Received: 6 February 2023 | Revised: 28 April 2023 | Accepted: 8 May 2023

DOI: 10.1002/ijc.34631

CANCER GENETICS AND EPIGENETICS

IJC INTERNATIONAL JOURNAL of CANCER

Targeted sequencing of cancer-related genes reveals a recurrent TOP2A variant which affects DNA binding and coincides with global transcriptional changes in glioblastoma

Bartlomiej Gielniewski[1] | Katarzyna Poleszak[1] | Adria-Jaume Roura[1] | Paulina Szadkowska[1] | Karol Jacek[1] | Sylwia K. Krol[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[4,5] | Cezary Szczylik[4] | Ryszard Czepko[6] | Mariusz Banach[6] | Wojciech Kaspera[7] | Wojciech Szopa[7] | Mateusz Bujko[5] | Bartosz Czapski[8] | Miroslaw Zabek[8,9] | Ewa Iżycka-Świeszewska[10] | Wojciech Kloc[11,12] | Pawel Nauman[13,14] | Joanna Cieslewicz[15] | Wieslawa Grajkowska[16] | Natalia Morosini[17] | Houtan Noushmehr[17] | 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]The Maria Sklodowska-Curie National Research Institute of Oncology, Warsaw, Poland

[6]Department of Neurosurgery, Andrzej Frycz Modrzewski Krakow University, Krakow, Poland

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

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

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

[10]Medical University of Gdansk, Gdansk, Poland

[11]Department of Neurosurgery, Copernicus PL, Gdansk, Poland

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

[13]Institute of Psychiatry and Neurology, Warsaw, Poland

[14]Faculty of Medical and Health Sciences, Siedlce University of Natural Sciences and Humanities, Siedlce, Poland

[15]Gdansk University of Technology, Faculty of Chemistry, Gdansk, Poland

[16]Department of Pathology, The Children's Memorial Health Institute, Warsaw, Poland

[17]Department of Neurosurgery, Henry Ford Cancer Institute, Detroit, Michigan, USA

Abbreviations: ABL1, ABL proto-oncogene 1, nonreceptor tyrosine kinase; ARID1A, AT-rich interaction domain1A; ATM, ATM serine/threonine kinase; ATRX, ATP-dependent helicase; AURK, Aurora kinase; BRAF, B-raf proto-oncogene, serine/threonine kinase; BRCA1, BRCA1 DNA repair associated; DNA, deoxyribonucleic acid; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetic acid; EGA, European Genome-Phenome Archive; EGFR, epidermal growth factor receptor; EMSA, electrophoretic mobility shift assay; FOXO3, Forkhead box O3; GBM, glioblastoma; HGG, high-grade glioma; IDH, isocitrate dehydrogenase; KMT2C, lysine methyltransferase 2C; LB, lysogeny broth; MAF, minor allele frequency; NF1, neurofibromin 1; PDE4DIP, phosphodiesterase 4D interacting protein; PDGFRA, platelet-derived growth factor receptor alpha; PTEN, phosphatase and tensin homolog deleted on chromosome 10; RAD, radiation-repair gene; RECQL4, RecQ like helicase 4; RNA, ribonucleic acid; RTK, receptor tyrosine kinases; SIFT, scale-invariant feature transform; SNP, single nucleotide polymorphism; TCGA, The Cancer Genome Atlas project; TOP2A, topoisomerase II alpha; TP53, tumor protein 53; UTR, untranslated region; VAF, variant allele frequency; VEP, variant effect predictor; WHO, World Health Organization; WT, wild type.

Bartlomiej Gielniewski and Katarzyna Poleszak contributed equally to this study.

Correspondence
Bozena Kaminska and Bartosz Wojtas, Laboratory of Molecular Neurobiology, Nencki Institute of Experimental Biology, Polish Academy of Sciences, 3 Pasteur Str., 02-093 Warsaw, Poland.
Email: b.kaminska@nencki.edu.pl (B. Ka.) and
Email: b.wojtas@nencki.edu.pl (B. W.)

Funding information
Fundacja na rzecz Nauki Polskiej

Abstract

High-grade gliomas are aggressive, deadly primary brain tumors. Median survival of patients with glioblastoma (GBM, WHO grade 4) is 14 months and <10% of patients survive 2 years. Despite improved surgical strategies and forceful radiotherapy and chemotherapy, the prognosis of GBM patients is poor and did not improve over decades. We performed targeted next-generation sequencing with a custom panel of 664 cancer- and epigenetics-related genes, and searched for somatic and germline variants in 180 gliomas of different WHO grades. Herein, we focus on 135 GBM *IDH*-wild type samples. In parallel, mRNA sequencing was accomplished to detect transcriptomic abnormalities. We present the genomic alterations in high-grade gliomas and the associated transcriptomic patterns. Computational analyses and biochemical assays showed the influence of TOP2A variants on enzyme activities. In 4/135 *IDH*-wild type GBMs we found a novel, recurrent mutation in the *TOP2A* gene encoding topoisomerase 2A (allele frequency [AF] = 0.03, 4/135 samples). Biochemical assays with recombinant, wild type (WT) and variant proteins demonstrated stronger DNA binding and relaxation activity of the variant protein. GBM patients carrying the altered *TOP2A* had shorter overall survival (median OS 150 vs 500 days, $P = .0018$). In the GBMs with the *TOP2A* variant we found transcriptomic alterations consistent with splicing dysregulation. luA novel, recurrent *TOP2A* mutation, which was found exclusively in four GBMs, results in the TOP2A E948Q variant with altered DNA binding and relaxation activities. The deleterious *TOP2A* mutation resulting in transcription deregulation in GBMs may contribute to disease pathology.

KEYWORDS
genomic landscape, gliomas, targeted NGS, topoisomerase 2A, transcriptome

What's new?

The prognosis of patients with glioblastoma (GBM) is poor and has not improved despite advances in treatment. Here, the authors identified genetic variants and transcription abnormalities in 135 GBM samples. They describe a novel variant of the *TOP2A* gene, present in four cases, that correlates with shorter survival. The mutated TOP2A protein showed stronger DNA binding and relaxation activity than the wild type protein. This mutation may contribute to the progression of the disease, and could be a viable target for new therapies.

1 | INTRODUCTION

High-grade gliomas (HGGs) are most frequent, primary malignant brain tumors in adults,[1] classified by the World Health Organization (WHO) as WHO grade 3 and 4 gliomas. The most aggressive is glioblastoma (GBM *IDH*-wild type), a recurrent and deadly tumor characterized by diffusive growth, numerous oncogenic alterations and therapy resistance. Overall survival of HGG patients ranges from 12 to 18 months from a time of diagnosis, despite aggressive treatments encompassing gross surgical resection, radiotherapy and chemotherapy.[2] Extensive, multiplatform genomic, epigenetic and proteomic characterization of HGGs by The Cancer Genome Atlas (TCGA) consortium[3] revealed many insights into molecular subtypes, genomic and epigenomic alterations along with potential markers and therapeutic cues slightly improving diagnostics and personalized therapy of these tumors. Recurrent somatic alterations in genes such as *TP53*, *PTEN*, *NF1*, *ATRX*, *EGFR*, *PDGFRA* and *IDH1/2* have been reported.[4,5] Detecting the mutation in the *IDH1* gene[6] resulting in distinct transcriptomic and DNA methylation profiles in *IDH*-mutant and *IDH*-wild type gliomas[7] improved diagnosis of HGGs.[8] Despite advancements in understanding gliomagenesis, most molecularly targeted therapies for GBMs failed in phase III trials. Thus, even rare but targetable genetic changes are of interest offering a personalized treatment for selected patients. For example, the *BRAF V600E* mutation occurring in 1% of GBM patients, is a valid therapeutic target in basket trials, including gliomas.[9] Aurora kinases (AURK) that are highly upregulated in proliferating cells in GBMs, could be targeted with specific inhibitors having synergistic or sensitizing effects with chemotherapy, radiotherapy or other targeted drugs in GBMs.[10,11]

Therefore, we applied targeted next generation sequencing to find somatic and germline variants in 180 gliomas of different WHO grades. We designed a unique panel of 664 cancer- and epigenetics-related genes intended to detect diagnostic, prognostic and therapeutic markers in gliomas. In this article we focus on 135 GBM (GBM *IDH*-wild type) samples. We found known alterations in *TP53*, *EGFR* genes and several new variants, predicted to be potential drivers. In four GBMs 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 is implicated in the maintenance of a high order DNA structure and other processes including chromatin folding and gene expression.[12] Using computational methods, we predicted negative consequences of the E948Q substitution on TOP2A. Biochemical assays showed altered functions of the recombinant variant proteins. Our findings show that the E948Q substitution in the TOP2A affects its DNA binding and relaxation activity contributing to gliomagenesis.

2 | MATERIALS AND METHODS

2.1 | 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 Gdansk, 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. In the analysis we used 205 glioma samples of WHO II, III and IV grades. We performed targeted DNA sequencing of 180 tumor samples (135 GBM *IDH*-wild type samples); in case of 67 samples, DNA isolated from blood was sequenced as a reference to detect somatic mutations. For the samples provided by the Canadian Human Tissue Bank (GL162-GL182) clinical data other than diagnosis were not available. Detailed description of the patient cohort is presented in the Table S2. A GBM acronym refers to an *IDH*-wild type GBM. Data collected in a frame of this study for all 205 samples was deposited in EGA repository (data availability section), but all further analysis concentrated on 135 GBM *IDH*-wild type samples.

2.2 | DNA/RNA isolation

Total DNA and RNA were extracted from fresh frozen tissue samples using Trizol Reagent (Thermo Fischer Scientific, Waltham, MA), 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 with Nanodrop (Thermo Fisher Scientific, Waltham, MA) and Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA).

2.3 | Targeted custom panel DNA sequencing and RNA sequencing

We used SeqCap EZ Custom Enrichment Kit (Roche, Basel, Switzerland, no 06266339001), which is an exome enrichment design that targets the latest genomic annotation GRCh38/hg38. The panel covers the exomic regions of 664 cancer- and epigenetics-related genes (the Supplementary Gene List Data S4). Detailed description of the gene panel and protocol has been previously published by Wojtas et al.[13]

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. The procedure started with fragmentation of the material with Covaris M220 Focused-ultrasonicator (Covaris Inc., Woburn, MA) for 240 seconds to obtain 180 to 220 bp DNA fragments. Next, ends of these fragments were repaired and A base was added to 3′ end. Then indexed adapters (SeqCap Adapter Kit A and B, Roche, Basel, Switzerland) were ligated (Ligase Enzyme from KAPA LTP Library Prep Kit, KAPA Biosystems, Woburn, MA) and libraries were amplified (7 PCR cycles, KAPA HiFi Hotstart Ready Mix, KAPA Biosystems, Woburn, MA). Obtained libraries were mixed eximolar with oligonucleotide probes (NimbleGen SeqCap EZ Library SR) recognizing the assigned genomic regions. After 72 hours incubation, the pooled libraries were purified using streptavidin beads (Dynabeads M270 Streptavidin, Life Technologies, Carlsbad, CA) and amplified (13 PCR cycles, KAPA HiFi Hotstart Ready Mix, KAPA Biosystems, Woburn, MA). Quality of libraries was evaluated using Agilent Bioanalyzer and High Sensitivity DNA chIPs (Agilent Technologies, Palo Alto, CA). Quantification was estimated using Quantus Fluorometer and QuantiFluor Double Stranded DNA kit (Promega, Madison, WI). Libraries were run in the rapid run flow cell and were paired-end sequenced (2 × 76 bp) on HiSeq 1500 (Illumina, San Diego, CA).

The mRNA sequencing libraries were prepared using KAPA Stranded mRNAseq Kit (Roche, Basel, Switzerland, cat no 07962207001) according to manufacturer's protocol and as previously described.[14]

Confirmation of the *TOP2A* mutation at the chromosomal position 17:40 400 367 was performed by amplification of the fragment using primers designed in our laboratory (sequences listed in the Table S1) with Taq polymerase using Eppendorf Mastercycler Nexus. The amplified fragments were subjected to Sanger sequencing at the Oligo.pl or ultra-deep sequencing in house.

The sequencing coverage and quality statistics for each sample are summarized in DNAseq QC and RNAseq QC supplementary tables in Data S3.

2.4 | Data analysis

2.4.1 | Somatic and rare germline variant pipeline

Sequencing reads were filtered using the Trimmomatic program.[15] Filtered and trimmed reads were mapped to the human genome version hg38 by the BWA aligner[16] (http://bio-bwa.sourceforge.net/). 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 only those with a variant allele frequency (VAF) higher than 20%. A separate analysis of each sample was performed with a Scalpel software[17] to enrich for discovery of short deletions and insertions. Variants obtained with a Scalpel were filtered to include only variants with VAF >20%. Variants obtained from bcftools and a Scalpel analysis were merged, and processed to maf format using vcf2maf program. Vcf2maf annotates genetic variants to existing databases by Variant Effect Predictor (VEP) program.[18] Subsequently, a 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 subsequently analyzed in maftools R library.[19] Final data contained putative somatic and rare germline variants.

2.5 | Somatic variants pipeline

FASTQ files from both tumor and blood DNA samples were mapped to the human genome (hg38) using NextGenMap aligner (http://cibiv.github.io/NextGenMap/). Read duplicates were marked and removed by Picard (https://broadinstitute.github.io/picard/). 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 removed. Only coding variants with damaging predicted SIFT values (>0.05) were selected. ProcessSomatic method from VarScan2[20] 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 versions of databases (refGene, clinvar, cosmic, avsnp150 and dbnsfp30a). Finally, maftools R library[19] was used to analyze the resulting somatic variants.

To test overlap of the results obtained with a somatic variant pipeline (including blood DNA as a reference) and somatic and rare germ-line variants pipelines, we compared 50 samples that were common in two analyses. We compared a number of Missense Mutations, Frame Shift Del, Frame Shift Ins and Nonsense mutations obtained with two pipelines.

2.6 | RNA sequencing analysis

RNA sequencing reads were aligned to the human genome by gap-aware algorithm—STAR and gene counts were obtained by feature counts.[21] Aligned reads (bam format) were processed by RseQC[22] 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,[23] DESeq2[24] and clusterProfiler[25] libraries.

2.7 | Computational prediction of the *TOP2A* mutant effect

To predict the consequences of: L916F, T932I, E948Q substitutions in the TOP2A, we used the crystal structure of the human TOP2A in a complex with DNA (PDB code: 4FM9). Substitutions in TOP2A were introduced using UCSF CHIMERA.[26] Free energy differences of DNA binding of TOP2A variants: L916F, T932I, E948Q in comparison to WT were calculated with HDOCK.[27] Protein structures were visualized and an electrostatic potential of TOP2A WT and E948Q variant were calculated using PyMOL.

2.8 | 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 re-cloned into a bacterial expression vector. The *TOP2A* fragment corresponding to 890 to 996 amino acid (aa) residues was amplified and cloned into pET28a vector as a BamHI—EcoRI fragment, resulting in a construct expressing His6tag fused to N-terminus of TOP2A 890 to 996 aa. The *TOP2A* fragment corresponding to residues 431 to 1193 aa was amplified and cloned into a pET28a vector resulting in a construct expressing His6tag fused to N-terminus of TOP2A 431 to 1193 aa. Site-directed mutagenesis of the *TOP2A* gene was performed using a PCR-based technique employing primers with mutated nucleotides (sequences listed in Table S1). All resulting constructs were sequenced using Sanger sequencing.

2.9 | Protein expression and purification

Proteins were expressed from the constructs carrying the *TOP2A* in *Escherichia coli* Rosetta (DE3) strain. Bacteria were grown in LB media with antibiotics at 37°C and when they reached OD_{600} 0.8, protein expression was induced with 1 mM IPTG at 37°C for 4 hours. Bacteria were harvested by centrifugation (4000 *g* for 5 minutes, 4°C). 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 processed using Bioruptor Plus sonicator (Diagenode). Proteins were purified using HisPur Ni-NTA Magnetic beads (Thermo Scientific) at 4°C. The beads were rinsed with wash buffer 1 (with 2 M NaCl) and wash buffer 2 (with 20 mM imidazole). The proteins were eluted with elution buffer containing 250 mM imidazole, pH 8.0. Next, the buffer was exchanged to phosphate-buffered saline with 10% (v/v) glycerol, 10 mM BME using G25 Sephadex resin (Thermo Scientific) and

Spin-X Centrifuge Tube (Costar). Final TOP2A 890 to 996 aa samples were analyzed using 15% SDS-PAGE and TOP2A 431 to 1193 aa samples in 6% SDS-PAGE. Homogeneity of the proteins was ~90%. Protein concentrations were estimated by measuring the absorbance at 280 nm.

2.10 | Electrophoretic mobility shift assay

DNA binding assay was performed using the LightShift Chemiluminescent electrophoretic mobility shift assay (EMSA) Kit (Thermo Scientific Cat no 20148) according to manufacturer's instructions. Binding reactions contained 60 pM biotin-labeled oligonucleotide 60 bp end-labeled duplex from LightShift Chemiluminescent EMSA Kit and 4.25 μM TOP2A 890 to 996 aa WT or E948Q or 0.1 μM TOP2A 431 to 1193 aa WT or E948Q proteins, in a 10 mM Tris pH 7.5 buffer with 50 mM KCl, 1 mM DTT, 5 mM $MgCl_2$ in 30 μL. The reaction mixtures were incubated for 30 minutes 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 no 77016) in a Tris-borate-EDTA buffer and detected by chemiluminescence using a Chemidoc camera (Bio-Rad).

2.11 | Supercoil relaxation assay

The supercoil relaxation reaction was conducted with 200 ng supercoiled (SC) pET28a plasmid and 0.1 or 0.2 μM WT or E948Q TOP2A variants in a buffer containing 20 mM Tris pH 8.0, 100 mM KCl, 0.1 mM EDTA, 10 mM $MgCl_2$, 1 mM adenosine triphosphate (ATP), 30 μg/mL bovine serum albumin (BSA), at 37°C for 30 minutes. Fragments corresponded to 890 to 996 and 431 to 1193 aa of the TOP2A. The resultant reaction mixtures were resolved by electrophoresis on 0.7% agarose gel followed by staining with a SimplySafe dye (EURX Cat no E4600-01) and visualization using a Chemidoc camera (Bio-Rad).

3 | RESULTS

3.1 | Identification of somatic and rare germline variants in gliomas

The study was performed on 205 glioma specimens: 184 were from the Polish glioma cohort and 21 were from the Canadian Brain Tumor Bank. Majority of tumors (170) were HGGs (WHO grades 3 and 4) and 69% (142/205 in total, 135/205 had DNA analysis done) were *IDH*-wild type GBMs. The clinical characteristics of patients are summarized in the Table S2 and Figure S1. We performed targeted DNA next generation sequencing and searched for somatic and rare germline variants in the dataset. Genetic variant databases have significantly improved in the recent years, which allows to identify variants occurring with a 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), these variants were considered likely pathogenic.[28] OncodriveCLUST spatial clustering[29] was used to find potential oncogenic drivers.

We focused our analyses on *IDH*-wild type GBMs (135 samples profiled by DNA sequencing) and identified top scoring hotspots, among a set of mutated genes including *FOXO3*, *PDE4DIP*, *ABL1* and *RECQL4* that were not previously reported as mutated in HGGs (Figure 1A,B). Interestingly, we found new mutations in the *TOP2A* gene in several GBMs. The expression of *TOP2A* was higher in HGGs, particularly in GBMs in the TCGA dataset (Figure 1C). In total, the *TOP2A* gene harbored seven mutations but the same variant in the same chromosomal position was found in four GBMs. The mutation was in the fragment encoding the TOWER domain of the TOP2A (Figure 1D). The presented lolliplot (Figure 1D) shows all variants found in the *TOP2A* gene, including the ones that did not pass criterion of variant allele frequency (VAF) higher than 20% (Table S3).

To verify concordance of the somatic and somatic/rare germline pipeline of the variant analysis, we computed overlap of these results. Most of somatic variants were also detected by a somatic/rare germline pipeline (Table S4).

3.2 | Novel mutations identified in the *TOP2A* gene and their pathological consequences

We found seven SNP in the *TOP2A* gene in nine samples of WHO grade 4 tumors (seven/nine were *IDH*-wild type GBMs from the Polish Glioma Cohort and two/nine in grade 4 astrocytoma; Table S3). We concentrated our analysis on GBM samples, therefore, astrocytoma samples were not taken for analysis and will not be discussed further. All identified SNPs give rise to missense variants. Transversion of guanine to cytosine at the chromosomal position 17:40 400 367, resulting in the substitution of E948 to Q, was found in four GBMs (AF = 4/135 GBMs = 0.03). The presence of the recurrent SNP was verified in three samples (GL3, GL135, GL151) using Sanger sequencing and in one (GL4) using ultra-deep sequencing. One of the detected variants was identified as a somatic variant in comparison to blood DNA, for other three patients no blood or archival sections were available to confirm or reject a somatic status of the alteration. In the sample, where reference blood was available, the mutation penetration was low ~13% suggesting that this is a somatic variant, in three other samples the mutation penetration was near 50% suggesting the heterozygous germline variant. These results combined with a very low frequency of this variant in the gnomAD (AF < 0.00001) suggest that this is a case of a rarely occurring, germline mutation that is enriched in tested GBM patients and at least in one of the patients occurred as a somatic mutation (Table S3).

The analysis of the newly released data from the "Thousands Polish Genome Project"[30] revealed that the E948Q variant was present

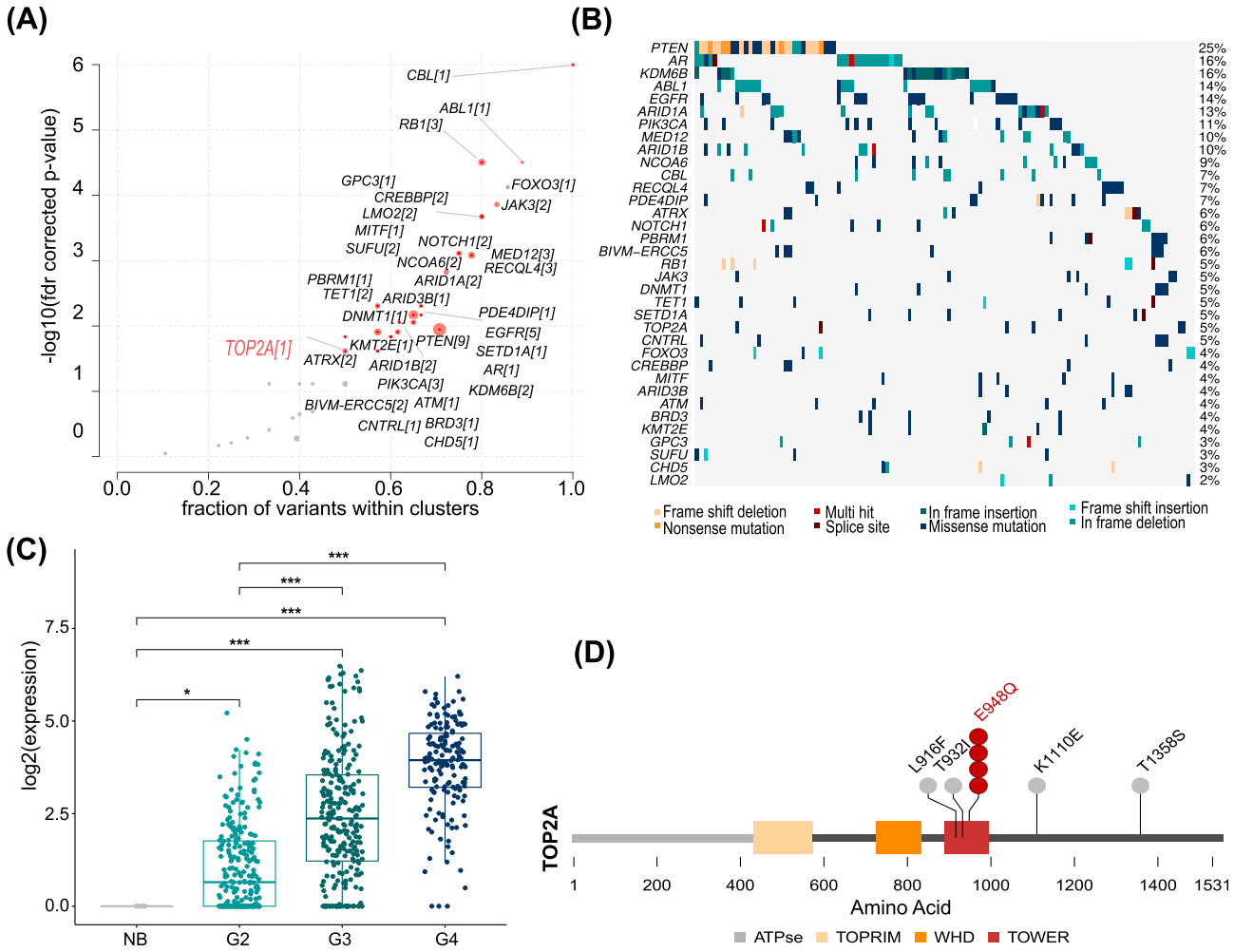

FIGURE 1 Targeted sequencing identified frequent mutations in *IDH*-wild type glioblastoma and revealed novel mutations in the *TOP2A* gene. (A) Plot showing potential oncogenic drivers (OncodriveCLUST function) in the cohort of *IDH*-wild type glioblastoma, red dots represent potential drivers, with FDR-corrected OncodriveCLUST *P* value <.05; size of the dot corresponds to a number of the alterations that were present in the cluster. (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 (G2, G3 and G4) in the TCGA dataset. *P*-values based on Wilcoxon signed-rank test were computed (**P* < .05; ***P* < .01 and ****P* < .001). (D) TOP2A lolliplot showing the identified substitutions mapped on the protein structure of the TOP2A. A number of dots represents the number of patients in which the particular alteration was found.

with AF = 0.000574, which is very low, although more frequent than in the gnomAD database. CADD algorithm prediction assigned the E948Q variant together with most of other TOP2A variants to a group of 1% most deleterious human variants—score >20 (Table S3). In the TCGA dataset *TOP2A* alterations resulting in changes in the position E948 were detected in three tumors, whereas the T215P residue was the most frequently substituted (seven non-CNS tumors; Figure S2A). Moreover, the *TOP2A* gene is affected by amplifications or missense mutations in most TCGA PanCancer cohorts (Figure S2B). To verify whether E948Q occurs as a confirmed somatic variant in other databases, we have interrogated the cBIO portal and found E948Q in one patient (Sample ID: XRN-GLE3RD)[31] (Figure S3). Finally, we verified if the E948Q variant was reported in the GLASS Consortium database,[32] but no alterations resulting in the somatic E949Q variant were detected. These results suggest that E948Q is a very rare germline variant enriched in the Polish cohort of GBM patients.

We studied exclusivity and co-occurrence of *TOP2A* mutations with frequently detected alterations. *TOP2A* mutations co-occurred with alterations in *ARID1A*, encoding a SWI/SNF family helicase and *KMT2C* (formerly *MLL3*), coding for histone methyltransferase, as well as with *EGFR* and *ABL1* mutations (Figure 2A). Alterations in *ARID1A*, *KMT2C* and other genes from the Figure 2A were also significantly enriched in *TOP2A* mutated samples in the PanCancer TCGA dataset (Figure S2C). Samples with *TOP2A* mutations have numerous genetic alterations when compared to other well-defined clusters characterized in *IDH*-wild type GBM, such as *EGFR*-, *NF1*-, *TP53*- and *PTEN*-mutated clusters shown in Figure S4. Most

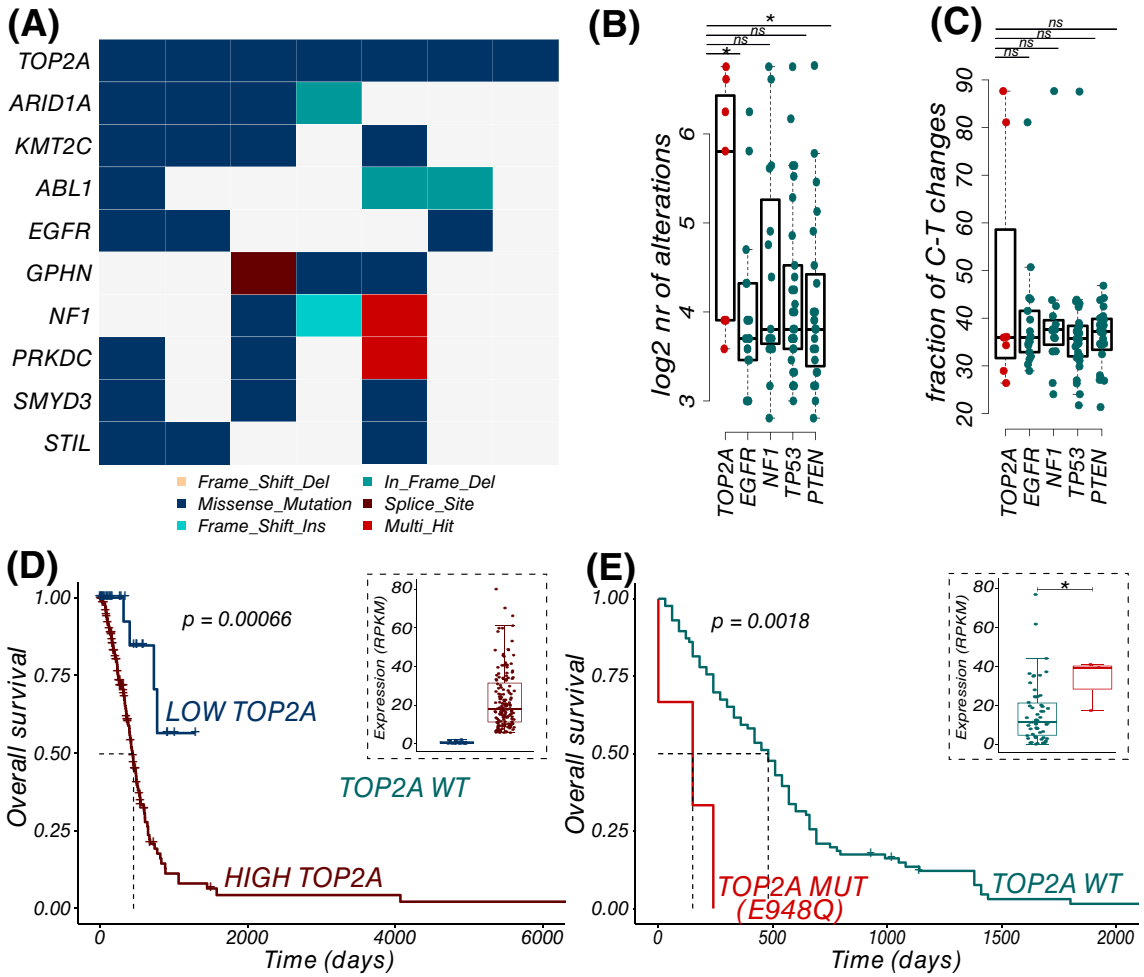

FIGURE 2 Co-occurrence of genomic alterations in *TOP2A* mutated samples and the effects of mutations or *TOP2A* expression on survival. (A) Oncoplot showing most frequently altered genes in all *TOP2A* mutated *IDH*-wildtype GBMs (7 patients), each column corresponds to a glioma sample, while each row corresponds to one altered gene. Log2 number of genetic alterations in the *TOP2A* (B) and a fraction of C to T nucleotide changes (C) in the *TOP2A* mutated samples vs the samples with other mutations: *EGFR*, *NF1*, *TP53* and *PTEN*. Sample assignment to each of described groups is explained in the Figure S5. (D) Kaplan-Meier overall survival curve for GBM patients with high (> *TOP2A* mRNA median expression × 1.25) or low (< *TOP2A* mRNA median expression × 0.5) *TOP2A* expression in the TCGA dataset. Patients alive at the time of the analysis were censored at the time they were followed up. The median overall survival was ~435 days in the HIGH *TOP2A* mRNA expression group (n = 164) and >6000 days in the LOW *TOP2A* mRNA expression group (n = 29). *P*-values based on the Log Rank Test were calculated. The boxplot shows the TOP2A mRNA expression for these patients in the TCGA. (E) Kaplan-Meier overall survival curve in this cohort. The median overall survival was ~150 days in GBM patients with the mutated *TOP2A* (TOP2A MUT E948Q, 3 patients) and ~480 days in the patients with wild-type *TOP2A* (TOP2A WT, 86 patients). *P*-values based on the Log Rank Test were calculated. The *TOP2A* mRNA expression and its relationship to the TOP2A status (TOP2A MUT E948Q n = 3; TOP2A WT n = 58) are shown in the associated boxplot (all IDH1 WT patients).

samples with TOP2A mutations show a higher number of genetic alterations when compared to other defined clusters in *IDH*-wild type GBMs (Figure 2B). Most frequent changes in the *TOP2A* altered samples are C > T transitions, most likely related to deamination of 5-methylcysteine (Figure 2C). There is a high variability within GBMs harboring *TOP2A* mutations with approximately a half of *TOP2A* samples having a very high fraction (>80%) of C > T changes, while the others have an expected fraction (<40%) of C > T changes (Figure 2C). Four GBM samples with TOP2A mutations have a high mutation rate (log nr of alterations >5, Figure 2B) and they harbor *MSH26* or *MLH1* mutations (data not shown). Out of these four samples two have a mutation, which results in E948Q substitution. Out of six samples with *TOP2A* alterations, two were in GBMs with a normal range of mutational burden (Figure 2A).

To determine if *TOP2A* expression levels or mutations may affect patient's survival, RNA sequencing (RNA-seq) data from the HGGs TCGA dataset were explored. We demonstrate that patients with the high *TOP2A* levels had shorter survival than other *IDH*-wild type GBM patients ($P = 0.00066$) (Figure 2D). Three GBM patients with TOP2A E984Q variant (one patient died after surgery) from the Polish Glioma Cohort survived 3, 5 and 8 months from the time of diagnosis,

respectively. Although, the number of inspected patients was low, the data indicate that the altered *TOP2A* GBM patients had a shorter overall survival compared to other GBM patients (Figure 2E).

3.3 | The TOP2A E948Q substitution may affect protein-DNA interactions

Consequences of the TOP2A E948Q substitution were predicted using the crystal structure of the human TOP2A in complex with DNA (PDB code: 4FM9)[33] (Figure 3A,B). The E948 aa is located in the TOWER domain of the TOP2A (Figure 3C), which with other domains (TOPRIM and WHD) forms the DNA-Gate involved in DNA binding[34] and is in close proximity to DNA (Figure 3B). The replacement of a negatively charged E residue to a polar Q residue changes the electrostatic potential of the protein (Figure 3C).

We calculated the free binding energy of DNA for the E948Q TOP2A in comparison to WT (ΔG −13.37 kcal/mol; Table S5). The analysis predicts that such change would affect protein-DNA interactions, and the TOP2A E948Q variant may have an increased affinity toward DNA. Similar analysis was performed for other discovered TOP2A variants: L916F and T932I. Variants K1110E and T1358S

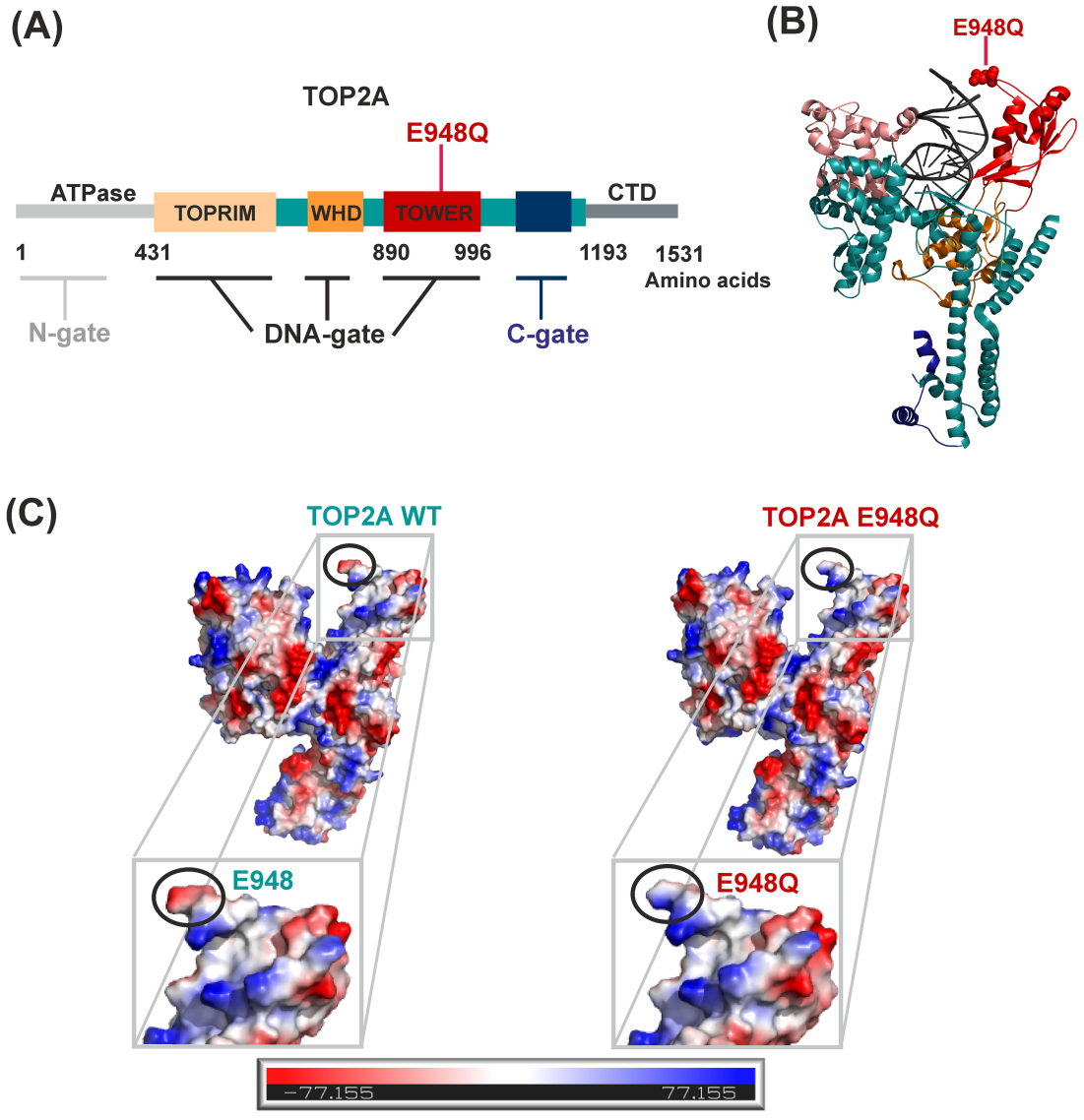

FIGURE 3 Substitution of E948Q to Q changes the electrostatic potential of the TOP2A. (A) Schematic representation of a TOP2A domain structure. Functional regions are colored and labeled: an ATPase domain, a TOPRIM—Mg^{2+} ion-binding topoisomerase/primase fold, a WHD—winged helix domain, a TOWER, CTD—C-terminal domain. The three dimerization regions are marked: N-gate, DNA-gate and C-gate. (B) Secondary structure representation of the TOP2A 431 to 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 the TOP2A variant is represented as spheres on side chains. (C) Electrostatic surface potential of TOP2A 431 to 1193 WT and E948Q variant. Electrostatic potential was calculated using the PyMOL program.

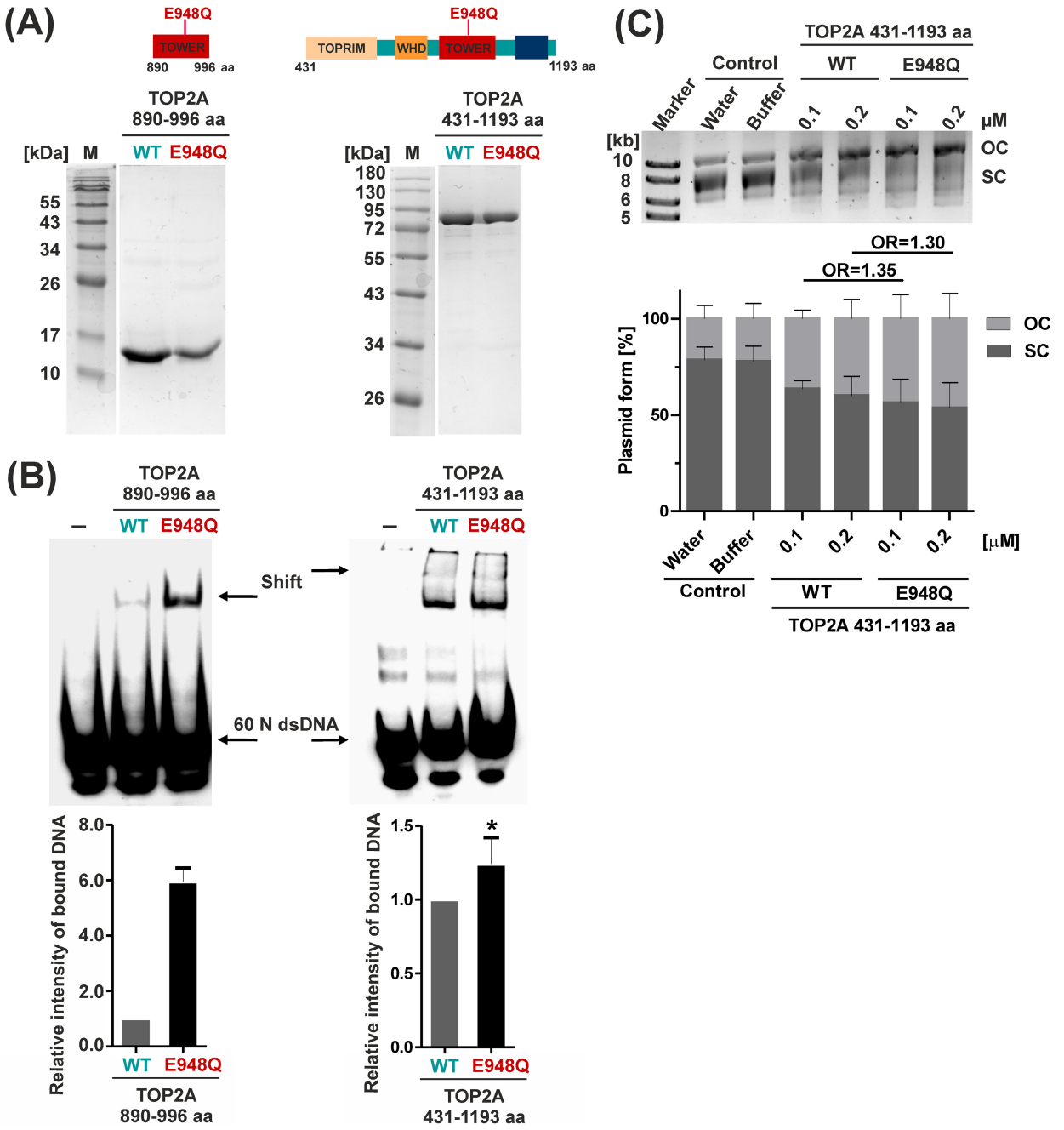

FIGURE 4 TOP2A E948Q protein exhibits a higher DNA binding capacity than WT. (A)Schematic representation of purified TOP2A proteins: TOP2A 890 to 996 aa (only the TOWER domain) and TOP2A 431 to 1193 aa (the DNA-gate). Amino acid residue E948, which is substituted to Q in the TOP2A, is labeled in the scheme. SDS-PAGE of TOP2A purified proteins. Lower panel shows purified proteins 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 ± SD (error bars) that were calculated from two independent measurements for the TOP2A 890–996 aa and five independent measurements for TOP2A 431–1193 aa (two preparations of the recombinant protein). Nonspecific binding was determined in a reaction without a protein. Statistical analysis was performed using Graphpad Prism software. Statistical significance was calculated from logarithmic raw data using *t*-test. (C) Supercoil relaxation assay of a plasmid DNA induced by TOP2A 431 to 1193 WT or E948Q proteins. The reaction mixtures were resolved by electrophoresis on 0.7% agarose gel followed by staining with a SimplySafe dye and visualization using a Chemidoc camera. The protein activities were measured by their ability to unwind a supercoiled plasmid DNA (SC) to an open circular DNA (OC). The data is presented as mean ± SD (error bars) that were calculated from three independent measurements. The signal intensities of SC and OC plasmids were measured and the fractions of both forms were calculated. The fractions (F) of OC were transformed using the logit function $\left(\text{logit} = \ln\frac{F}{1-F}\right)$. Next, the difference between the means of logit for TOP2A E948Q variant and WT in two concentrations: 0.1 and 0.2 µM were calculated (*x*). The odds ratio (OR) was determined using the function $\text{OR} = e^{x}$. The 95% confidence interval (95% CI) were calculated using the GraphPad software. [Color figure can be viewed at wileyonlinelibrary.com]

were excluded from the analysis because these residues are not present in the crystal structure. Residues L916 and T932 in the crystal structure of the TOP2A are not in the vicinity of DNA, therefore, the substitution of these residues should not influence DNA binding (Table S5).

3.4 | The TOP2A E948Q has increased binding to DNA and enzymatic activity

To analyze the effect of the TOP2A E948Q substitution on enzymatic activity, we produced and purified recombinant human WT and variant TOP2A proteins as His6 tag-fusion proteins (Figure 4A). We evaluated DNA binding and DNA relaxing activities of WT and E948Q TOP2A purified proteins. We tested binding of the TOP2A 890 to 996 aa (comprising only the TOWER domain) and the TOP2A 431 to 1193 aa (the entire DNA-Gate domain) fragments using an electrophoretic mobility shift assay (EMSA). The representative EMSA gel shows that E948Q TOP2A variant binds to DNA more strongly (Figure 4B). Densitometric assessment of shifted bands for TOP2A 890 to 996 aa (n = 2) and TOP2A 431 to 1193 aa (n = 5) confirmed statistically significant stronger binding of the longer E948Q TOP2A variant to DNA in comparison to the WT TOP2A.

The catalytic activity of the proteins was determined as a supercoil DNA relaxation activity of WT and TOP2A E948Q proteins with the plasmid DNA. The disappearance of the lower band representing a supercoiled plasmid DNA (SC bands), and the appearance of the higher band representing the untwisted plasmid DNA (OC) represent the TOP2A DNA relaxation activity and show differences in the enzymatic activity of TOP2A proteins. The shifted bands were evaluated by densitometry. While the results are not significant, odds ratio for 0.1 μM proteins in relaxation assay is OR = 1.35 and for 0.2 μM OR = 1.3 (Figure 4C). These results indicate that the E948Q TOP2A variant is functional and exhibits higher activity.

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

Transcriptomic data from 77 *IDH*-wild type GBM were analyzed to compare transcriptional patterns of the WT and *TOP2A* mutated

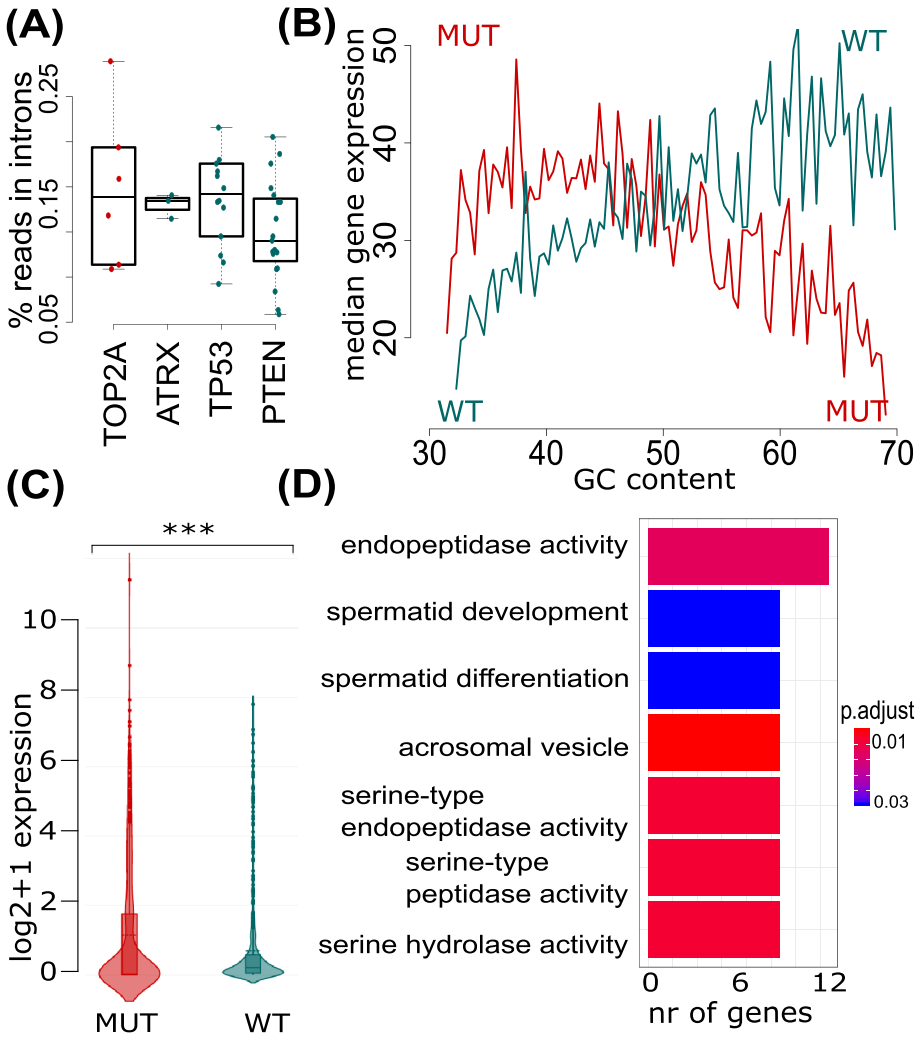

FIGURE 5 Transcriptomic alterations in *TOP2A* mutated GBMs. (A) Fractions of RNA-seq reads mapped to intron regions. (B) Profiles of the GC content in expressed mRNAs in *TOP2* mutant (MUT) and wild type (WT) GBMs. (C) Differences in expression of the genes that are not expressed in the normal brain in the *TOP2A* MUT and WT gliomas. Genes were preselected by filtering out all the genes that were expressed in any of the cells from the Stanford Human Brain RNA-seq dataset 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 GBMs. *P*-values for plots based on Wilcoxon signed-rank test, for the panel C on Kolmogorov-Smirnov test (**P* < .05; ***P* < .01 and ****P* < .001). [Color figure can be viewed at wileyonlinelibrary.com]

gliomas. Transcriptional patterns of seven GBMs with *TOP2A* mutations were compared to GBMs harboring the mutations in the *ATRX*, *TP53* or *PTEN* genes.

We found that higher percentages of the detected sequencing reads are mapped in noncoding (intron) regions in the *TOP2A* altered GBMs when compared to the *ATRX*, *TP53* or *PTEN* mutated GBMs (Figure 5A). This finding suggests a higher level of unspliced transcripts in samples with the mutated *TOP2A*; however, the difference was not significant due to small sample size and high variations. Percentages of G-C pairs in the coding sequences of the expressed genes in *TOP2A* altered gliomas vs other *IDH*-wild type GBMs were calculated (Figure 5B). The genes with the coding sequences rich in A-T pairs were more often expressed in the *TOP2A* altered GBMs, while the genes with coding sequences rich in G-C pairs were less frequently present in these samples (Figure 5B).

Furthermore, we compared the expression of a set of genes not expressed in the normal brain in the WT and mutated *TOP2A* GBMs. RNA-seq data were filtered to remove all brain-specific 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* altered samples than in *TOP2A* WT GBM samples (Figure 5C). Gene Ontology (GO) analyses of differentially expressed non-brain specific genes showed over-representation of pathways involved in an endopeptidase activity and testis development, suggesting acquisition of new functionalities related to aberrant differentiation and migratory processes (Figure 5D). These newly acquired functions of the altered *TOP2A* may contribute to a shorter survival of the patients with the *TOP2A* mutation. Lastly, RNAseq analysis without pre-filtering for brain-specific genes revealed four pathways up-regulated in *TOP2A* mutated samples: Herpes simplex virus 1 infection, Ubiquitin-mediated proteolysis, Colorectal cancer and Homologous recombination (Figure S5A). Focusing on Homologous recombination, we found many classical repair genes such as *BRCA1*, *ATM* or *RAD* genes up-regulated in samples with the *TOP2A* mutation (Figure S5B).

4 | DISCUSSION

The targeted NGS-based genomic profiling of 664 cancer-and epigenetics-related genes performed in the present study revealed genomic alterations occurring in HGGs at high sensitivity and specificity. We identified a number of well-known and frequent alterations in *TP53*, *EGFR*, *RB1* and *PTEN* genes previously detected in GBMs, which authenticates our approach and the results. However, using an OncodriveCLUST, a new nucleotide sequence-based clustering algorithm to detect cancer drivers in genomic regions,[29] we identified several novel genomic variants that represent potentially deleterious germline, somatic mosaicism or loss-of-heterozygosity variants. Mutations in genes such *FOXO3*, *PDE4DIP*, *ABL*, *TOP2A* and *RECQL4* were not previously reported in HGGs.

Among top 50 genes most frequently mutated in this cohort of particular interest are genetic alterations in the genes coding for epigenetic enzymes and modifiers: *KDM6B*, *ARID1A/B*, *KMT2D* and *NCOA6*. KDM6B is critical for maintenance of glioma stem cells that upregulate primitive developmental programs.[35] KDM6B has been targeted with a specific inhibitor GSK-J4 in acute myeloid leukemia,[36] and GSK-J4 was active in parental and temozolomide-resistant glioblastoma cells.[37] NCOA6 can associate with three co-activator complexes containing CBP/p300 and Set methyltransferases, and may contribute to transcriptional activation by modulating chromatin structure through histone methylation.[35] The biological relevance of these genomic alterations in HGGs requires further investigation.

In the current study, we evaluated biological effects of a new, recurrent mutation in the *TOP2A* gene identified in four GBM patients from the Polish Glioma cohort. *TOP2A* mutations (although not mutations resulting in substitution E948Q specifically) were found in many cancers, including breast and lung cancer,[38] but not in gliomas. As this recurrent mutation in the *TOP2A* was discovered only in GBMs from the Polish Glioma cohort (but not in 21 GBMs from the Canadian Brain Tumor Bank), this variant might represent a putative founder mutation. A somatic status of the mutation in the *TOP2A* gene causing E948Q substitution was confirmed only in one patient, due to a lack of corresponding blood samples, therefore it cannot be excluded that the SNP may occur both as a germline and somatic variant, similarly to some cancer-related variants (ie, *BRCA1*). Nevertheless, the mutation causing E948Q was found more frequently in the Polish population dataset (AF = 0.000574)[39] than in the gnomAD (AF = 0.000008) population suggesting a cancer-related or predisposing variant specific to the Polish or Central European population. There is an obvious limitation of our study due to a small sample size and the fact that only four samples with the E948Q substitution were found. A larger cohort of patients, preferentially with the Central European origin, should be collected and interrogated regarding E948Q variants in the future.

Topoisomerases regulate DNA topology by introducing transient single- or double-strand DNA breaks, and may unwind DNA, which is a critical step for replication or transcription.[40,41] The E948 aa is located in the TOWER domain of the TOP2A which binds DNA. Computational analyses revealed that substitution of a negatively charged E to a polar Q changes the electrostatic potential of the protein. This prediction was confirmed experimentally when DNA binding activities of recombinant WT and E948 TOP2A proteins were compared. As TOP2A is a large protein, we tested shorter TOP2A 890–996 aa (only the TOWER domain) and longer TOP2A 431–1193 aa (the entire DNA-Gate domain) fragments of TOP2A. EMSA assays showed that the TOP2A E948Q binds DNA more strongly than WT TOP2A.

The presented evidence suggests that the TOP2A variant may acquire new functions and exhibits a gain-of-function phenotype. Global gene expression profiling showed different expression of genes with a high AT content in the altered *TOP2A* GBMs in comparison with WT *TOP2A* GBMs which indicates global transcriptional alterations. Moreover, *TOP2A* altered GBMs showed activation of genes that encode proteins with an endopeptidase activity that are implicated in cancer invasion and genes related with homologous recombination. GBM patients with the *TOP2A* alteration had shorter survival than others. The analysis of the TCGA dataset shows increased

expression of *TOP2A* in HGGs, which negatively correlates with patient's survival. Altogether, we demonstrate that the *TOP2A* alteration results in enhanced enzymatic activities, which are associated with transcription deregulation and altered transcriptomic profiles.

More analyses are required to distinguish whether the alterations in transcription are caused by *TOP2A* variants or other acquired mutation, as many *TOP2A* mutated samples are hypermutated. However, if *TOP2A* variants alter the transcription, GBM patients with altered *TOP2A* could be selectively treated. Anthracyclines, such as daunorubicin, doxorubicin, which are widely used in chemotherapy of lymphoma, breast, ovarian cancers and others[42] could be potentially used in GBM patients harboring the *TOP2A* variant. These drugs intercalate to DNA and inhibit topoisomerase 2 leading to the induction of DNA breaks and cell death. The use of anthracyclines causes severe side effects, including cardiotoxicity. However, recent attempts to increase the brain delivery of such drugs with convection-enhanced delivery or using brain-targeting liposomes would enable the use of therapeutic agents at higher concentrations without a systemic toxicity.

AUTHOR CONTRIBUTIONS

The work reported in the article has been performed by the authors, unless clearly specified in the text. **Bartlomiej Gielniewski:** Experimental part; Data interpretation; Article writing; **Katarzyna Poleszak:** Experimental part; Data interpretation; Article writing; **Adria-Jaume Roura:** Data interpretation; Data analysis; **Paulina Szadkowska:** Experimental part; **Karol Jacek:** Data analysis; **Sylwia K. Krol:** Experimental part; **Rafal Guzik:** Experimental part; **Paulina Wiechecka:** Experimental part; **Marta Maleszewska:** Experimental part; **Beata Kaza:** Experimental part; **Andrzej Marchel:** Clinical data preparation; **Tomasz Czernicki:** Clinical data preparation; **Andrzej Koziarski:** Clinical data preparation; **Grzegorz Zielinski:** Clinical data preparation; **Andrzej Styk:** Clinical data preparation; **Maciej Kawecki:** Clinical data preparation; **Cezary Szczylik:** Clinical data preparation; **Ryszard Czepko:** Clinical data preparation; **Mariusz Banach:** Clinical data preparation; **Wojciech Kaspera:** Clinical data preparation; **Wojciech Szopa:** Clinical data preparation; **Mateusz Bujko:** Clinical data preparation; **Bartosz Czapski:** Clinical data preparation; **Miroslaw Zabek:** Clinical data preparation; **Ewa Iżycka-Świeszewska:** Clinical data preparation; **Wojciech Kloc:** Clinical data preparation; **Pawel Nauman:** Clinical data preparation; **Joanna Cieslewicz:** Experimental part; **Wieslawa Grajkowska:** Clinical data preparation; **Natalia Morosini:** Data analysis; **Houtan Noushmehr:** Data analysis; **Bartosz Wojtas:** Conception and design; Experimental part; Data interpretation; Data analysis; Article writing; **Bozena Kaminska:** Conception and design; Data interpretation; Article writing; Final approval of article: All authors. Accountable for all aspects of the study: All authors.

ACKNOWLEDGEMENTS

We thank Kamil Wojnicki for help with statistical analysis of biochemical tests. We thank all the patients for the consent for the use of their biological material. 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 to 2013 is highly appreciated.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflict of interest.

DATA AVAILABILITY STATEMENT

All processed sequencing data generated in this study have been submitted to The European Genome-phenome Archive (EGA; https://ega-archive.org/) under accession number EGAS00001004556. Further information is available from the corresponding author upon request.

ETHICS STATEMENT

All procedures were approved by the responsible authorities as described in Materials and Methods. Each patient provided a written consent for the use of tumor tissues. The study has been approved by the Institutional Review Board or Ethics Committee (#8/2012 from the Institute of Psychiatry and Neurology; #73/KBL/2015 from the Cracov Regional Medical Chamber; #KB/54/2016 of the Medical University of Warsaw).

ORCID

Bartosz Wojtas https://orcid.org/0000-0002-9967-910X

REFERENCES

1. Ostrom QT, Gittleman H, Xu J, et al. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2009-2013. *Neuro Oncol*. 2016;18:v1-v75.
2. Stupp R, Mason WP, van den Bent MJ, et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med*. 2005;352:987-996.
3. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455:1061-1068.
4. Brennan CW, Verhaak RG, McKenna A, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013;155:462-477.
5. Verhaak RG, Hoadley KA, Purdom E, 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.
6. Parsons DW, Jones S, Zhang X, et al. An integrated genomic analysis of human glioblastoma multiforme. *Science*. 2008;321:1807-1812.
7. Ceccarelli M, Barthel FP, Malta TM, et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell*. 2016;164:550-563.
8. Louis DN, Perry A, Reifenberger G, et al. The 2016 World Health Organization classification of tumors of the central nervous system: a summary. *Acta Neuropathol*. 2016;131:803-820.
9. Kanemaru Y, Natsumeda M, Okada M, et al. 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*. 2019;7:119.
10. Borges KS, Castro-Gamero AM, Moreno DA, et al. Inhibition of Aurora kinases enhances chemosensitivity to temozolomide and causes radiosensitization in glioblastoma cells. *J Cancer Res Clin Oncol*. 2012;138:405-414.

11. de Almeida Magalhães T, de Sousa GR, Alencastro Veiga Cruzeiro G, Tone LG, Valera ET, Borges KS. The therapeutic potential of Aurora kinases targeting in glioblastoma: from preclinical research to translational oncology. *J Mol Med (Berl)*. 2020;98:495-512.
12. Roca J. Topoisomerase II: a fitted mechanism for the chromatin landscape. *Nucleic Acids Res*. 2009;37:721-730.
13. Wojtas B, Gielniewski B, Wojnicki K, et al. Gliosarcoma is driven by alterations in PI3K/Akt, RAS/MAPK pathways and characterized by collagen gene expression signature. *Cancers (Basel)*. 2019;11 (3):284.
14. Rajan WD, Wojtas B, Gielniewski B, Gieryng A, Zawadzka M, Kaminska B. Dissecting functional phenotypes of microglia and macrophages in the rat brain after transient cerebral ischemia. *Glia*. 2019; 67:232-245.
15. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30:2114-2120.
16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754-1760.
17. Fang H, Bergmann EA, Arora K, et al. Indel variant analysis of short-read sequencing data with scalpel. *Nat Protoc*. 2016;11:2529-2548.
18. McLaren W, Gil L, Hunt SE, et al. The Ensembl variant effect predictor. *Genome Biol*. 2016;17:122.
19. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28:1747-1756.
20. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568-576.
21. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923-930.
22. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics*. 2012;28:2184-2185.
23. Tarazona S, Furió-Tarí P, Turrà D, et al. Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res*. 2015;43:e140.
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014; 15:550.
25. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics*. 2012;16: 284-287.
26. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera? A visualization system for exploratory research and analysis. *J Comput Chem*. 2004;25:1605-1612.
27. Yan Y, Zhang D, Zhou P, Li B, Huang SY. HDOCK: a web server for protein-protein and protein-DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res*. 2017;45:W365-W373.
28. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. *Genome Biol*. 2017;18:77.
29. Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013;29:2238-2244.
30. Kaja E, Lejman A, Sielski D, et al. The Thousand Polish Genomes –A Database of Polish Variant Allele Frequencies. *Int J Mol Sci*. 2022;23(9): 4532.
31. Ho AS, Ochoa A, Jayakumaran G, et al. Genetic hallmarks of recurrent/metastatic adenoid cystic carcinoma. *J Clin Invest*. 2019; 129:4276-4289.
32. Barthel FP, Johnson KC, Varn FS, et al. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature*. 2019;576:112-120.
33. Wendorff TJ, Schmidt BH, Heslop P, Austin CA, Berger JM. The structure of DNA-bound human topoisomerase II alpha: conformational mechanisms for coordinating inter-subunit interactions with DNA cleavage. *J Mol Biol*. 2012;424:109-124.
34. Bax BD, Murshudov G, Maxwell A, Germe T. DNA topoisomerase inhibitors: trapping a DNA-cleaving machine in motion. *J Mol Biol*. 2019;431:3427-3449.
35. Mahajan MA, Samuels HH. Nuclear receptor coactivator/coregulator NCoA6(NRC) is a pleiotropic coregulator involved in transcription, cell survival, growth and development. *Nucl Recept Signal*. 2008;6:e002.
36. Liau BB, Sievers C, Donohue LK, et al. Adaptive chromatin remodeling drives glioblastoma stem cell plasticity and drug tolerance. *Cell Stem Cell*. 2017;20:233-246.
37. Romani M, Daga A, Forlani A, Pistillo MP, Banelli B. Targeting of histone demethylases KDM5A and KDM6B inhibits the proliferation of Temozolomide-resistant glioblastoma cells. *Cancers (Basel)*. 2019;11:11.
38. Chen T, Sun Y, Ji P, Kopetz S, Zhang W. Topoisomerase IIα in chromosome instability and personalized cancer therapy. *Oncogene*. 2015; 34:4019-4031.
39. Hu ZY, Liu YP, Xie LY, et al. AKAP-9 promotes colorectal cancer development by regulating Cdc42 interacting protein 4. *Biochim Biophys Acta*. 2016;1862:1172-1181.
40. Pommier Y, Sun Y, Huang SN, Nitiss JL. Roles of eukaryotic topoisomerases in transcription, replication and genomic stability. *Nat Rev Mol Cell Biol*. 2016;17:703-721.
41. Wang JC. Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol*. 2002;3:430-440.
42. da Ros M, Iorio AL, Lucchesi M, Stival A, de Martino M, Sardi I. The use of anthracyclines for therapy of CNS tumors. *Anticancer Agents Med Chem*. 2015;15:721-727.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Gielniewski B, Poleszak K, Roura A-J, et al. Targeted sequencing of cancer-related genes reveals a recurrent TOP2A variant which affects DNA binding and coincides with global transcriptional changes in glioblastoma. *Int J Cancer*. 2023;153(5):1003-1015. doi:10.1002/ijc.34631