsaw, Poland;

[4]Department of Oncology, Military Institute of Medicine, Warsaw, Poland;

[5]Andrzej Frycz Modrzewski Kraków University, Clinical Department of Neurosurgery St. Raphael Hospital, Krakow, Poland;

[6]Department of Neurosurgery, Medical University of Silesia, Regional Hospital, Sosnowiec, Poland;

[7]The Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland;

[8]Department of Neurosurgery, Mazovian Brodnowski Hospital, Warsaw, Poland;

[9]Medical University of Gdansk, Gdansk, Poland;

[10]Department of Neurosurgery, Copernicus PL, Gdansk, Poland;

[11]Department of Psychology and Sociology of Health and Public Health School of Public Health Collegium Medicum, University of Warmia – Mazury, Olsztyn, Poland;

[12]Institute of Psychiatry and Neurology, Warsaw, Poland;

[13]Department of Neurosurgery and Nervous System Trauma, Centre of Postgraduate Medical Education, Warsaw, Poland

* Authors contributed equally to this work.

#Corresponding authors: Prof. Bozena Kaminska, Dr. Bartosz Wojtas

Laboratory of Molecular Neurobiology, Neurobiology Center

Nencki Institute of Experimental Biology,

3 Pasteur Str., 02-093 Warsaw, Poland

Fax: (48 22) 8225342

Phone: (48 22) 5892223

E-mail: b.kaminska@nencki.edu.pl; b.wojtas@nencki.edu.pl

Running title: Novel mutations in *TOP2A* in gliomas

Abstract

High grade gliomas (HGGs) are aggressive, primary brain tumors with poor clinical outcomes. To better understand glioma pathobiology and find potential therapeutic susceptibilities, we designed a custom- panel 664 cancer- and epigenetics-related genes and employed targeted next generation sequencing to study the genomic landscape of somatic and germline variants in 182 glioma samples of different malignancy grades. Besides known alterations in *TP53*, *IDH1*, *ATRX*, *EGFR* genes, we found several novel variants that can be potential drivers in gliomas. In four patients from the Polish glioma cohort, we identified a novel recurrent mutation in the *TOP2A* gene coding for Topoisomerase 2A in glioblastomas (GBM, WHO grade IV gliomas). The mutation results in a substitution of glutamic acid (E) 948 to glutamine (Q) of TOP2 A and we predicted this E948Q substitution may affect DNA binding and a TOP2A enzymatic activity. Topoisomerases are enzymes that control the higher order DNA structure by introducing transient breaks and rejoining DNA strands. Using recombinant proteins we demonstrated stronger DNA binding and DNA supercoil relaxation activities of the variant proteins. Glioblastoma (GBM) patients with the mutated *TOP2A* had shorter overall survival than wild type *TOP2A* GBM patients. Computational analyses of transcriptomic data showed that the GBM samples with the mutated *TOP2A* have different transcriptomic patterns suggesting higher transcriptomic activity. The results suggest that TOP2A E948Q variant strongly binds to DNA and is more active in comparison to the wild-type protein. Altogether, our findings suggest that the E948Q substitution leads to gain of function by TOP2A.

Introduction

High-grade gliomas (HGGs) are most frequent, primary brain tumors in adults.[34, 35, 54, 55] The World Health Organization (WHO) classifies HGGs as WHO grade III and IV gliomas. Among HGGs, the most aggressive is glioblastoma (GBM, a WHO grade IV tumor) which has a high recurrence and mortality rate due to highly infiltrative growth, genetic alterations and therapy resistance. Post-surgery radiotherapy combined with temozolomide in glioblastomas or with procarbazine, CCNU (lomustine) and vincristine in diffuse low-grade (WHO grade II) and anaplastic (WHO grade III) gliomas showed some survival benefits. Overall survival of HGG patients from a time of diagnosis ranges from 12-18 months despite an extensive all available therapies.[20, 43, 53] Recent advancements in high-throughput genomic and transcriptomic studies performed by The Cancer Genome Atlas (TCGA) consortium[6] brought emerging insights into etiology, molecular subtypes and putative therapeutic cues in HGGs. Recurrent somatic alterations in genes such as *TP53*, *PTEN*, *NF1*, *ATRX*, *EGFR*, *PDGFRA* and *IDH1/2* have been reported.[5, 17, 33, 49] Discovery of the recurrent mutation in the *IDH1* gene[36] in gliomas and resulting distinct transcriptomic and DNA methylation profiles in *IDH*-mutant and *IDH*-wild type gliomas[7] improved the WHO classification.[28, 37, 57] Phosphoinositide 3-kinase (PI3K) signaling which is affected by several oncogenic alterations such as *PTEN* deletion, mutations in *PDGFR* (encoding platelet derived growth factor receptor) or *EGFR* (encoding epidermal growth factor receptor) emerged as a vital target in HGGs. Several PI3K inhibitors including pan-PI3K, isoform-selective or dual PI3K/mammalian target of rapamycin (mTOR) inhibitors have displayed promising preclinical results and entered clinical trials.[21, 63] Regrettably, inhibitors against aberrant EGFR failed in the phase III clinical trials[52, 58] and none of the receptor tyrosine kinase inhibitors have been approved for GBM patients. Due to the lack of an effective treatment in HGGs, even rare, but targetable genetic changes are of interest as they may offer a personalized treatment for selected patients. *BRAF V600E* mutations found in approximately

1% of glioma patients are considered as a therapeutic target in basket trials including gliomas.[19, 60] Inhibitors of Aurora kinases, which play important roles in the cell division through regulation of chromosome segregation, had a synergistic or sensitizing effect with chemotherapy drugs, radiotherapy, or with other targeted molecules in GBM and several of them are currently in clinical trials.[3, 13]

In this study, we applied targeted next generation sequencing of a custom panel of 664-cancer- and epigenetics related genes to study the genomic landscape of both somatic and germ-line variants in 182 high grade gliomas. Besides previously described alterations in *TP53*, *IDH1*, *ATRX*, *EGFR* genes, we found several new variants that were predicted to be potential drivers. In four GBM patients from the Polish HGG cohort, we identified a novel, recurrent mutation in the *TOP2A* gene, resulting in a substitution of glutamic acid (E) 948 to glutamine (Q). The *TOP2A* gene encodes Topoisomerase 2A which belongs to a family of Topoisomerases engaged in the maintenance of a higher order DNA structure. These enzymes are also implicated in other cellular processes such as chromatin folding and gene expression, in which the topological transformations of higher-order DNA structures are catalyzed.[39] GBMs with *TOP2A* mutations had shorter survival and displayed the increased transcriptomic activity. Using computational methods we predicted the consequences of the E948Q substitution on TOP2A and verified them with biochemical assays using a recombinant wild type and variant TOP2A. Our results show that the E948Q substitution in TOP2A affects its DNA binding and enzymatic activity, which could lead to a “gain of function”.

Materials and Methods

Tumor Samples

Glioma samples and matching blood samples were obtained from the: St. Raphael Hospital, Scanmed, Cracow, Poland; Medical University of Silesia, Sosnowiec, Poland; Medical University of Warsaw, Poland; Military Institute of Medicine, Warsaw, Poland; Copernicus Hospital PL, Gdansk, Poland; Medical University of Gdańsk, Poland; Institute of Psychiatry and Neurology, Warsaw, Poland; Mazovian Brodnowski Hospital, Warsaw, Poland; Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland; Canadian Human Tissue Bank. Each patient provided a written consent for the use of tumor tissues. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committees, with the 1964 Helsinki declaration and its later amendments. Detailed description of our patient cohort is presented in Supplementary Table 1. In the analysis we used 182 II, III and IV grade glioma samples. For 67 of them we were able to sequence DNA isolated from blood as a reference for analysis of somatic mutations. Most of the patients from our cohort suffered from glioblastoma (132/182). For the samples provided by Canadian Human Tissue Bank (GL162-GL182) we do not have access to clinical data, excluding diagnosis.

DNA/RNA Isolation

Total DNA and RNA were extracted from fresh frozen tissue samples using Trizol Reagent (Thermo Fischer Scientific, Waltham, MA, USA), following manufacturer's protocol. DNA from blood samples was isolated using QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany), following manufacturer's protocol. Quality and quantity of nucleic acids were determined by

Nanodrop (Thermo Fisher Scientific, Waltham, MA, USA) and Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Design of Targeted Cancer-Related Gene Enrichment Panel

More than 90% of known pathogenic mutations registered in the NCBI ClinVar database are located in protein-coding DNA sequence (CDS) regions. We used SeqCap EZ Custom Enrichment Kit, which is an exome enrichment design that targets the latest genomic annotation GRCh38/hg38. Our SeqCap EZ Choice Library covers exomic regions of 664 genes frequently mutated in cancer. A majority of genes (578) were selected from the Roche Nimblegene Cancer Comprehensive Panel (based on Cancer Gene Consensus from Sanger Institute and NCBI Gene Tests). Additionally, we included 86 epigenetics-related genes (genes coding for histone acetylases/deacetylases, histone methylases/demethylases, DNA methylases/demethylases and chromatin remodeling proteins) based on a literature review.[12, 30, 47]

DNA and RNA Sequencing

DNA isolated from tumor samples was processed for library preparation according to a NimbleGen SeqCap EZ Library SR (v 4.2) user guide. Libraries were prepared from 1 µg of total DNA. Initial step in the procedure was fragmentation of the material to obtain 180-220 bp DNA fragments. Next ends of these fragments were repaired and A base was added to 3’ end. Then indexed adapters were ligated and libraries were amplified. Obtained libraries were mixed eximolar and special oligonucleotide probes were added to enrich the libraries with earlier fragments of interest. After 72 h incubation of the mixture pooled libraries were purified using special beads and amplified. Quality control of obtained libraries was done using Agilent Bioanalyzer with High Sensitivity DNA Kit (Agilent Technologies, Palo Alto, CA, USA).

Quantification of the libraries was done using Quantus Fluorometer and QuantiFluor Double Stranded DNA System (Promega, Madison, Wisconsin, USA). Libraries were run in the rapid run flow cell and were paired-end sequenced (2x76bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

mRNA sequencing libraries were prepared using KAPA Stranded mRNAseq Kit according to manufacturer's protocol. Briefly, mRNA molecules were enriched from 500 ng of total RNA using poly-T oligo-attached magnetic beads (Kapa Biosystems, MA, USA). Obtained mRNA was fragmented and first and second strand of cDNA were synthesized. Then adapter was ligated and loop structure of adapter was cut by USER enzyme (NEB, Ipswich, MA, USA). Amplification of obtained dsDNA fragments containing the adapters was performed using NEB starters (Ipswich, MA, USA). Quality control of obtained libraries was done using Agilent Bioanalyzer with High Sensitivity DNA Kit (Agilent Technologies, Palo Alto, CA, USA). Quantification of the libraries was done using Quantus Fluorometer and QuantiFluor Double Stranded DNA System (Promega, Madison, Wisconsin, USA). Libraries were run in the rapid run flow cell and were paired-end sequenced (2x76bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

Total RNA sequencing libraries were prepared using NEBNext Ultra II Directional RNA Library Prep Kit (NEB, Ipswich, MA, USA). rRNA molecules were depleted using NEBNext rRNA Depletion Kit (NEB, Ipswich, MA, USA). Libraries were prepared from 100 ng of total RNA. After rRNA depletion first and second strand of cDNA were synthesized, adapter was ligated and finally libraries were amplified using NEB starters (NEB, Ipswich, MA, USA). Quallity evaluation of the libraries was done using Agilent Bioanalyzer with High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA) and concentrations were checked with Quantus Fluorometer and QuantiFluor Double Stranded DNA System (Promega, Madison, Wisconsin, USA). Libraries were run in the rapid run flow cell and were paired-end sequenced (2x76bp) on HiSeq 1500 (Illumina, San Diego, CA 92122 USA).

Data Analysis

Somatic plus rare germline variants pipeline

Fastq files from both tumor and blood samples were prepared as follows. Sequencing reads were filtered by trimmomatic program.[2] Filtered and trimmed reads were mapped to human genome version hg38 by BWA aligner[24] (http://bio-bwa.sourceforge.net/). The mapped reads as the BAM files were sorted and indexed. Next, sample read pileup was prepared by samtools mpileup and then variants were called by bcftools and filtered by varfilter from vcfutils to include just variants with coverage higher than 12 reads per position and with mapping quality higher than Q35. Next variants were filtered to include just the ones with variant allele frequency (VAF) higher than 20%, meaning that more than 20% of reads had to support the variant. Separate analysis on the same sample was prepared with Scalpel software[16] to enrich short deletions and insertions discovery. Variants obtained with scalpel were filtered as well to include just variants with VAF>20%. Variants obtained from bcftools and scalpel analysis were merged and processed to maf format by vcf2maf program. Vcf2maf tool annotates genetic variants to existing databases by Variant Effect Predictor (VEP) program.[32] Subsequently maf object was filtered by the Minor Allele Frequency (MAF) value, it had to be lower than 0.001 in all populations annotated by VEP to be included in the final analysis, meaning that in any of well defined populations, including gnomAD, 100 genomes, EXAC it was not found in more than 1 in 1000 individuals. Filtered maf object was subsequantly analyzed in maftools R library.[31] At the end our glioma cohort included putative somatic and rare germline variants.

Somatic variants pipeline

FASTQ files from both tumor and blood samples were processed with trimmomatic program[2] to remove low quality reads and sequencing adapters. Filtered and trimmed reads were mapped to human genome (hg38) by NextGenMap aligner (http://cibiv.github.io/NextGenMap/). Read

duplicates were marked and removed by Picard (https://broadinstitute.github.io/picard/) and only properly oriented and uniquely mapping reads were considered for further analysis. For somatic calls, a minimum coverage of 10 reads was established for both normal and tumor samples. Additionally, variants with strand supporting reads bias were discarded. Only coding variants with damaging predicted SIFT values (>0.05) were selected.

ProcessSomatic method from VarScan2[23] was applied to extract high confidence somatic calls based on variant allele frequency and Fisher's exact test p-value. The final subset of variants was annotated with Annovar (http://annovar.openbioinformatics.org/en/latest/) employing the latest databases versions (refGene, clinvar, cosmic, avsnp150 and dbnsfp30a). Finally, maftools R library[31] was used to analyze the resulting somatic variants.

To test overlap between a somatic analysis pipeline (including blood DNA as a reference) and somatic andrare germ-line variants pipelines, we compared 50 samples that were common in two analyses. A number of Missense Mutations, Frame Shift Del, Frame Shift Ins and Nonsense mutation obtained was compared between these two pipelines.

RNA sequencing analysis

RNA sequencing reads were aligned to the human genome by gap-aware algorithm - STAR (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3530905/) and gene counts were obtained by featurecounts.[26] Aligned reads (bam format) were processed by RseQC[51] program to evaluate the quality of obtained results, but also to estimate the proportion of reads mapping to protein coding exons, 5'-UTRs, 3'-UTRs and introns. From the gene counts subsequent statistical analysis was performed in R using NOIseq[46] and clusterProfiler[61] libraries.

Prediction of the TOP2A mutant effect

To predict the consequences of E948Q substitution on the TOP2A, we used the crystal structure of the human TOP2A in a complex with DNA (PDB code: 4FM9). Protein structures were visualized and an electrostatic potential was calculated using PyMOL.

Cloning and mutagenesis

The *TOP2A* gene (NCBI nucleotide accession number (AC) – NM_001067, protein – NP_001058.2) cloned into pcDNA3.1+/C-(K)-DYK vector (Clone ID OHu21029D) was purchased from Genscript. The cDNAs were cloned into a bacterial expression vector. *TOP2A* fragment corresponding to amino acid residues 890-996 was amplified and cloned into pET28a vector as a BamHI - EcoRI fragment, resulting in a construct expressing His6 tag fused to N-terminus of TOP2A 890-996 aa. *TOP2A* fragment corresponding to amino acid residues 431-1193 was amplified and cloned into a pET28a vector as a SalI - NotI fragment, resulting in a construct expressing His6 tag fused to N-terminus of TOP2A 431-1193 aa. Site-directed mutagenesis of *TOP2A* gene was performed by a PCR-based technique. All mutant genes were sequenced and found to contain only the desired mutations.

Protein expression and purification

Proteins were expressed from bacterial plasmid DNA carrying *TOP2A* in *E. coli* Rosetta (DE3) strain. Cells were grown in LB medium with appropriate antibiotics at 37°C. When the bacteria reached OD_{600} 0.8, protein expression was induced by addition of 1 mM IPTG. The induction of TOP2A 890-996 aa was carried out at 37°C for 4 h. One hour before the induction of TOP2A 431-1193 aa the temperature was changed to 37°C and the induction was carried out overnight. Cells were harvested by centrifugation (4000 g for 5 min, 4°C) and pelleted. Proteins were purified using the HisPur Ni-NTA Magnetic beads (Thermo Scientific). The pellet was resuspended in binding buffer (10 mM Na_2HPO_4, 1,8 mM KH_2PO_4, 2,7 mM KCl, 300 mM NaCl, 10 mM imidazole, pH 8.0, 10 % (v/v) glycerol, 10 mM 2-Mercaptoethanol (BME), 1 mM PMSF)

and lysed using Bioruptor Plus sonicator (Diagenode). The purification was carried out at 4°C. Proteins from the clarified lysates were bound to the Ni-NTA resin, the beads were washed with binding buffer, wash buffer 1 (binding buffer with 2 M NaCl) and wash buffer 2 (binding buffer with 20 mM imidazole). The proteins were eluted with elution buffer (binding buffer with 250 mM imidazole, pH 8.0). Next, the buffer was exchanged to (phosphate-buffered saline, 10 % (v/v) glycerol, 10 mM BME) using G25 Sephadex resin (Thermo Scientific) and Spin-X Centrifuge Tube (Costar). Final protein samples of TOP2A 890-996 aa were analyzed in 15% SDS-PAGE and TOP2A 431-1193 aa samples in 6% SDS-PAGE. The homogeneity of the proteins was approximately 90%. Protein concentrations were estimated by measuring the absorbance at 280 nm.

Electrophoretic mobility shift assay (EMSA)

EMSA was performed using the LightShift Chemiluminescent EMSA Kit (Thermo Scientific Cat # 20148) according to manufacturer's instructions. Binding reactions contained 20 pM biotin-labeled oligonucleotide 60 bp end-labeled duplex from LightShift Chemiluminescent EMSA Kit and 4.25 µM TOP2A 431-1193 aa WT or E948Q variant proteins, in the 10 mM Tris pH 7.5 buffer with 50 mM KCl, 1 mM DTT, 5 mM $MgCl_2$. DNA binding reactions were performed in 30 µl. The reaction mixtures were incubated for 30 min at room temperature and subjected to electrophoresis (100 V, 8°C) on 6% polyacrylamide gels with 10% glycerol and Tris–borate–EDTA buffer. Then, complexes were transferred onto a 0.45 µm Biodyne nylon membrane (Thermo Scientific Cat # 77016) in a Tris–borate–EDTA buffer and detected by chemiluminescence using a Chemidoc camera (Bio-Rad).

Supercoil relaxation assay

The supercoil relaxation reaction was conducted at 37°C for 30 min in a buffer containing 20 mM Tris pH 8.0, 100 mM KCl, 0.1 mM EDTA, 10 mM $MgCl_2$, 1 mM adenosinetriphosphate

(ATP), 30 µg/ml bovine serum albumin (BSA) and 200 ng supercoiled (SC) pET28a plasmid with *TOP2A* fragment corresponding to amino acid residues 890-996 DNA. 0.1 and 0.2 µM TOP2A 431-1193 aa WT or E948Q variant were added to start the reaction. The resultant reaction mixtures were resolved by electrophoresis on 0.7% agarose gel followed by staining with SimplySafe dye (EURX Cat # E4600-01) and visualization using Chemidoc camera (Bio-Rad).

Results

Landscape of somatic and rare germ-line variants in the studied cohort

A total of 182 glioma specimens were collected in this study, 21 were from Canadian Brain Tumor Bank, 161 were from the Polish glioma cohort. Majority of tumors were glioblastomas (74%). The clinical characteristics of patients are summarized in the Supplementary Table 1 and Supplementary Fig.1. We searched for somatic and rare germline variants, as genetic variants databases have significantly improved in the last few years allowing to identify variants occurring with very low frequency in existing databases (based on EXAC, TOPMED, gnomAD and 1000 genomes projects) even in the absence of a reference blood DNA. As the identified germline variants were exceptionally rare in the general population (threshold of MAF<0.001 in all external databases), it is likely that these variants are pathogenic.

Our approach to find potential new oncogenic drivers was based on OncodriveCLUST spatial clustering method.[45] OncodriveCLUST estimates a background model from coding-silent mutations and tests protein domains containing clusters of missense mutations that are likely to alter protein structure. The results show *IDH1* gene as a top scoring hotspot, but among identified genes we found also genes such as *FOXO3*, *PDE4DIP*, *HOXA11, ABL1* and *RECQL4*

that were not previously reported as mutated in HGGs (Fig.1A, B). When the identified genes were compared to the results obtained in the TCGA project, most of the genes (shown in bold in the Fig.1B) were found to be potentially new oncogenic genes.

One of the most interesting genes was *TOP2A,* which expression is highly up-regulated in gliomas in a grade-specific manner (Fig. 1C). The *TOP2A* gene harbored 7 different mutations and interestingly, one mutation occurred in 4 GBM patients exactly in the same chromosomal position. This recurrent mutation in the *TOP2A* was located inside a gene fragment encoding TOWER domain of TOP2A (Fig. 1D). The presented lolliplot (Fig. 1D) shows all variants found in the *TOP2A* gene, including the ones that did not pass QC analysis (Supplementary Table 2).

The results of the targeted sequencing indicate a high frequency of altered genes (Supplementary Fig. 2A) and are in agreement with previous findings on the landscape of pathogenic genetic changes in gliomas.[41] Annotations of variants to genes revealed that top 10 of the most frequently altered genes have genetic changes in more than 80% of samples (Supplementary Fig. 2A). The most altered gene was *TP53*, followed by *IDH1*, *PTEN*, *ATRX* and *EGFR* (Supplementary Fig. 2A). Other genes that were found to be frequently altered included *KDM6B*, *ABL1*, *ARID1A*, *AR* and *NCOA6* (Supplementary Fig. 2A). One of the reasons, why these changes have not been noticed previously, is the fact that these changes are most likely rare germline alterations and represent mostly in-frame insertions (*KDM6B*) or in-frame deletions (*ABL1*, *ARID1A*, *AR, NCOA6)*. These genetic alterations were found at a very low frequency in the general population (MAF<0.001) and would result in insertion or deletion one or more amino acids to the protein, therefore we consider them potentially pathogenic. The highest number of genetic alterations occur in genes coding for receptor tyrosine kinases (RTK) of RAS pathway (RTK-RAS), less frequently in genes coding for proteins from NOTCH, WNT and PI3K pathways (Supplementary Fig. 2B). More than 30 genes from the RTK-RAS pathway were found to be altered (Supplementary Fig. 2C). The identification of RTK-RAS, MAPK

(mitogen activated protein kinases), PI(3)K (phosphatidylinositol-3-OH kinase) and Wnt/β-catenin signaling pathways as deregulated in HGGs is consistent with the results of previous cancer studies. Notably, new categories (for example, transcription regulators, epigenetic enzymes and modifiers) emerge as exciting, potential targets for the development of new therapeutics. In this cohort, the *ABL1* gene was found to be altered more frequently than the *EGFR* gene (Supplementary Fig. 2C). Most frequently mutated genes encoding specific protein domains from the PFAM database are: the TP53 domain, the PTZ00435 domain related to *IDH1/2* mutations, conserved domains of protein kinases (PKc_like, S_TKc) and kinase phosphatases (PTPc) (Supplementary Fig. 2D).

Identification of somatic variants in high grade gliomas

Fifty glioma samples of WHO grades II, III and IV and reference blood DNAs were subjected to targeted sequencing and detection of SNVs were performed on both a tumor and blood DNA, which allowed to confirm if the identified variants were somatic. Considering a set of the somatic mutations with deleterious consequences (Fig. 2A), the most frequently altered gene is *TP53*, which is mutated in 48% of the patients, followed by *IDH1*, *ATRX*, *EGFR* and *PTEN*, among others.

We investigated the mutational signatures of hotspot mutations and illustrated the common and distinct sequence contexts under which the hotspot mutations were enriched. Two novel mutational hotspots with deleterious consequences were detected in the *AKAP9* gene (Fig. 2A,B). A-kinase anchor protein 9 (AKAP9), a cAMP-responsive scaffold protein AKAP9, regulates microtubule dynamics and nucleation at the Golgi apparatus[48], promotes development and growth of colorectal cancer[18] and is associated with enhanced cancer risk and metastatic relapses in breast cancer[22] The analysis of *AKAP9 e*xpression in gliomas from the TCGA dataset shows its lower expression in HGGs, in particular in glioblastomas (Fig. 2C). Low expression of *AKAP9* associates with poor survival of HGG patients (Fig. 2D). This finding

suggests that *AKAP9* could be an important mutated gene in a subset of HGGs. Using the Sorting Intolerant from Tolerant (SIFT) algorithm which predicts whether the detected frameshift insertion could affect the protein structure, we established that the somatic variation in the *AKAP9* gene could have a damaging effect on protein structure.

To verify concordance of the somatic and somatic/rare germline pipeline of the variant analysis, we have computed overlap of these results and found out that most of somatic variants are also found by a somatic/rare germline pipeline (Supplementary Table 3).

Novel mutations identified in the *TOP2A* gene

We found 7 SNPs in the *TOP2A* gene in 10 glioblastoma samples (Supplementary Table 2). All identified SNPs give rise to missense variants. Transversion of guanine to cytosine at the chromosomal position 17:40400367 which results in the substitution of E948 to Q, was found in four GBM patients, and several other mutations in the *TOP2A* gene were found in GBM patients from the Polish HGG cohort. These SNVs have been verified by Sanger sequencing, apart from one sample which had a low variant allele frequency. The *TOP2A* gene encodes Topoisomerase 2A which belongs to a family of topoisomerases that relax, loosen, and disentangle DNA by transiently cleaving and re-ligating both strands of DNA, which modulate DNA topology.[10, 39] The novel, recurrent mutation in the *TOP2A* gene results in a substitution of E948 to Q in the Tower domain of the TOP2A.

We studied exclusivity and co-occurrence of *TOP2A* mutations among detected top mutations. *TOP2A* mutations co-occurred with *ARID1A* mutations, encoding a SWI/SNF family helicase, and *KMT2C* (formerly *MLL3*), coding for histone methyltransferase, as well as with *EGFR* and *NOTCH3* mutations (Fig. 3A). One of the *TOP2A* mutated samples (first from the left in Fig. 3A) is a hypermutated sample, some of the other samples have also a high number of genetic alterations when compared to other well-defined clusters, such as *IDH1_TP53-, IDH1-, TP53- and PTEN*-mutated clusters (Fig. 3B). All *TOP2A* mutated samples show a higher

number of genetic alterations when compared to other defined clusters of glioma samples (Fig. 3B). Moreover, most of the changes in *TOP2A* mutated samples are C>T transitions, linked most likely to deamination of 5-methylcysteine[42] (Fig. 3C). There is a high variability within GBM samples with the *TOP2A* mutated, showing roughly a half of *TOP2A* samples having a very high fraction (>80%) of C>T changes, while the other half have an expected fraction (<40%) of C>T changes (Fig. 3C).

Using RNA sequencing (RNA-seq) data from HGGs in the TCGA dataset, we demonstrate that patients with the mutated *TOP2A* had shorter survival than other HGG patients within that cohort (p=0.00027) (Fig. 3D). The mutation in the *TOP2A* gene at the position 17:40400367 was identified in four GBM samples in the studied cohort. Three GBM patients with the mutated *TOP2A* (one patient died after surgery) survived 5, 3 and 8 months from the time of diagnosis, respectively, which indicates that the GBM patients harboring this *TOP2A* variant had a shorter overall survival compared to other GBM patients (Fig. 3E). This suggests a negative effect of mutations in the *TOP2A* gene on the overall survival of GBM patients.

E948Q substitution in the TOP2A changes its functions

To predict the consequences of E948Q substitution on the TOP2A, we used the crystal structure of the human TOP2A in a complex with DNA (PDB code: 4FM9[56]) (Fig.4A, B). Amino acid residue E948 is located in the Tower domain of the TOP2A protein (Fig. 4C), which with other domains (TOPRIM and WHD) form the DNA-Gate involved in DNA binding.[1] In the crystal structure of the human TOP2A in a complex with DNA E948 is in a close proximity to DNA (Fig. 4B). The replacement of a negatively charged E residue to a polar Q residue changes an electrostatic potential of the protein (Fig. 4C). Therefore, we predict that this change would affect protein-DNA interactions and the TOP2A E948Q variant may have an increased affinity

towards DNA. Some of the other discovered *TOP2A* mutations may also have the effect on TOP2A binding to DNA, as described in Supplementary Table 4.

To analyze the effect of the TOP2A E948Q substitution on an enzymatic activity, we produced and purified recombinant wild-type (WT) and variant TOP2A proteins as His6 tag-fusion proteins (Fig. 5A). We compared DNA binding activity of WT and E948Q TOP2A using electrophoretic mobility shift assay (EMSA) and purified proteins: TOP2A 890-996 aa comprising only the Tower domain and TOP2A 431-1193 aa, which includes the entire DNA-Gate domain (see Materials and Methods). The representative EMSA gel shows that the TOP2A variant E948Q binds to DNA more strongly (Fig. 5B). The lower panel shows the results of densitometric assessment of shifted bands from two experiments.

Subsequently, we analyzed the effects of the TOP2A E948Q substitution on a catalytic activity of the protein. Topoisomerases are enzymes that regulate DNA topology by making temporary single- or double-strand DNA breaks. The DNA double helix needs to be unwound during DNA replication or transcription. Topoisomerases relax supercoiled DNA and decatenate DNA to allow DNA transcription and replication.[38, 50] *In vitro* measurements of a protein activity towards supercoiled DNA substrates may reflect the *in vivo* situation, at least with respect to the impact on DNA topology. We studied a supercoil relaxation activity of WT and TOP2A E948Q proteins. Disappearance of the lower band representing a supercoiled plasmid (SC bands) and appearance of higher band representing an untwisted plasmid DNA (OC) shows differences in the enzymatic activity of TOP2A proteins. The E948Q variant is functional and might have a higher activity than the WT protein (Fig. 6). The lower panel shows the results of densitometric assessment of shifted bands from these experiments.

Transcriptomic alterations in glioblastomas with the mutated *TOP2A* point to dysfunction of splicing

Bulk RNA sequencing was performed on 105 samples of low (WHO grade II) and high-grade gliomas (WHO grades III and IV), and we searched for peculiarities of transcriptional patterns in the wild type and mutated *TOP2A* gliomas. Among the analyzed samples were 7 GBM samples with mutations in the *TOP2A* gene, including 4 samples with a recurrent mutation at position 17: 40400367. Transcriptional patterns of GBMs with mutations in the *TOP2A* gene were compared with samples with mutations in the *IDH1, ATRX, TP53 and PTEN* genes.

We found that a larger percentage of the detected sequencing reads are mapped in non-coding (intron) regions in GBMs with the mutated the *TOP2A* gene when compared to the other gliomas with the *IDH1, ATRX, TP53,* and *PTEN* mutations (Fig. 7A). This observation suggests a higher level of transcription in samples with the mutated *TOP2A*. The higher content of intron regions in the RNA sequencing data may indicate a higher level of *de novo* transcription or deficiency in mRNA splicing (Fig. 7A). An analysis of the content of G-C pairs in the coding sequences of the expressed genes in gliomas with the mutated *TOP2A* versus other samples was performed (Fig. 7B). The analysis showed significant differences in the content of G-C pairs in the coding sequences of the genes expressed in the compared groups. In GBMs with the *TOP2A* mutation, the genes with the coding sequences rich in A-T pairs were more often expressed, while the genes with coding sequences rich in G-C pairs were less frequently present in these samples (Fig. 7B).

Additionally, we tested the expression of some genes that are not expressed in a normal brain and *TOP2A*-wild type (WT) and *TOP2A* mutated GBMs. RNA-seq data were filtered to remove all brain-specific genes, by filtering out genes expressed in the brain uploaded from the Stanford Brain RNA-Seq database and our data on human normal brain samples. Expression of non-brain specific genes was higher in the *TOP2A* mutated samples than in *TOP2A* WT glioma samples (Fig. 7C). We performed a Gene Ontology (GO) analysis of differentially expressed, non-brain specific genes and the results showed over-representation of

pathways involved in endopeptidase activity and spermatid development, suggesting acquisition of some new functionalities related to aberrant differentiation and migratory processes (Fig. 7D).

Further, we performed a GO terms analysis of non-brain specific genes that are upregulated in TOP2A mutants. Analysis has showed increased expression of genes associated with endopeptidase activity and germ cells development in samples with *TOP2A* mutations (Fig.7D). These acquired functions of TOP2A mutants may contribute to a shorter survival time of *TOP2A* mutated GBM patients.

Discussion

In this study we present the results of next generation sequencing with a custom-designed panel 664 cancer- and epigenetics-related genes of 182 II, III and IV grade glioma samples. This focused NGS approach has brought a novel information regarding a spectrum of genomic alterations in high grade gliomas and revealed a set of known and novel mutations, The results considerably broaden our understanding of glioma pathobiology and disclose therapeutically targetable genes. Previous efforts of TCGA and other consortia to molecularly characterize a genomic landscape of gliomas revealed highly recurrent somatic alterations and provided a rationale for a molecular classification of gliomas.[4, 7, 15, 44]

We identified a number of well-known and described GBM frequent alterations in *TP53*, *IDH1*, *ATRX*, *EGFR*, *CDK8, RB1, PTEN* genes, which validates our approach. However, using OncodriveCLUST[45] we identified several novel variants that represent potentially deleterious germ-line, somatic mosaicism or loss-of-heterozygosity variants.. Mutations in genes such *FOXO3*, *PDE4DIP*, *HOXA11, ABL1, TOP2A* and *RECQL4* were not previously reported as mutated in HGGs (Fig.1A,B). In three HGGs we found a recurrent mutation in the *AKAP9* gene coding for A-kinase anchor protein 9/AKAP9), a scaffold protein which regulates microtubule dynamics and nucleation at the Golgi apparatus.[48] AKAP9 promotes development and growth of

colorectal cancer[18] and is associated with enhanced risk and metastatic relapses in breast cancer.[22] *AKAP9 e*xpression was lower in HGGs, in particular in GBMs and low expression of *AKAP9* associated with poor survival of HGG patients (Fig. 2D). The somatic variation in the *AKAP9* gene could have a damaging effect on four different transcripts and result in a non-functional or truncated protein.

When a set of mutated genes from this study was compared with the results obtained in the TCGA project, most of them were found to be potentially oncogenic genes. There are several potential reasons for detecting novel mutations in this study: updated databases with novel definitions of SNVs, targeted sequencing focusing on specific genes and a unique national cohort. Among top 50 genes most frequently mutated in this cohort, of particular interest are are genetic alterations in the genes coding for epigenetic enzymes and modifiers: *KDM6B, NCOA2*, *NCOA3* and *NCOA6. KDM6B* encodes histone 3 lysine 27 demethylase KDM6B/JMJD3 which controls the induction of a mature neuronal gene expression program in cerebellar granule neurons.[59] KDM6B is critical for maintenance of glioma stem cells that upregulate primitive developmental programs.[27] KDM6B has been targeted with a specific inhibitor GSK-J4 in acute myeloid leukemia[25] and GSK J4 is active against native and TMZ-resistant glioblastoma cells.[40] The presence of mutations in genes encoding three different members of the NCOA (nuclear receptor co-activator) family of co-activators conveys potential significance. NCOA2 is a co-activator of diverse nuclear receptors such as estrogen receptor, retinoic acid receptor/retinoid X receptor, thyroid receptor and vitamin D receptor, and regulates a wide variety of physiological responses.[8, 29] NCOA6 has been shown to associate with three co-activator complexes containing CBP/p300 and Set methyltransferases, and may play a crucial role in transcriptional activation by modulating chromatin structure through histone methylation.[29] NCOA2 inhibited Wnt/β-catenin signaling through upregulation of inhibitors and downregulation of stimulators of Wnt/β-catenin pathway.[62] The biological relevance in HGGs requires further investigations.

In this study, we focused on a newly identified, recurrent mutation in the *TOP2A* discovered in four GBM patients from the Polish HGG cohort. *TOP2A* mutations were described in many cancers, including breast and lung cancer[9], but not in gliomas. As this recurrent mutation in the *TOP2A* was discovered only in GBMs from the Polish HGG cohort (not in 21 GBMs from the Canadian Brain Tumor Bank), we consider this mutation as a putative founder mutation. Topoisomerases regulate DNA topology by introducing transient single- or double-strand DNA breaks, and may relax and unwind DNA, which is a critical step for proper DNA replication or transcription.[38, 50] Our computational predictions showed substitution of E948 is located in the Tower domain of the TOP2A protein in a close proximity to DNA and the replacement of a negatively charged residue E to a polar residue Q changes an electrostatic potential of the protein. This prediction has been confirmed by comparing DNA binding and enzymatic activities of recombinant WT and variant TOP2A. As TOP2A is a large protein, we prepared constructs expressing the 890-996 aa N-terminus of TOP2A (comprising only the Tower domain) and the 431-1193 aa N-terminus of TOP2A (encompassing the entire DNA-Gate domain). Biochemical assays showed that the TOP2A variant E948Q binds to DNA more strongly and has a higher supercoil relaxation activity than WT TOP2A proteins.

The presented evidence suggests that TOP2A mutant acquires new functions and exhibits a gain-of-function phenotype. Mutant TOP2A binds more strongly to DNA, exhibits a higher unwinding activity resulting in expression of genes with a high AT content and abundance of mRNAs with more intronic regions. These transcriptomic changes suggest either an increased transcriptomic rate or splicing aberrations. TOP2A mutant GBMs show activation of genes that encode proteins with an endopeptidase activity implicated in cancer invasion. Acquisition of these properties may lead to an increased migratory and invasive phenotype that leads to poor outcome for the patients.

In summary, this study revealed new insights into a landscape of genomic alterations in the high grade gliomas. In particular, a novel, pathogenic mutation in the *TOP2A* gene has been found in glioblastomas from The Polish HGG cohort.

Acknowledgements

Studies were supported by the Foundation for Polish Science TEAM-TECH Core Facility project “NGS platform for comprehensive diagnostics and personalized therapy in neuro-oncology”. The use of CePT infrastructure financed by the European Union—The European Regional Development Fund within the Operational Programme "Innovative economy" for 2007-2013 is highly appreciated.

Declarations of interest

The authors declare that they have no conflict of interest

Data and Code Availability

The datasets supporting the current study have not been deposited in a public repository because they are in part a subject for other ongoing project but are available from the corresponding author on request.

References

1. Bax, B.D., Murshudov, G., Maxwell, A., and Germe, T. (2019). DNA Topoisomerase Inhibitors: Trapping a DNA-Cleaving Machine in Motion. J Mol Biol 431, 3427-3449.
2. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.
3. Borges, K.S., Castro-Gamero, A.M., Moreno, D.A., da Silva Silveira, V., Brassesco, M.S., de Paula Queiroz, R.G., de Oliveira, H.F., Carlotti, C.G., Scrideli, C.A., and Tone, L.G. (2012). Inhibition of Aurora kinases enhances chemosensitivity to temozolomide and causes radiosensitization in glioblastoma cells. J Cancer Res Clin Oncol 138, 405-414.
4. Brat, D.J., Verhaak, R.G., Aldape, K.D., Yung, W.K., Salama, S.R., Cooper, L.A., Rheinbay, E., Miller, C.R., Vitucci, M., Morozova, O., et al. (2015). Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med 372, 2481-2498.
5. Brennan, C.W., Verhaak, R.G., McKenna, A., Campos, B., Noushmehr, H., Salama, S.R., Zheng, S., Chakravarty, D., Sanborn, J.Z., Berman, S.H., et al. (2013). The somatic genomic landscape of glioblastoma. Cell 155, 462-477.
6. Network, C.G.A.R. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature 455, 1061-1068.
7. Ceccarelli, M., Barthel, F.P., Malta, T.M., Sabedot, T.S., Salama, S.R., Murray, B.A., Morozova, O., Newton, Y., Radenbaugh, A., Pagnotta, S.M., et al. (2016). Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. Cell 164, 550-563.
8. Chen, H., Lin, R.J., Xie, W., Wilpitz, D., and Evans, R.M. (1999). Regulation of hormone-induced histone hyperacetylation and gene activation via acetylation of an acetylase. Cell

98, 675-686.

9. Chen, T., Sun, Y., Ji, P., Kopetz, S., and Zhang, W. (2015). Topoisomerase IIα in chromosome instability and personalized cancer therapy. Oncogene 34, 4019-4031.

10. Corbett, K.D., and Berger, J.M. (2004). Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. Annu Rev Biophys Biomol Struct 33, 95-118.

11. Daguenet, E., Dujardin, G., and Valcárcel, J. (2015). The pathogenicity of splicing defects: mechanistic insights into pre-mRNA processing inform novel therapeutic approaches. EMBO Rep 16, 1640-1655.

12. Dawson, M.A., and Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. Cell 150, 12-27.

13. de Almeida Magalhães, T., de Sousa, G.R., Alencastro Veiga Cruzeiro, G., Tone, L.G., Valera, E.T., and Borges, K.S. (2020). The therapeutic potential of Aurora kinases targeting in glioblastoma: from preclinical research to translational oncology. J Mol Med (Berl) 98, 495-512.

14. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15-21.

15. Eckel-Passow, J.E., Lachance, D.H., Molinaro, A.M., Walsh, K.M., Decker, P.A., Sicotte, H., Pekmezci, M., Rice, T., Kosel, M.L., Smirnov, I.V., et al. (2015). Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. N Engl J Med 372, 2499-2508.

16. Fang, H., Bergmann, E.A., Arora, K., Vacic, V., Zody, M.C., Iossifov, I., O'Rawe, J.A., Wu, Y., Jimenez Barron, L.T., Rosenbaum, J., et al. (2016). Indel variant analysis of short-

read sequencing data with Scalpel. Nat Protoc 11, 2529-2548.

17. Frattini, V., Trifonov, V., Chan, J.M., Castano, A., Lia, M., Abate, F., Keir, S.T., Ji, A.X., Zoppoli, P., Niola, F., et al. (2013). The integrated landscape of driver genomic alterations in glioblastoma. Nat Genet 45, 1141-1149.

18. Hu, Z.Y., Liu, Y.P., Xie, L.Y., Wang, X.Y., Yang, F., Chen, S.Y., and Li, Z.G. (2016). AKAP-9 promotes colorectal cancer development by regulating Cdc42 interacting protein 4. Biochim Biophys Acta 1862, 1172-1181.

19. Kanemaru, Y., Natsumeda, M., Okada, M., Saito, R., Kobayashi, D., Eda, T., Watanabe, J., Saito, S., Tsukamoto, Y., Oishi, M., et al. (2019). Dramatic response of BRAF V600E-mutant epithelioid glioblastoma to combination therapy with BRAF and MEK inhibitor: establishment and xenograft of a cell line to predict clinical efficacy. Acta Neuropathol Commun 7, 119.

20. Khasraw, M., and Lassman, A.B. (2010). Advances in the treatment of malignant gliomas. Curr Oncol Rep 12, 26-33.

21. Kim, G., and Ko, Y.T. (2020). Small molecule tyrosine kinase inhibitors in glioblastoma. Arch Pharm Res 43, 385-394.

22. Kjällquist, U., Erlandsson, R., Tobin, N.P., Alkodsi, A., Ullah, I., Stålhammar, G., Karlsson, E., Hatschek, T., Hartman, J., Linnarsson, S., et al. (2018). Exome sequencing of primary breast cancers with paired metastatic lesions reveals metastasis-enriched mutations in the A-kinase anchoring protein family (AKAPs). BMC Cancer 18, 174.

23. Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., and Wilson, R.K. (2012). VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res 22, 568-576.

24. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754-1760.

25. Li, Y., Zhang, M., Sheng, M., Zhang, P., Chen, Z., Xing, W., Bai, J., Cheng, T., Yang, F.C., and Zhou, Y. (2018). Therapeutic potential of GSK-J4, a histone demethylase KDM6B/JMJD3 inhibitor, for acute myeloid leukemia. J Cancer Res Clin Oncol 144, 1065-1077.

26. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923-930.

27. Liau, B.B., Sievers, C., Donohue, L.K., Gillespie, S.M., Flavahan, W.A., Miller, T.E., Venteicher, A.S., Hebert, C.H., Carey, C.D., Rodig, S.J., et al. (2017). Adaptive Chromatin Remodeling Drives Glioblastoma Stem Cell Plasticity and Drug Tolerance. Cell Stem Cell 20, 233-246.e237.

28. Louis, D.N., Perry, A., Reifenberger, G., von Deimling, A., Figarella-Branger, D., Cavenee, W.K., Ohgaki, H., Wiestler, O.D., Kleihues, P., and Ellison, D.W. (2016). The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol 131, 803-820.

29. Mahajan, M.A., and Samuels, H.H. (2008). Nuclear receptor coactivator/coregulator NCoA6(NRC) is a pleiotropic coregulator involved in transcription, cell survival, growth and development. Nucl Recept Signal 6, e002.

30. Maleszewska, M., and Kaminska, B. (2013). Is glioblastoma an epigenetic malignancy? Cancers (Basel) 5, 1120-1139.

31. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res 28,

1747-1756.

32. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol 17, 122.

33. Noushmehr, H., Weisenberger, D.J., Diefes, K., Phillips, H.S., Pujara, K., Berman, B.P., Pan, F., Pelloski, C.E., Sulman, E.P., Bhat, K.P., et al. (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell 17, 510-522.

34. Omuro, A., and DeAngelis, L.M. (2013). Glioblastoma and other malignant gliomas: a clinical review. JAMA 310, 1842-1850.

35. Ostrom, Q.T., Gittleman, H., Xu, J., Kromer, C., Wolinsky, Y., Kruchko, C., and Barnholtz-Sloan, J.S. (2016). CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2009-2013. Neuro Oncol 18, v1-v75.

36. Parsons, D.W., Jones, S., Zhang, X., Lin, J.C., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.M., Gallia, G.L., et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. Science 321, 1807-1812.

37. Paul, Y., Mondal, B., Patil, V., and Somasundaram, K. (2017). DNA methylation signatures for 2016 WHO classification subtypes of diffuse gliomas. Clin Epigenetics 9, 32.

38. Pommier, Y., Sun, Y., Huang, S.N., and Nitiss, J.L. (2016). Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. Nat Rev Mol Cell Biol 17, 703-721.

39. Roca, J. (2009). Topoisomerase II: a fitted mechanism for the chromatin landscape. Nucleic Acids Res 37, 721-730.

40. Romani, M., Daga, A., Forlani, A., Pistillo, M.P., and Banelli, B. (2019). Targeting of

Histone Demethylases KDM5A and KDM6B Inhibits the Proliferation of Temozolomide-Resistant Glioblastoma Cells. Cancers (Basel) 11.

41. Shah, M.A., Denton, E.L., Arrowsmith, C.H., Lupien, M., and Schapira, M. (2014). A global assessment of cancer genomic alterations in epigenetic mechanisms. Epigenetics Chromatin 7, 29.

42. Shen, J.C., Rideout, W.M., and Jones, P.A. (1994). The rate of hydrolytic deamination of 5-methylcytosine in double-stranded DNA. Nucleic Acids Res 22, 972-976.

43. Stupp, R., Mason, W.P., van den Bent, M.J., Weller, M., Fisher, B., Taphoorn, M.J., Belanger, K., Brandes, A.A., Marosi, C., Bogdahn, U., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. N Engl J Med 352, 987-996.

44. Suzuki, H., Aoki, K., Chiba, K., Sato, Y., Shiozawa, Y., Shiraishi, Y., Shimamura, T., Niida, A., Motomura, K., Ohka, F., et al. (2015). Mutational landscape and clonal architecture in grade II and III gliomas. Nat Genet 47, 458-468.

45. Tamborero, D., Gonzalez-Perez, A., and Lopez-Bigas, N. (2013). OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. Bioinformatics 29, 2238-2244.

46. Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A.D., Nueda, M.J., Ferrer, A., and Conesa, A. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. Nucleic Acids Res 43, e140.

47. Tessarz, P., and Kouzarides, T. (2014). Histone core modifications regulating nucleosome structure and dynamics. Nat Rev Mol Cell Biol 15, 703-708.

48. Venkatesh, D., Mruk, D., Herter, J.M., Cullere, X., Chojnacka, K., Cheng, C.Y., and Mayadas, T.N. (2016). AKAP9, a Regulator of Microtubule Dynamics, Contributes to

Blood-Testis Barrier Function. Am J Pathol 186, 270-284.

49. Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell 17, 98-110.

50. Wang, J.C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. Nat Rev Mol Cell Biol 3, 430-440.

51. Wang, L., Wang, S., and Li, W. (2012). RSeQC: quality control of RNA-seq experiments. Bioinformatics 28, 2184-2185.

52. Weller, M., Butowski, N., Tran, D.D., Recht, L.D., Lim, M., Hirte, H., Ashby, L., Mechtler, L., Goldlust, S.A., Iwamoto, F., et al. (2017). Rindopepimut with temozolomide for patients with newly diagnosed, EGFRvIII-expressing glioblastoma (ACT IV): a randomised, double-blind, international phase 3 trial. Lancet Oncol 18, 1373-1385.

53. Weller, M., van den Bent, M., Tonn, J.C., Stupp, R., Preusser, M., Cohen-Jonathan-Moyal, E., Henriksson, R., Le Rhun, E., Balana, C., Chinot, O., et al. (2017). European Association for Neuro-Oncology (EANO) guideline on the diagnosis and treatment of adult astrocytic and oligodendroglial gliomas. Lancet Oncol 18, e315-e329.

54. Weller, M., Wick, W., Aldape, K., Brada, M., Berger, M., Pfister, S.M., Nishikawa, R., Rosenthal, M., Wen, P.Y., Stupp, R., et al. (2015). Glioma. Nat Rev Dis Primers 1, 15017.

55. Wen, P.Y., Macdonald, D.R., Reardon, D.A., Cloughesy, T.F., Sorensen, A.G., Galanis, E., Degroot, J., Wick, W., Gilbert, M.R., Lassman, A.B., et al. (2010). Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology

working group. J Clin Oncol 28, 1963-1972.

56. Wendorff, T.J., Schmidt, B.H., Heslop, P., Austin, C.A., and Berger, J.M. (2012). The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. J Mol Biol 424, 109-124.

57. Wesseling, P., and Capper, D. (2018). WHO 2016 Classification of gliomas. Neuropathol Appl Neurobiol 44, 139-150.

58. Westphal, M., Maire, C.L., and Lamszus, K. (2017). EGFR as a Target for Glioblastoma Treatment: An Unfulfilled Promise. CNS Drugs 31, 723-735.

59. Wijayatunge, R., Liu, F., Shpargel, K.B., Wayne, N.J., Chan, U., Boua, J.V., Magnuson, T., and West, A.E. (2018). The histone demethylase Kdm6b regulates a mature gene expression program in differentiating cerebellar granule neurons. Mol Cell Neurosci 87, 4-17.

60. Woo, P.Y.M., Lam, T.C., Pu, J.K.S., Li, L.F., Leung, R.C.Y., Ho, J.M.K., Zhung, J.T.F., Wong, B., Chan, T.S.K., Loong, H.H.F., et al. (2019). Regression of. Oncotarget 10, 3818-3826.

61. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284-287.

62. Yu, J., Wu, W.K., Liang, Q., Zhang, N., He, J., Li, X., Zhang, X., Xu, L., Chan, M.T., Ng, S.S., et al. (2016). Disruption of NCOA2 by recurrent fusion with LACTB2 in colorectal cancer. Oncogene 35, 187-195.

63. Zhao, H.F., Wang, J., Shao, W., Wu, C.P., Chen, Z.P., To, S.T., and Li, W.P. (2017). Recent advances in the use of PI3K inhibitors for glioblastoma multiforme: current preclinical and clinical development. Mol Cancer 16, 100.

64. Zhao, S., Zhang, Y., Gamini, R., Zhang, B., and von Schack, D. (2018). Evaluation of two main RNA-seq approaches for gene quantification in clinical RNA sequencing: polyA+ selection versus rRNA depletion. Sci Rep 8, 4781.

Figure legend

Fig. 1. **Identification of most frequent mutations in high grade gliomas revealed novel mutations in the *TOP2A* gene**. A – Plot showing potential oncogenic drivers (OncodriveCLUST function) in the cohort of high grade gliomas (HGGs), red dots represent potential drivers, with FDR-corrected OncodriveCLUST P value<0.025; size of the dot corresponds to a number of the alterations that were present in the cluster. **A, B** - Genes marked in bold are new potential oncodrivers in HGGs, as these genes were not found to be mutated in TCGA LGG/GBM datasets. **B** – Oncostrip plot showing genes discovered in the OncodriveCLUST analysis, each column corresponds to a glioma patient, while each row corresponds to one altered gene. C –*TOP2A* expression in normal brain (NB) and WHO glioma grades (GII, GIII and GIV) in the TCGA dataset. P-values based on Wilcoxon signed-rank test were obtained (* P < 0.05, ** P < 0.01 and *** P < 0.001). D – TOP2A lolliplot showing the identified substitutions mapped on the protein structure of TOP2A. A number of dots represents the number of patients in which the particular alteration was found.

Fig. 2. Identification of a novel, recurrent mutation in the *AKAP9* gene in HGGs. **A** – Oncostrip showing most frequently altered genes in high-grade gliomas, color coded by the specific mutation type. **B** – AKAP9 lolliplot showing the identified frameshift mapped on the protein structure. Only potential protein damaging somatic mutations (scores < 0.05) and low (ExAC < 0.001) or non-existent minor allele frequencies in population are represented here. **C** – mRNA *AKAP9* expression in normal brain (NB) and WHO glioma grades (GII, GIII and GIV) in the TCGA dataset. P-values based on Wilcoxon signed-rank test were obtained (* P < 0.05, ** P < 0.01 and *** P < 0.001). **D** – Kaplan-Meier overall survival curve from TCGA. Patients alive at the time of the analysis were censored at the time they were last followed up. The medial overall survival was ~480 days in the low *AKAP9* mRNA expression and ~1950 days in the high *AKAP9*

mRNA expression. P-values based on Log Rank Test were obtained (* $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$).

Fig. 3. **Genomic alterations of *TOP2A* mutated samples. A** – Oncoplot showing most frequently altered genes in *TOP2A* mutated HGGs, each column corresponds to a gliomasample, while each row corresponds to one altered gene. Log2 number of genetic alterations in *TOP2A* (**B**) and a fraction of C to T nucleotide changes (**C**) in the *TOP2A* mutated samples versus the samples with other mutations: *IDH1*, *IDH1* and *TP53*, *TP53* and *PTEN*. Sample assignment to each of described groups is explained in the Supp. Fig. 3. **D** - Kaplan-Meier overall survival curve for patients with high or low expression in the TCGA dataset. Patients alive at the time of the analysis were censored at the time they were last followed up. The median overall survival was ~500 days in the HIGH *TOP2A* mRNA expression group and ~2000 days in the LOW *TOP2A* mRNA expression group. P-values based on Log Rank Test were obtained. E – Kaplan-Meier overall survival curve in this cohort. The median overall survival was ~150 days in the HGG patients with the *TOP2A* mutated (TOP2A MUT (E948Q) and ~400 days in the patients with wild-type *TOP2A* (TOP2A WT). P-values based on Log Rank Test were obtained.

Fig. 4. Substitution of E948 to Q changes the electrostatic potential of TOP2A. A – Schematic representation of a TOP2A domain structure. Functional regions are colored and labeled: an ATPase domain, a TOPRIM - Mg^{2+} ion-binding topoisomerase/primase fold, a WHD – winged helix domain, a TOWER, CTD – C-terminal domain. The three dimerization regions are marked: N-gate, DNA-gate and C-gate. **B** – Secondary structure representation of TOP2A 431-1193 bound to DNA (crystal structure PDB code: 4FM9). For simplification only a TOP2A protomer is shown. An amino acid residue E948 which is substituted to Q in a TOP2A variant is

represented as spheres on side chains. **C** – Electrostatic surface potential of TOP2A 431-1193 WT and E948Q variant. Electrostatic potential was calculated using a PyMOL program.

Fig. 5. TOP2A E948Q variant exhibits a higher DNA binding capacity than WT. A –Purified TOP2A proteins. Schematic representation of purified proteins is placed in the top of the figure: TOP2A 890-996 aa (only the TOWER domain) and TOP2A 431-1193 aa (the DNA-gate). Amino acid residue E948 which is substituted to Q in the TOP2A variant is labeled in the schemes. SDS-PAGE of TOP2A purified proteins. Purified proteins were resolved in a 6% and 15% SDS–PAGE and stained with Coomassie brilliant blue. **B** – DNA-binding activity of TOP2A WT and E948Q. EMSA was performed using the LightShift Chemiluminescent EMSA Kit. The binding reactions were resolved in 6% polyacrylamide gels, transferred onto a membrane and detected by chemiluminescence. Binding activities are expressed in relation to that of TOP2A WT (which was set to 1) and are presented as mean ± standard deviations (error bars) that were calculated from two independent measurements. Nonspecific binding level was measured in a reaction without a protein and was subtracted from the results.

Fig. 6. Activity of TOP2A WT and E948Q variant. Supercoil relaxation assay of a plasmid DNA induced by TOP2A 431-1193 WT or E948Q proteins. The reaction mixtures were resolved by electrophoresis on 0.7% agarose gel followed by staining with SimplySafe dye and visualization using a Chemidoc camera. The protein activities were measured by its ability to unwind a supercoiled plasmid DNA (SC) to an open circular DNA (OC). The intensities of SC and OC plasmid DNA were measured in relation to a control DNA without an added protein (which was set to 100%) and are presented as means ± standard deviations (error bars) that were calculated from three independent measurements.

Fig.7. Transcriptomic alterations in *TOP2A* mutated HGGs. A – Fractions of RNA-seq reads mapped to intron regions. B - GC content profile of gene expression in *TOP2* mutant (MUT) and wild type (WT) HGGs. C – Differences in expression of genes that are not expressed in the normal brain in *TOP2A* MUT and WT gliomas. Genes were preselected by filtering out all the genes that were expressed in any of the cells from Stanford Human Brain RNAseq containing data from sorted healthy brain cells. Additionally, all genes expressed in normal brain samples were filtered out. D – Gene Ontology terms show functional annotations of genes not expressed in the normal brain that became activated in *TOP2A* mutant HGGs. P-values for plots A based on Wilcoxon signed-rank test, for the panel C on Kolmogonorov-Smirnov test (* $P < 0.05$, ** $P < 0.01$ and *** $P < 0.001$).

Supp. Fig. 1. **Clinical characteristics of patients**. A – patient age distribution. B – patient sex distribution. C – percentages of HGG patients of a specific WHO grade). D – percentages of patients with a specific localization of the tumor in the brain.

Supp. Fig. 2. The detailed landscape of somatic and rare germ-line variants in high-grade gliomas. A – Oncostrip showing most frequently altered genes in high-grade gliomas, each column corresponds to a glioma sample, while each row corresponds to one altered gene. B – Plot of oncogenic pathways more frequently altered in the studied cohort of high-grade gliomas. Size and color intensity of dots mark how many genes from a given pathway were mutated in the studied cohort. C – Oncostrip plot of RTK-RAS pathway which emerges as the most frequently altered in this cohort. D – Most commonly mutated pfam domains, size of a dot corresponds to a number of genes with that particular domain being altered in the studied cohort.

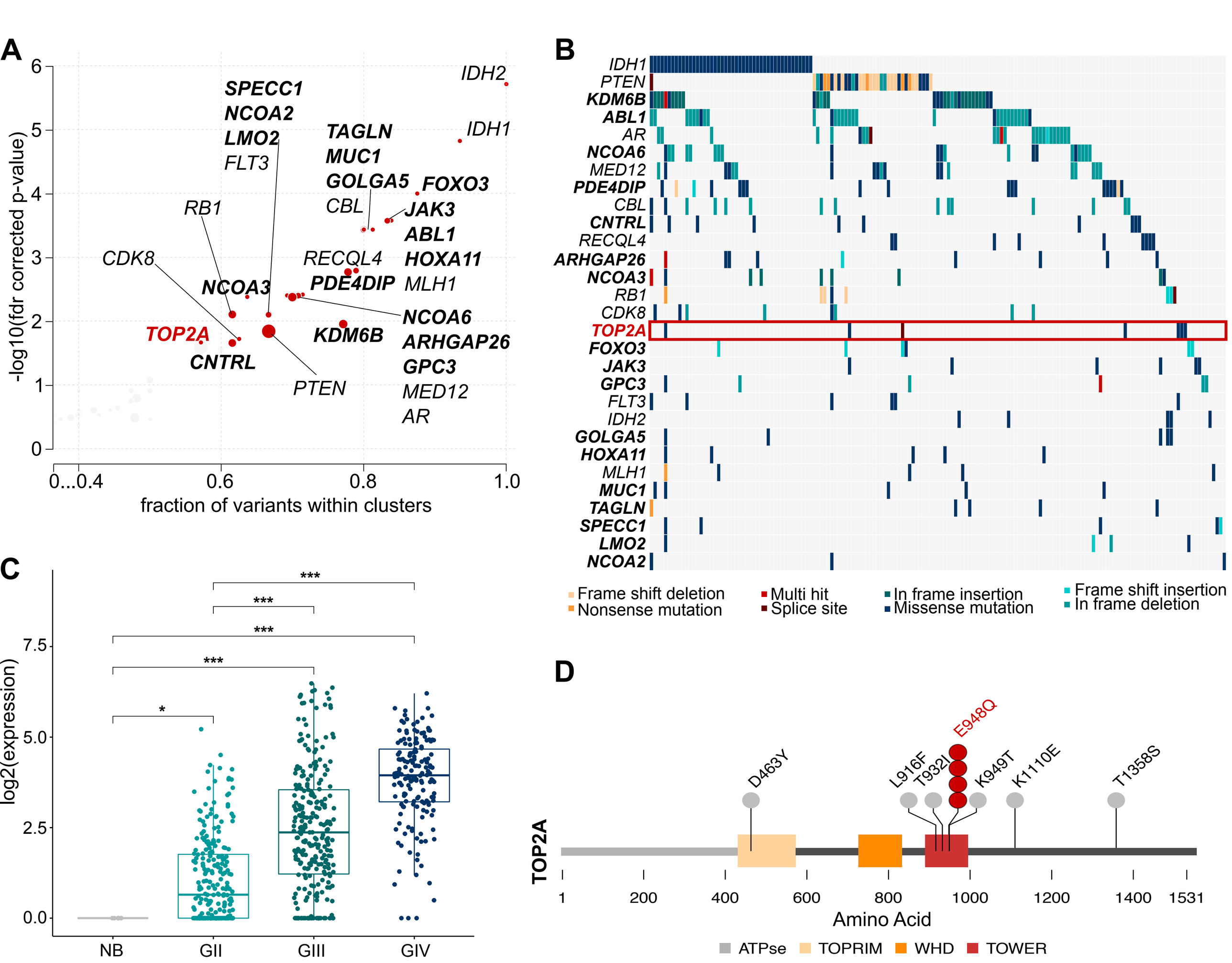
Supp. Fig. 3. Oncostrip showing most frequently altered genes in high-grade gliomas, each column corresponds to a glioma sample, while each row corresponds to one altered gene. Vertical lines demark the groups with mutated *IDH1* and*TP53*, *TP53*, *IDH1* and *PTEN* that were used to achieve data presented in the Figs. 3B and C.

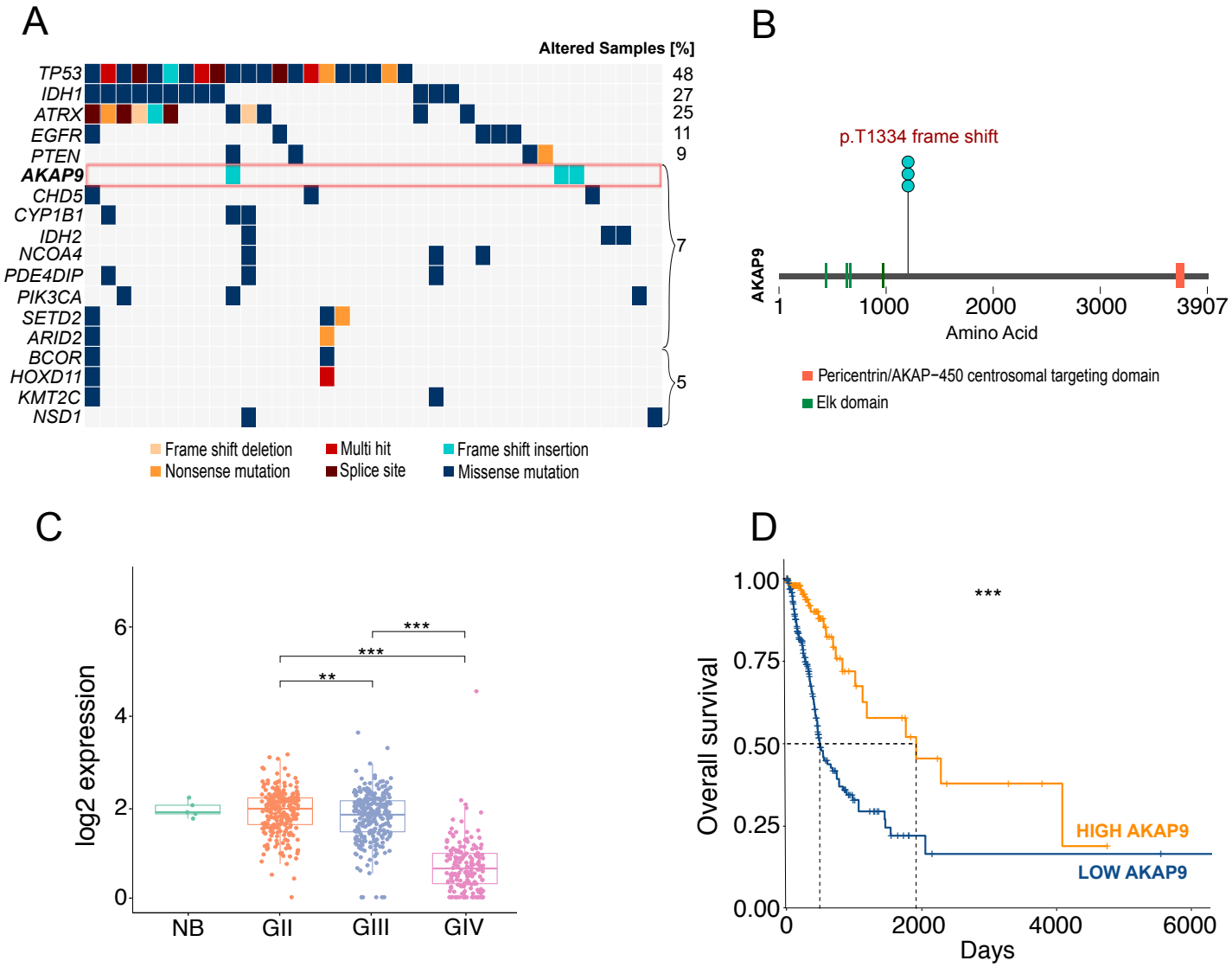
Supplementary Table 1. Clinical characteristics of the tumor samples used in the study. The following coding was used: "M" - male, "F" - female, "ND" - no data

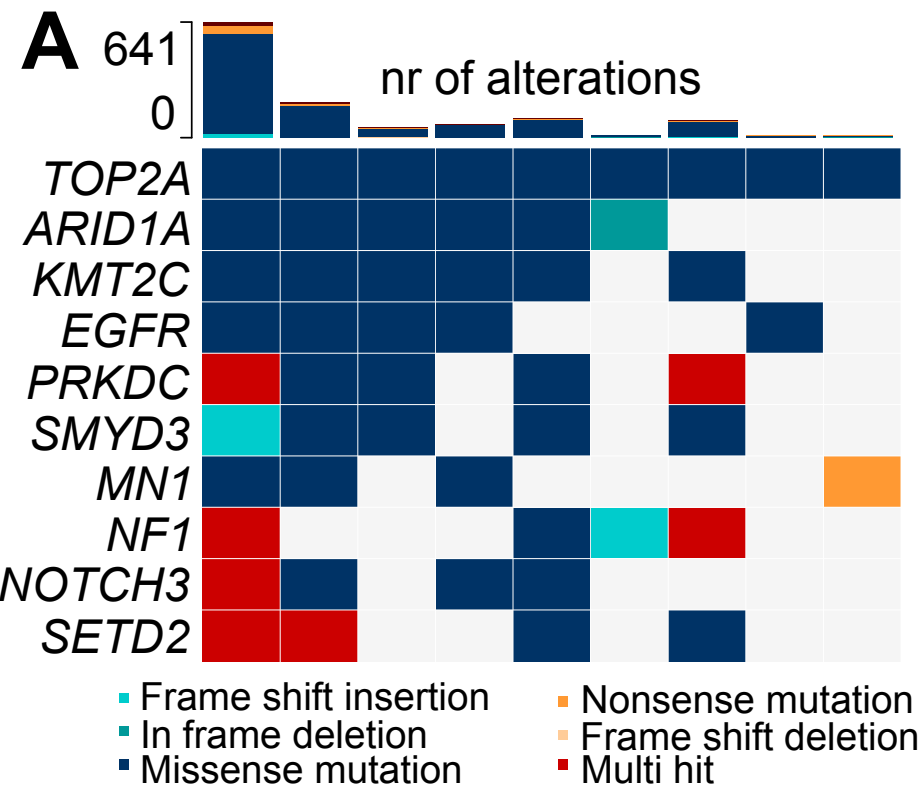
Supplementary Table 2. List of genetic changes identified in the TOP2A gene using targeted next generation sequencing of 182 glioma samples. The table contains the most important information regarding the identified variants

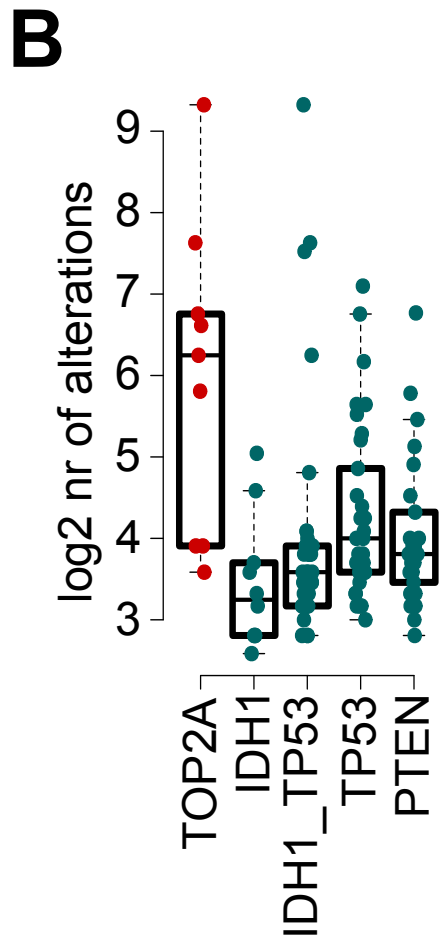
Supplementary Table 3. Comparison of the number of identified somatic variants using two different data analysis pipelines. Data analysis pipeline for identifying only somatic variants vs data analysis pipeline for identifying somatic and rare germline variants were compared.

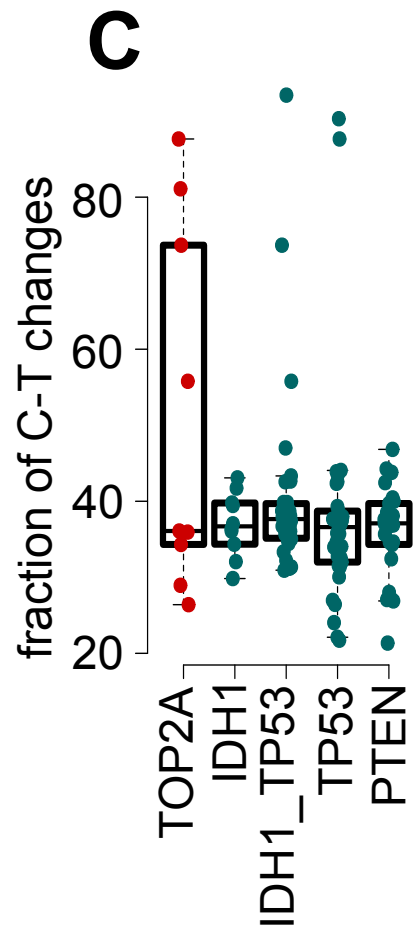
Supplementary Table 4. The most important features of the amino acid substitutions caused by mutations identified in *TOP2A* gene. Some of the identified changes may increase or decrease DNA binding.

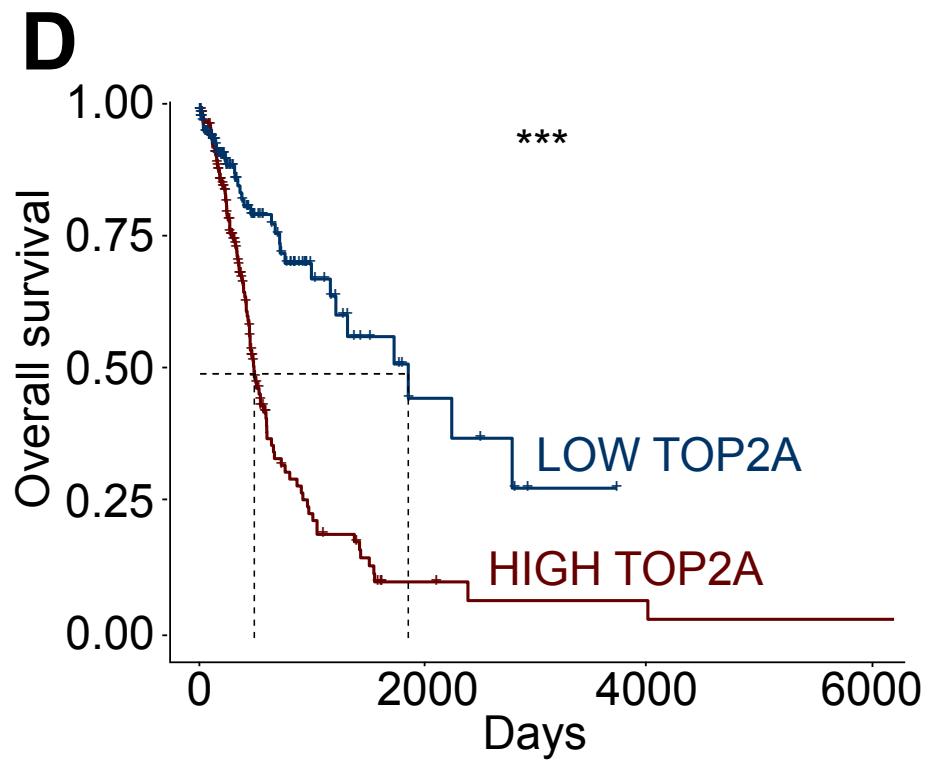

A
-log10(fdr corrected p-value)
fraction of variants within clusters
0...0.4
0.6
0.8
1.0
IDH2
IDH1
SPECC1
NCOA2
LMO2
FLT3
TAGLN
MUC1
GOLGA5
FOXO3
CBL
JAK3
ABL1
HOXA11
MLH1
RB1
CDK8
RECQL4
PDE4DIP
NCOA3
NCOA6
ARHGAP26
GPC3
MED12
AR
TOP2A
CNTRL
KDM6B
PTEN
B
IDH1
PTEN
KDM6B
ABL1
AR
NCOA6
MED12
PDE4DIP
CBL
CNTRL
RECQL4
ARHGAP26
NCOA3
RB1
CDK8
TOP2A
FOXO3
JAK3
GPC3
FLT3
IDH2
GOLGA5
HOXA11
MLH1
MUC1
TAGLN
SPECC1
LMO2
NCOA2
Frame shift deletion
Nonsense mutation
Multi hit
Splice site
In frame insertion
Missense mutation
Frame shift insertion
In frame deletion
C
log2(expression)
NB
GII
GIII
GIV
D
TOP2A
Amino Acid
D463Y
L916F
T932I
E948Q
K949T
K1110E
T1358S
ATPse
TOPRIM
WHD
TOWER

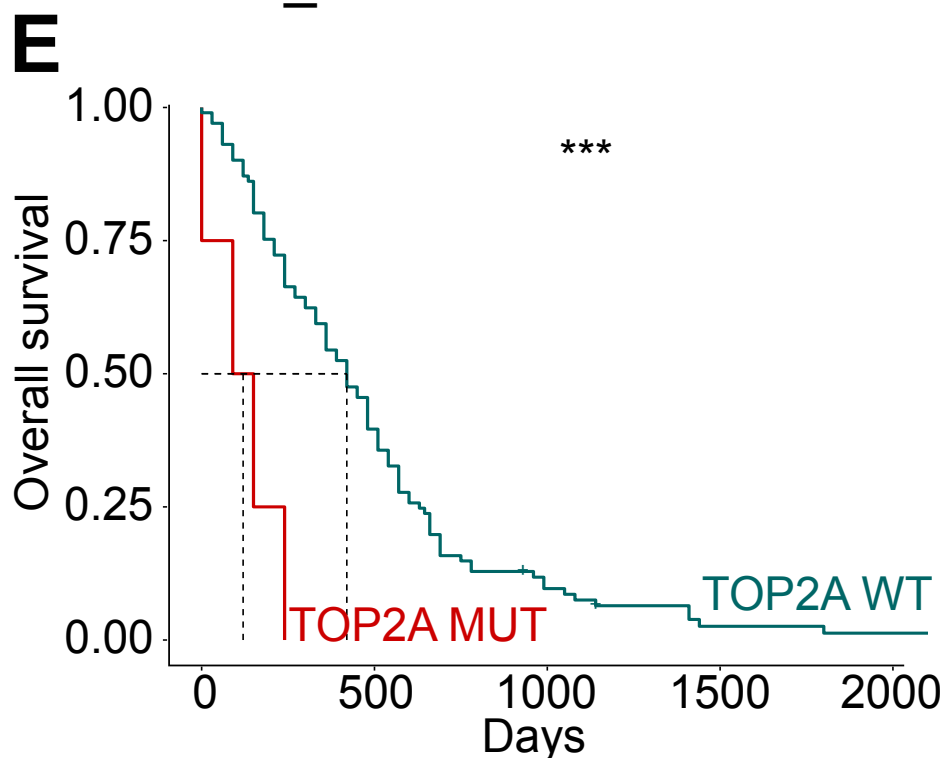

A
Altered Samples [%]
TP53
IDH1
ATRX
EGFR
PTEN
AKAP9
CHD5
CYP1B1
IDH2
NCOA4
PDE4DIP
PIK3CA
SETD2
ARID2
BCOR
HOXD11
KMT2C
NSD1
48
27
25
11
9
7
5
Frame shift deletion
Multi hit
Frame shift insertion
Nonsense mutation
Splice site
Missense mutation
B
p.T1334 frame shift
AKAP9
1
1000
2000
3000
3907
Amino Acid
Pericentrin/AKAP−450 centrosomal targeting domain
Elk domain
C
log2 expression
0
2
4
6
**
***
***
NB
GII
GIII
GIV
D
Overall survival
1.00
0.75
0.50
0.25
0.00
***
HIGH AKAP9
LOW AKAP9
0
2000
4000
6000
Days

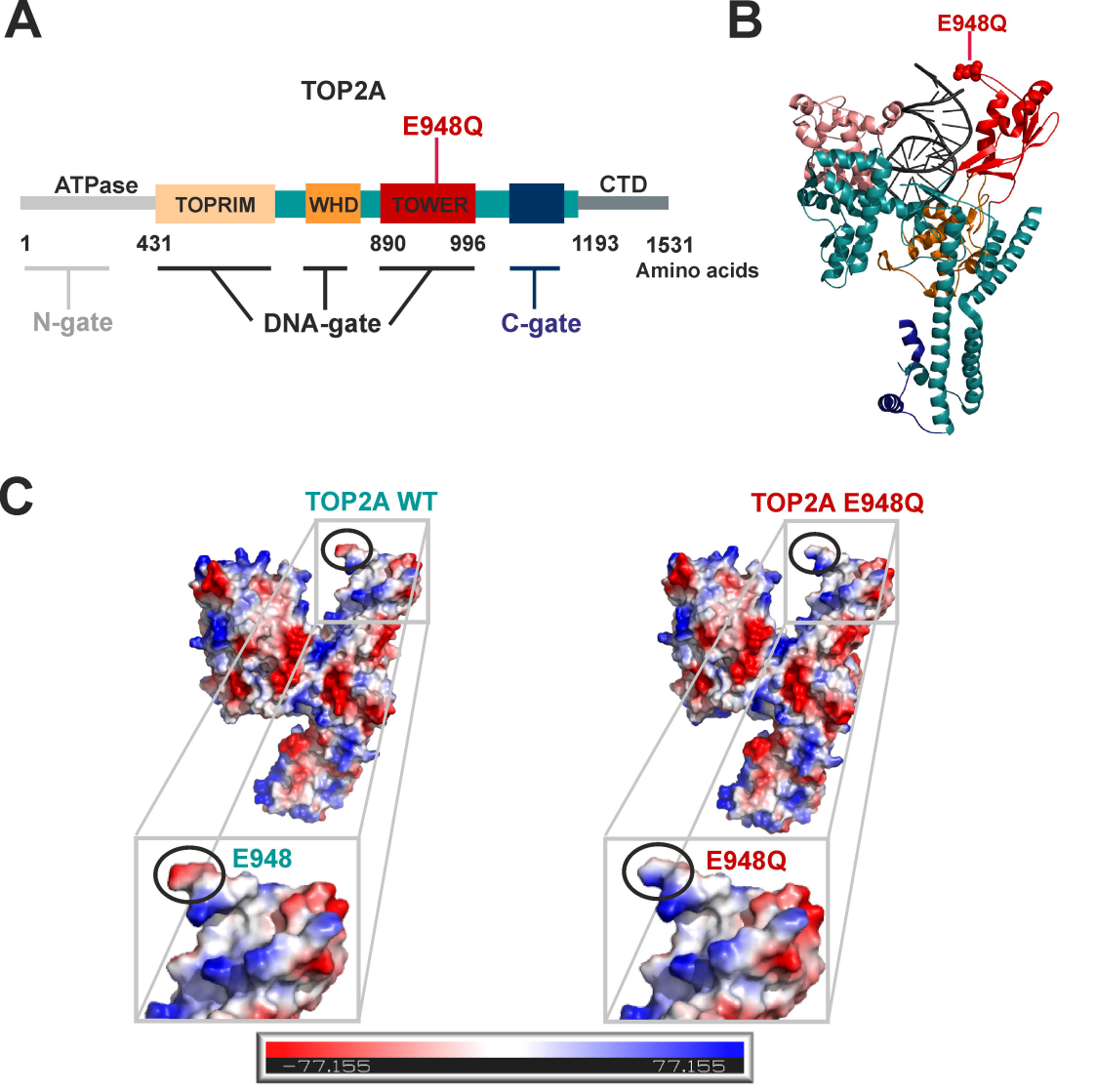

A
641
0
nr of alterations
TOP2A
ARID1A
KMT2C
EGFR
PRKDC
SMYD3
MN1
NF1
NOTCH3
SETD2
Frame shift insertion
In frame deletion
Missense mutation
Nonsense mutation
Frame shift deletion
Multi hit

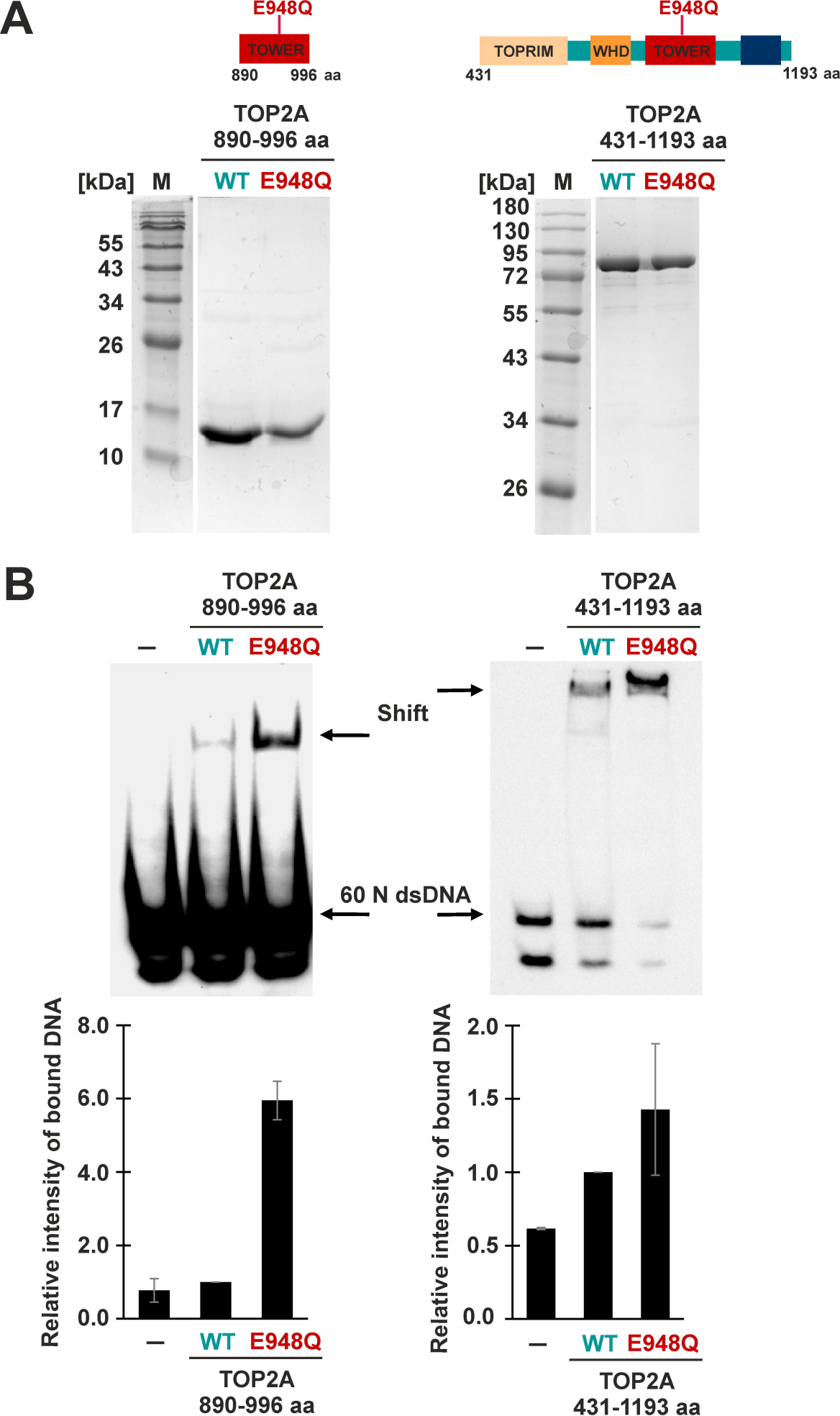

B
log2 nr of alterations
9
8
7
6
5
4
3
TOP2A
IDH1
IDH1_TP53
TP53
PTEN

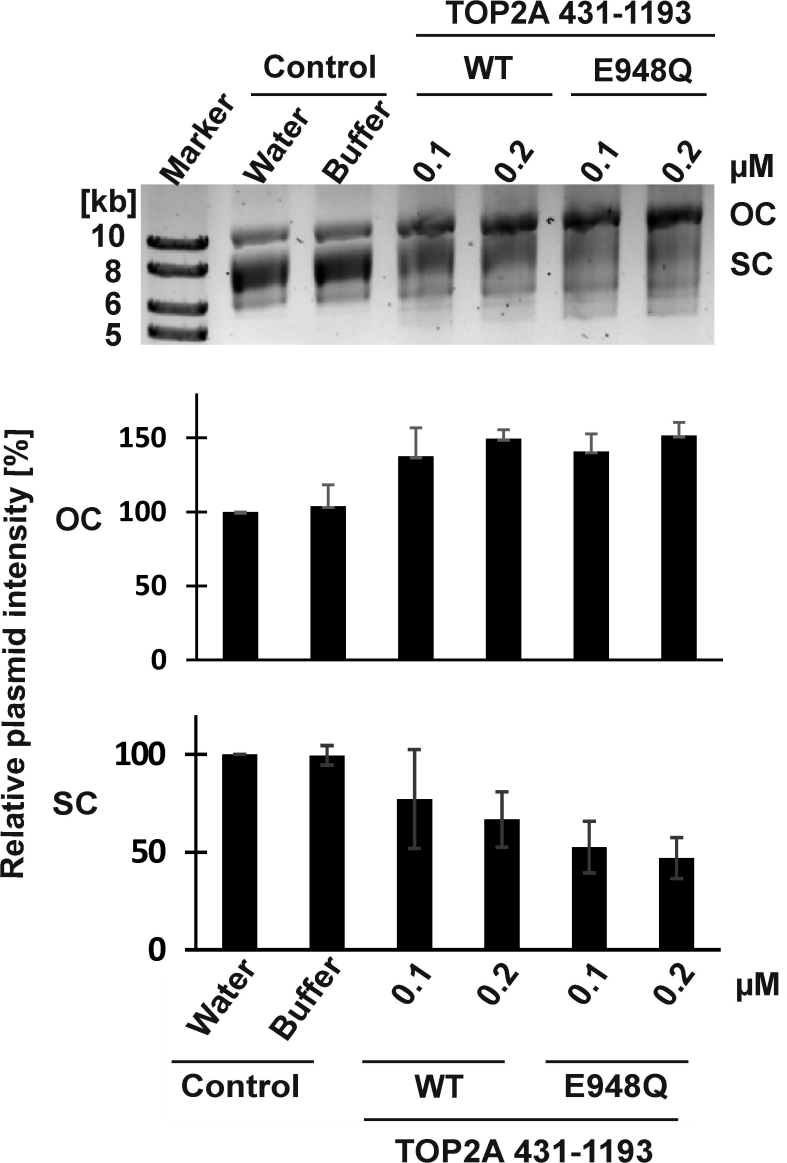

C
fraction of C-T changes
80
60
40
20
TOP2A
IDH1
IDH1_TP53
TP53
PTEN

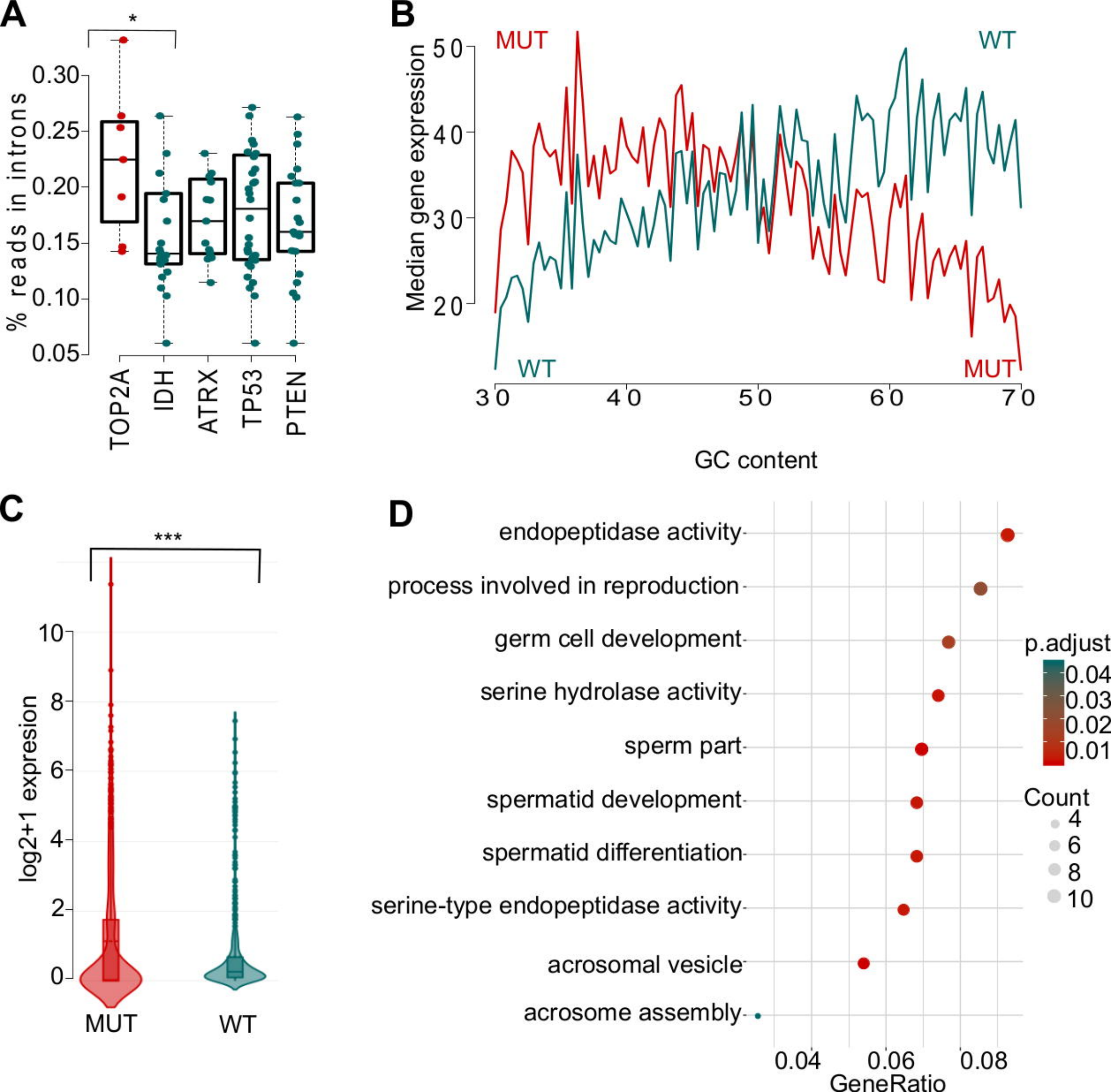

D
Overall survival
1.00
0.75
0.50
0.25
0.00
***
LOW TOP2A
HIGH TOP2A
0
2000
4000
6000
Days


E
Overall survival
1.00
0.75
0.50
0.25
0.00
***
TOP2A MUT
TOP2A WT
0
500
1000
1500
2000
Days


A
TOP2A
E948Q
ATPase
TOPRIM
WHD
TOWER
CTD
1
431
890
996
1193
1531
Amino acids
N-gate
DNA-gate
C-gate
B
E948Q
C
TOP2A WT
TOP2A E948Q
E948
E948Q
−77.155
77.155


A
E948Q
TOWER
890
996 aa
E948Q
TOPRIM
WHD
TOWER
431
1193 aa
TOP2A
890-996 aa
[kDa]
M
WT
E948Q
55
43
34
26
17
10
TOP2A
431-1193 aa
[kDa]
M
WT
E948Q
180
130
95
72
55
43
34
26
B
TOP2A
890-996 aa
–
WT
E948Q
TOP2A
431-1193 aa
–
WT
E948Q
Shift
60 N dsDNA
Relative intensity of bound DNA
8.0
6.0
4.0
2.0
0.0
–
WT
E948Q
TOP2A
890-996 aa
Relative intensity of bound DNA
2.0
1.5
1.0
0.5
0.0
–
WT
E948Q
TOP2A
431-1193 aa


TOP2A 431-1193
Control
WT
E948Q
Marker
Water
Buffer
0.1
0.2
0.1
0.2
µM
[kb]
10
8
6
5
OC
SC
Relative plasmid intensity [%]
OC
150
100
50
0
SC
100
50
0
Water
Buffer
0.1
0.2
0.1
0.2
µM
Control
WT
E948Q
TOP2A 431-1193


A
*
% reads in introns
0.30
0.25
0.20
0.15
0.10
0.05
TOP2A
IDH
ATRX
TP53
PTEN
B
MUT
WT
WT
MUT
Median gene expression
50
40
30
20
30
40
50
60
70
GC content
C
***
log2+1 expresion
10
8
6
4
2
0
MUT
WT
D
endopeptidase activity
process involved in reproduction
germ cell development
serine hydrolase activity
sperm part
spermatid development
spermatid differentiation
serine-type endopeptidase activity
acrosomal vesicle
acrosome assembly
0.04
0.06
0.08
GeneRatio
p.adjust
0.04
0.03
0.02
0.01
Count
4
6
8
10