Landscape of Driver Gene Events, Biomarkers and Druggable Targets Identified by Whole Genome Sequencing of Glioblastomas

Wesley S. van de Geer1,2, Youri Hoogstrate2,3, Kaspar Draaisma4, Pierre A. Robe4, Sander Bins1, Ron H.J. Mathijssen1, Pim French1,*, Harmen J.G. van de Werken2,5,6,*, Filip Y.F. de Vos1,*

1Dept. of Medical Oncology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands; 2Cancer Computational Biology Center, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands; 3Dept. of Neurology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands; 4Dept. of Neurosurgery, University Medical Center Utrecht, Utrecht, the Netherlands; 5Dept. of Urology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands; 6Dept. of Immunology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands; 7Dept. of Medical Oncology, Cancer Center, University Medical Center Utrecht, Utrecht, the Netherlands.

Equal contribution *

Corresponding authors
Pim French, PhD, Dept. of Neurology, Erasmus MC Cancer Institute, University Medical Center, Rotterdam, the Netherlands. Telephone: +31107044333, E-mail p.french@erasmusmc.nl;
Harmen J.G. van de Werken, PhD, Cancer Computational Biology Center & Department of Urology & Department of Immunology, Erasmus MC Cancer Institute, University Medical Center Rotterdam, the Netherlands. PO Box 2040, 3000 CA, BE355. Telephone: +31107044467, E-mail h.vandewerken@erasmusmc.nl;

© The Author(s) 2021. Published by Oxford University Press, the Society for Neuro-Oncology and the European Association of Neuro-Oncology.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (https://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
Funding
No funding was received for this work.

Conflict of Interest
The authors declare that they have no competing interests related to this study.

Authorship
PF, HvdW, FdV contributed equally to this work.
Conceptualization, FdV, HvdW, PF;
Formal analysis, WvdG;
Supervision, PF, HvdW, FdV;
Writing—original draft, WvdG, PF, HvdW, FdV;
Writing—review and editing, YH, KD, PR, SB, RM.
Abstract

Background: The survival of glioblastoma patients is poor. Median survival after diagnosis is 15 months, despite treatment involving surgical resection, radiotherapy and/or temozolomide chemotherapy. Identification of novel targets and stratification strategies of glioblastoma patients to improve patient survival is urgently needed. Whole genome sequencing (WGS) is the most comprehensive means to identify such DNA-level targets. We report a unique set of WGS samples along with comprehensive analyses of the glioblastoma genome and potential clinical impact of WGS.

Methods: Our cohort consisted of 42 glioblastoma tumor tissue and matched whole-blood samples, which were whole-genome sequenced as part of the CPCT-02 study. Somatic single-nucleotide variants, small insertions/deletions, multi-nucleotide variants, copy-number alterations (CNAs) and structural variants were analyzed. These aberrations were harnessed to investigate driver genes, enrichments in CNAs, mutational signatures, fusion genes and potential targeted therapies.

Results: Tumor mutational burden (TMB) was similar to other WGS efforts (1-342 mutations per megabase pair). Mutational analysis in low TMB samples showed that the age-related CpG demethylation signature was dominant, while hyper- and ultramutated tumors had additional defective DNA mismatch repair signatures and showed microsatellite instability in their genomes. We detected chromothripsis in 24% of our cohort, recurrently on chromosomes 1 and 12. Recurrent non-coding regions only resulted in TERT promoter variants. Finally, we found biomarkers and potentially druggable changes in all but one of our tumor samples.

Conclusions: With high quality WGS data and comprehensive methods, we identified the landscape of driver gene events and druggable targets in glioblastoma patients.

Keywords
Glioblastoma; Whole Genome Sequencing; Genomics; Precision Medicine
Key points

List two to three key points that summarize the most important findings of your manuscript.

- Mutational signatures varied greatly in low and hyper- or ultramutated tumors
- Chromothripsis was found in 24% of our patients, repeatedly on chromosomes 1 and 12
- We found biomarkers and potentially druggable targets in nearly all our tumor samples

Importance of the Study

We characterized the genomes of 42 glioblastoma samples. By utilizing whole-genome sequencing data, identification of structural variants, copy-number alterations, mutational signatures, catastrophic events and microsatellite instability is much more accurate than based on panel-based or whole exome data. We compared our results with previously published glioblastoma studies (panel-based, whole-exome sequencing (WES) and whole-genome sequencing (WGS)) and showed the heterogeneity through concordant (mutational burden) and discordant (chromothripsis, prevalence of fusion genes) findings. By linking our molecular data with current clinical practice and trials, we show researchers and neuro-oncologists the frequencies, biomarkers and potential targeted treatment options for glioblastoma patients.
Introduction

Glioblastoma is the most common and aggressive malignant primary brain tumor in adults. Patients have a median survival of less than 15 months after diagnosis, and standard treatment consists primarily of surgical resection and subsequent radiation therapy with concurrent and adjuvant temozolomide chemotherapy\(^1\)–\(^3\). The short survival is due to the aggressiveness of the tumor, its infiltrative nature and therefore cannot be resected completely, and its refractoriness to treatments, which may be due to intratumoral heterogeneity of glioblastoma\(^4\).

Research of the genomic landscape of glioblastomas has primarily focused on changes in the coding regions of genes. These studies revealed mutations in common cancer markers such as TP53 and PTEN, and glioblastoma associated hallmarks as TERT promoter, IDH1 and PDGFRA mutations, MGMT promoter methylation, EGFR amplification and mutations and homozygous deletion of the CDKN2A/CDKN2B locus\(^5\). However, as only a fraction of the entire genome is protein coding, genetic changes outside the coding genome require further investigation in relation to driving glioblastoma. By assessing a patient’s entire cancer genome, tumor mutational burden can be quantified more precisely compared to panel and WES based data, optimizing cut-offs for hypermutation and potential immunotherapy susceptibility\(^6\). Analyzing non-coding regions may provide new insights into tumor biology and may point towards new directions for diagnosis and treatments\(^7\). Moreover, detection of structural variants (SVs), copy-number alterations (CNAs), microsatellite instability (MSI) and common mutational patterns are more uniformly covered and more accurate using whole genome sequencing compared to the commonly used exome and panel sequencing technologies. Whole genome analyses can lead to the identification of novel treatment targets, which may ultimately improve on the standard treatment modalities for glioblastoma patients.

We therefore investigated whole-genome sequencing data from 42 glioblastoma tumor tissue samples that were sequenced as part of the CPCT-02 study\(^8\). Here, we show the genomic characterization of the CPCT-02 glioblastoma cohort, compare our data to other WGS datasets and discuss which genetic characteristics are potential targets for current or novel treatments. In the end, we remodel our genetic insights into clinically relevant hypotheses which could be applied in precision medicine strategies to increase survival at an individual patient level.

Methods

Center for Personalized Medicine: Patient cohort, study procedures, sample collection. Details about the study protocol, sampling and sequencing have recently been discussed by Priestley et al.\(^5\) and van Dessel et al.\(^9\). This study and its protocol were approved by the medical ethical committee (METC) of the University Medical Center Utrecht, as well as the Netherlands Cancer Institute. In accordance with the Declaration of Helsinki, patients with
glioblastoma provided their written informed consent for participation in the CPCT-02 (NCT01855477) clinical study before procedures. Core needle biopsies from the tumor lesion and peripheral blood samples were collected, along with clinical data in the same manner across Dutch hospitals.

**Hartwig Medical Foundation: Whole genome sequencing and preprocessing.** Tumor and whole-blood pairs were whole-genome sequenced on a HiSeqX system generating 2 × 150 base read pairs using standard settings (Illumina, San Diego, CA, USA) at the Hartwig Medical Foundation central sequencing center. Patients included in this study, were whole-genome sequenced between the 5th of June 2017 and the 14th of September 2018. Preprocessing was performed as described by Priestley et al.⁸, as well as tumor profiling of 32 of these samples. Briefly, read pairs were mapped to the human reference genome (GRCh37) using BWA-mem¹⁰ and subsequent variant calling was applied (see below). Next, several quality control or correction steps were performed systematically. SV calling was performed using the Genome Rearrangement IDentification Software Suite (GRIDSS)¹¹,¹². Computational ploidy estimation and copy-number (CN) analysis was performed using the custom pipeline PURPLE (PURity & PLoidy Estimator)⁸,¹², estimating tumor purity and CN profile by combining B-allele frequency (BAF), read depth, and SVs. Gene fusion events were called using LINX¹² (v1.11). MSI status was determined by assessing small insertions/deletions (InDels) per million bases in repeat regions with the tool MSISeq¹³ (MSI with a score > 4).

**Variant annotation and filtering.** Somatic variants were determined using Strelka and provided by the Hartwig Medical Foundation as part of the data request. Somatic Variant Call Format (VCF) files were annotated based on GRCh37 with HUGO gene symbols, HGVS notations, gnomAD¹⁴ frequencies and COSMIC¹⁵ (9,733,455 entries) and dbSNP¹⁶ (686,600,501 entries) identifiers using VEP¹⁷ (database release 95, merged cache) with setting --per_gene. Exclusively somatic single-nucleotide variants (SNVs), small InDels, multi-nucleotide variants (MNVs) with >= 3 alternative read observations and passing variant caller quality control were included in the analyses. Furthermore, population variants were removed to prevent germline leakage, based on the gnomAD database (v2.0.2): gnomAD exome (ALL) allele frequency ≥ 0.001; and gnomAD genome (ALL) ≥ 0.005. Variants specific for the Dutch CPCT cohort were removed based on a panel-of-normals from 1,762 representative normal blood HMF samples. The most deleterious mutation was used to annotate the overlapping gene for each sample.
Tumor Mutational Burden calculation. The number of mutations per megabase pair was calculated as the amount of somatic genome-wide SNVs, MNVs, and InDels divided by the amount of callable nucleotides in the human reference genome (GRCh37) FASTA file:

\[
TMB = \frac{(SNVs_g + MNVs_g + InDels_g)}{(2858674662 \times 10^6)}
\]

Ploidy and copy-number analysis. Broad and focal somatic CN alterations were identified by GISTIC2.0 (v2.0.23), using the following parameters: genegistic 1, gcm extreme, maxseg 4000, broad 1, brlen 0.98, conf 0.95, rx 0, cap 3, saveseg 0, armpeel 1, smallmem 0, res 0.01, ta 0.1, td 0.1, savedata 0, savegene 1, qvt 0.1. Distinction between shallow and deep CN events per region was based on thresholding performed by GISTIC2.0¹⁸. The alterations are assigned a score which takes both the amplitude and the frequency of its occurrence across samples into account (G-score). Thresholding divides into five CN categories, two for deletions (-2 = deep, possibly homozygous loss, -1 = shallow, possibly heterozygous loss), one for diploid (0 = diploid) and two for amplifications (1 = few additional copies, often broad gain, 2 = more copies, often focal gain). Annotation of GISTIC2.0 peaks was performed as follows: A) Wide peaks were annotated with all overlapping canonical UCSC genes within its limits. B) Focal peaks were annotated based on overlapping genomic coordinates, using custom R scripts and UCSC gene annotations.

Structural variant analysis. SVs were imported using custom R scripts, overlapping genes on at least one breakpoint, using GRCh37 genomic coordinates. SVs with an upstream or downstream Tumor Allele frequency (TAF) below 0.1 as determined by PURPLE and GRIDSS were discarded along with SVs covering all exons of a gene. In the case of both (multiple) mutations and/or SVs in the same gene, these were annotated as ‘multiple mutations’.

Fusion gene analysis. WGS-based TSV files were imported using R and overlapped with the three pillars of ChimerDB¹⁹; deep sequencing data (ChimerSeq), text mining of PubMed publications (ChimerPub), with extensive manual annotations (ChimerKB). Events that were not present in any pillar of ChimerDB and intra-gene fusions were filtered out. RNA-Seq based fusion genes detected with Isofox (https://github.com/hartwigmedical/hmftools/tree/master/isofox) were imported using R and overlapped with the fusion events detected in the DNA sequencing.
**Somatic Driver Genes Analysis.** We utilized the dN/dS model^{20} (192 Poisson rate parameters; under the full trinucleotide model) to identify genes undergoing mutational selection with the R package dndscv^{20} (v0.0.0.9). Both the substitution model and InDel model were used and were corrected for sequence composition, gene length and mutational signatures. These models test the ratio between nonsynonymous (missense, nonsense and essential splice site) and background (synonymous) mutations. To identify genes that drive selection, a q-value < 0.05 (both including and excluding the InDel model) was used.

**Mutational signatures analysis.** Mutational signatures analysis was performed using the MutationalPatterns R package^{21} (v1.10.0). The mutational signatures based on single base substitutions (N=90 v3 signatures), as established by Alexandrov et al.^{22}, were downloaded from COSMIC. SNVs were categorized according to their respective trinucleotide context (GRCh37) into a mutational spectrum matrix M_{ij} (where i represents 1:96 trinucleotide contexts and j represents the number of 1:42 samples) and subsequently, a constrained linear combination of the ninety mutational signatures was constructed per sample using non-negative least squares regression implemented in the R package pracma^{23} (v2.2.9).

**Catastrophic mutational processes.** Detection of chromothripsis-like events and localized hypermutation (kataegis) were performed as described in van Dessel et al.^{9}. Chromothripsis, also known as chromosomal shattering, followed by seemingly random re-ligation, was detected using Shatterseek (v0.4) with default settings. The following definition of chromothripsis was applied: (a) Total number of intrachromosomal structural variants involved in the event ≥25; (b) max. number of oscillating CN segments (two CN states) ≥7 or max. number of oscillating CN segments (three CN states) ≥14; (c) Total size of chromothripsis event ≥20 Mb; (d) Satisfying the test of equal distribution of SV types (p > 0.05); and (e) Satisfying the test of non-random SV distribution within the cluster region or chromosome (p ≤ 0.05). Kataegis was detected using a piecewise constant fit method on SNVs for each chromosome separately. The resultant segments with a mean intermutational distance of ≤2000 bp and at least five SNVs were taken into account. Samples with >200 events were ignored, as these show global hypermutation instead of local.

**Recurrent somatic variants.** All filtered somatic mutations occurring in at least four samples were selected as recurrent. Non-coding constraint mutations (classified as ‘related to key GBM genes’ and ‘select genes outside the Key GBM genes’) retrieved from Sakthikumar et al.^{24} were compared to all variants by overlapping genomic ranges with the IRanges R package^{25}.
**Actionable targets.** Iclusion (https://iclusion.com) data, which connects specific or gene-level aberrations to clinical cancer studies, while also integrating clinical interpretations from Precision Oncology Knowledge Base (OncoKB)\(^{26}\), Clinical Interpretation of Variants in Cancer (CIViC)\(^{27}\) and Cancer Genome Interpreter (CGI)\(^{28}\) was provided by Hartwig Medical foundation. These data were imported into R, attaching a label based on the actionability for each sample and gene. On-Label indicates treatment registered by federal authorities for glioblastoma, whereas Off-Label indicates a registration for other tumor types. The Off-Label data was filtered for cancer types fitting gliomas and solid tumors, excluding entries for chemotherapy (for its general applicability). All targets and biomarkers were overlapped with filtered molecular data to verify presence and assessed by our neuro-oncological experts.

**Hartwig Medical Foundation: RNA Sequencing**

RNA was isolated from biopsy using the QIAsymphony RNA Kit (Qiagen, Hilden, Germany) for tissue and quantified on the Qubit. A total of 50 - 100 ng of RNA was used as input for KAPA RNA HyperPrep Kit with RiboErase (Human/Mouse/Rat) library preparation (Roche) on an automated liquid handling platform (Beckman Coulter). RNA was fragmented (high temperature in the presence of magnesium) to a targeted fragment length of 300bp. Barcoded libraries were sequenced as pools on either a NextSeq 500 (V2.5 reagents) generating 2 x 75 base read pairs or on a NovaSeq 6000 generating 2 x 150 base read pairs using standard settings (Illumina, San Diego, CA, USA). BCL output from the sequencing platform was converted to FASTQ using Illumina’s bcl2fastq tool (versions 2.17 to 2.20) using default parameters. RNA-Seq data was aligned to GRCh37 using STAR\(^{29}\) to unsorted BAMs with chimeric reads included. Gene and transcript counts were generated and used for subsequent fusion detection using Isofox (https://github.com/hartwigmedical/hmftools/tree/master/isofox).

**Results**

**Genomic overview**

We started off by characterizing the genomes of the 42 glioblastoma samples in our cohort. We detected 1,246,206 Single Nucleotide Variants (SNVs), 218,519 InDels and 2,224 Multi-Nucleotide Variants (MNVs) in the coding regions of the genome (**Table 1**/**Supplementary figure 1**). The types of SNV are unequally distributed (**Supplementary figure 1B**). Transitions are, for instance, more frequent than transversions (**Supplementary figure**
All these small genetic aberrations lead to two dominant types of variants at protein level 1) missense variants (54%) but also 2) synonymous variants (26%), which have no effect on the function of the protein (Supplementary figure 1G). With a mean tumor purity of 70% (Supplementary 1C) we were able to assess, accurately, the genome ploidy (Supplementary figure 1D), which is mostly diploid and the number and types of Structural Variants (1,408 Deletions, 1,099 Translocations, 1,200 Tandem duplications, 599 Insertions, 1,456 inversions: Supplementary Fig 1F). Next, we compared our results to an independent glioblastoma set from the Pan-Cancer Analysis of Whole Genomes (PCAWG) consortium (N=39) (Table 1). Our findings are in concordance with PCAWG, however, we detected 966,176 more noncoding variants than the PCAWG dataset in our CPCT-02 dataset. Of these noncoding variants, 99% were detected in a single, ultramutated sample.

As a natural measure of mutational load, we determined Tumor Mutational Burden (TMB; number of mutations per megabase pair) and performed mutational signature calling. Ninety percent (N=38) of tumors showed a TMB below 5.0 (Figure 1A), with a mean of 2.6. Three tumors had a TMB exceeding 10.0 (hypermutated) and one tumor exceeded 100 (ultramutated). Mutational signatures of tumors with low TMB included clock-like and age related signature 1, whereas tumors with high TMB presented additional defective DNA Mismatch Repair (MMR) signatures (Figure 1E). The ultramutated sample was the only sample with a contribution from the alkylating agent resembling mutational signature 11. All high-TMB samples also showed microsatellite instability profiles, with genomes enriched for InDels in satellite regions. The genome-wide ploidy (Figure 1B) is generally diploid, but five samples showed whole-genome duplication events, even though many chromosomal arms are altered in copy-number.

Copy-number alterations

We detailed both arm-level (Figure 2) and focal CNAs (Figure 3). Amplification of chromosome 7 (79%) and deletion of chromosome 10 (86%) were most prevalent in our cohort. Loss of 9p occurred in 26%, making it the only chromosome to have statistically significant alterations limited to the small arm. Eight chromosomes (1, 7, 10, 13, 14, 19, 20, 22) were significantly altered (q-value < 0.05) in copy-number on both arms.

Structural variants

The median number of SVs per sample was 98 (ranging from 20 to 433) and was similar to that of the PCAWG glioblastoma set (Table 1). The absolute number of the different types of SVs per sample and the associated genomic widths are depicted in Supplementary figure 2. While the absolute number of SVs differed from sample to sample, the relative frequency was relatively narrow with quartile 1 (Q1) from 9.67 to Q3 of 12.6. Deletions showed the highest Interquartile Range (IQR), as well as the highest mean (28.2%) and max (80%) relative to the
sample. Insertions were detected least, with a mean of 7.9% and a max of 30.7% relative to the sample.

Chromothripsis (CT) occurred in ten patients (24%, Figure 1G) and kataegis in seven patients (16%, Figure 1H). While CT in other cancer types typically shatters entire chromosomes or chromosome-arms, events tend to affect smaller regions on single chromosomes, as previously reported in glioblastoma. CT of chromosomes 1 and 12 were the most common events in our cohort (Supplementary figure 3, Supplementary data file 1). We detected no CT on chromosomes 7 and 9, whereas previous studies linked this to a higher incidence of EGFR, MDM2 and/or CDK4 (hyper-)amplifications and CDKN2A homozygous deletions. All these amplifications/deletions were independent of chromothripsis events in our cohort based on X² tests (alpha <= 0.05). Five samples contained an FGFR3-TACC3 fusion and there was one instance of a CCDC6-RET fusion (Supplementary figure 4). With the access to two RNA-seq samples we could confirm actual fusion transcripts of both fusions in their respective samples; one with a FGFR3-TACC3 and one with a CCDC6-RET fusion. This 100% concordance was identified with a minimum of 38 fragment counts supporting the fusion transcripts in RNA-Seq data. The incidence of the FGFR3-TACC3 fusion (11.9%) is higher (p-value 0.00435; Fisher’s exact test) than previous reports, which ranged 1.2-3.1% (Singh et al., Brennan et al.).

Non-coding variants

WGS is well suited to study both coding and non-coding variants. We therefore scrutinized recurrent variants in our cohort and found two reliable variants in the TERT promoter (C228T in 19 samples and C250T in 10 samples). Sakthikumar et al. pose that non-coding constraint mutations (NCCM) could play an essential role in glioblastoma, finding enrichments for these variants in their cohort. They split these NCCMs between “related to key glioblastoma genes” (354 variants in 60 genes) and “other protein-coding genes” (591 variants in 43 genes). We also assessed which of their posited NCCMs were present in our cohort, resulting in just two occurrences: one in SULT1B1 and one in MET.

Driver gene analysis

Next, we screened for enrichments (q-value < 0.05) of coding mutations and copy-number alterations, which presented nine well known driver genes of glioblastomas (Figure 4D). The most noticeable drivers were CDKN2A (homozygous deletion in 69%), TERT (promoter variants in 67%), EGFR (amplified in 42.9% and enriched in mutations) and PTEN (deletions in 9.5% and enriched in mutations). Most EGFR amplified samples also contained SVs or missense variants within the gene. Furthermore, we detected deletions in the ERBB receptor feedback inhibitor 1 encoding gene ERRFI1 (homozygous deletions, N=2), amplifications of CDK4 (14.3%), PDGFRα (11.9%) and CCND2 (9.5%), and disruptive TP53 mutations. This last gene was the only driver exclusively
significantly enriched based on small mutations. Mutation status of genes commonly related to glioblastoma (such as IDH1/IDH2) not passing our threshold value of significance (p-adjusted < 0.05) are depicted in Supplementary figure 5.

**Potential drug targets**

Finally, we assessed potential targets detectable through whole-genome sequencing by utilizing databases with clinical trial data and drug-target combinations in cancer patients, such as iClusion data. We found biomarkers and potentially targetable events in all but one of our 42 samples (Figure 5A, 5C), with a median of three targetable genes per sample. Results of clinical trials have been published for some of these (e.g., EGFR). The type of targetable change or biomarker is represented in Figure 5B, with the specific change and treatment listed in Supplementary Datafile 1.

**Discussion**

By investigating driver gene events, copy-number alterations, we found that our cohort contained most of the prevalent glioblastoma hallmarks, such as TERT promoter mutations, EGFR amplification and CDKN2A deletion. These findings are in line with other glioblastoma datasets, as all events that drive mutational selection in our set of glioblastoma tumors have been previously described. Non-coding variants presented the largest difference in mutational frequencies between the CPCT-02 study and PCAWG. Since we found just two non-coding variants out of 945 Sakthikumar et al. previously reported in glioblastoma, these non-coding variants were not of relevance in our cohort. Also in line with previous reports for glioblastoma, chromothripsis (24% of our cohort) generally occurs in smaller regions than in other cancer types. Contrary to the findings of Ah-Pine et al., no CT occurred on chromosomes 7 or 9, and most commonly detected on chromosomes 1 and 12 in our cohort. On chromosome 12, MDM2 and CDK4 amplification were independent of CT events. It should be noted that not only different platforms were used (WGS versus SNP arrays), but also different definitions of CT. While it is encouraging to see the hallmarks represented in our dataset, it does raise the question if the differences in chromothripsis are purely due to disparity in definition. However, this can likely be answered when the field settles on a gold-standard chromothripsis definition in general and in glioblastoma specifically.

We investigated mutational signatures, which greatly differed between low and high-TMB samples. This is akin to the (panel-based) data presented by Touat et al., with two types of high-TMB tumors (treatment-induced and de novo). One patient pre-treated with temozolomide showed signature 11 (resembling the mutational pattern of alkylating agents) in its mutational landscape. The larger contribution of defective DNA MMR signatures in the four high-TMB samples align with microsatellite instability, which is a marker of this phenomenon. Touat et al. suggested a high-TMB cut-off of 17.0 based on the MSK-IMPACT panels. Based on MSI-status prediction and the
larger contribution of defective DNA MMR signatures in our cohort, four samples fit the high-TMB mold in our cohort, passing a WGS cut-off of 10 somatic mutations per Mb. Additionally, we have one genomically-distinct sample with 342 mutations per Mb, which was categorized as ultramutated\textsuperscript{31}. This ultramutated sample shows a similar mutational signature pattern as the treatment-induced hypermutated samples Touat et al. present, while the other three high-TMB match the \textit{de novo} samples. The fact that the genomic MSI-status overlaps perfectly with our high-TMB samples, suggests that this might be an alternative means of determining TMB status in glioblastoma\textsuperscript{33}.

We presented an overview of potential targets and biomarkers based on genomic characteristics of our cohort where 41 out of 42 samples showed at least one targetable gene alteration. In an effort to remain comprehensive, we did not exclude any targets that have not shown promise in clinical trials in the past. The fact that there is a median of three potentially targetable gene alterations for these samples, provides optimism for the improvement on the standard treatment modalities for glioblastoma patients at a personal level. Since this list consists of targets (genes or alterations) linked with drugs and is not glioblastoma-specific, there is no guarantee these therapeutics have the capacity to cross the blood–brain barrier\textsuperscript{38}. For instance, \textit{EGFR} has been posited as a prime target for glioblastoma patients, due to the amplification of the gene and protein overexpression. However, its promise is as of yet unfulfilled, with many hypotheses for the reasons behind previously unsuccessful trials\textsuperscript{36}. Drug delivery across the blood-brain barrier seems to be the most problematic issue, but intratumoral adaptations at a pharmacodynamic level may also constitute resistance to therapy. The negative impact of these limitations need to be considered before the observed genetic alterations are targeted clinically. For example, \textit{CDKN2A} and \textit{CDKN2B} were frequently deleted in our cohort. Clearly, a homozygously deleted gene cannot be targeted, but alternate means of reactivating the pathway can prove useful. While not applicable for most patients, our five samples with the \textit{FGFR3-TACC3} fusion events show a higher incidence (11.9\%) than the previously reported 1-3.1\%. This oncogenic fusion might facilitate targeted treatment with \textit{FGFR3} kinase inhibition\textsuperscript{39}. Our other reported fusion (\textit{CCDC6-RET}) is rather prevalent in thyroid and lung cancers and shows selective sensitivity to \textit{RET} inhibitors in clinical trials with those cancer types\textsuperscript{35,40}. We confirmed fusion gene detection data based on RNA-Seq from two of the five samples we reported. As multiple potential targetable genes are present in these samples, in-depth genomic knowledge by WGS will improve clinical study design and treatment allocation with targeted therapies in glioblastoma patients.
In conclusion, we have described the genomic landscape by means of the tumor mutational burden, driver gene events, and mutational signatures in a cohort of 42 glioblastoma patients included in CPCT-02. Several factors (such as mutational frequencies) coincide with other datasets, while others do not (chromothripsis locations, noncoding mutations). Finally, we have alluded to several potential treatment options and biomarkers for 41 out of 42 patients, including within ongoing clinical trials, based on molecular traits. This reveals new avenues for clinical trials and can improve patient management through potential personalized medicine in glioblastoma patients.

Acknowledgements

This publication and the underlying study have been made possible partly on the basis of the data that Hartwig Medical Foundation (HMF) and the Center of Personalised Cancer Treatment (CPCT) have made available to the study.
References


Figure Captions

Figure 1. Overview of genomic characteristics of CPCT-02 glioblastoma whole-genome-sequencing cohort (N=42). Track A shows the tumor mutational burden (mutations per Mb; blue for Single Nucleotide Variants (SNVs), yellow for Insertions/Deletions (Indels) and red for Multi-Nucleotide Variants (MNVs)). Track B confers the mean genome-wide ploidy for each sample. Tracks C and D illustrate the abundance of structural variants and the relative frequency of the type of variants. Tracks E and F show the relative mutational signature contribution (COSMIC signatures v3) and the relative frequency of mutational changes at base level. Tracks G and H displays which samples exhibit predicted catastrophic events, namely chromothripsis and kataegis. Finally, track I shows predicted microsatellite instability profiles of the sample. Figure is ordered descendingly by tumor mutational burden on the x-axis.

Figure 2. Arm-level copy-number alterations and significantly altered (q-value < 0.05) chromosome arms according to GISTIC2.0 in the CPCT-02 glioblastoma whole-genome-sequencing cohort (N=42). Each column is a glioblastoma sample (ordered descendingly by tumor mutational burden), with the chromosome arm listed on the y-axis. Increase in copy-number is displayed in the grid cells in yellow and decrease in purple, white squares show no change in copy-number. Highlighted rows with bold chromosome-arm names indicate statistically significant arm-level changes.

Figure 3. Circular overview of the copy-number alterations in the CPCT-02 glioblastoma whole-genome sequencing cohort (N=42). The outer ring shows the chromosomal ideogram, followed by a cohort-wide GISTIC2.0 G-score track with large peaks rounded to 1.5 and -1.5. Negative copy-numbers on the y-axis (blue) indicate deletions, with positive (green) indicating amplifications. The darker color is indicative passing the statistical q-value threshold of 0.05. Known cancer driver genes overlapping copy-number peaks found to be significant by
GISTIC2.0 are labelled in the center of the circle, utilizing the same color scheme as the G-score track.

Figure 4. Overview of coding mutations and copy-number alterations in driver genes (determined by dN/dS and GISTIC2.0) in the CPCT-02 glioblastoma whole-genome-sequencing cohort (N=42). Track A shows the tumor mutational burden (mutations per Mb; blue for Single Nucleotide Variants (SNVs), yellow for Insertions/Deletions (Indels) and red for Multi-Nucleotide Variants (MNVs)). Tracks B and C show predicted Microsatellite Instability (MSI) status of the sample and chromothripsis status. The oncoplot in track D shows mutations (filled center) and copy-number alterations (grid cell background). Figure is ordered descendingly by tumor mutational burden on the x-axis.

Figure 5. Overview of DNA-based biomarkers and potential treatment options in 42 whole-genome sequenced glioblastoma samples (CPCT-02). A highest level of potential treatment options available based on genomic characteristics. On-Label indicates treatment registered by federal authorities for glioblastoma, whereas Off-Label indicates a registration for other tumour types. B the distribution of actionable target types in percentages. C potentially actionable alterations at gene-level with every column portraying a sample, ordered descendingly by tumor mutational burden on the x-axis. A tile is colored pink in case of an On-Label treatment and blue for Off-Label.

Data and material availability

The data that support the findings of this study are available from Hartwig Medical Foundation, which were used under data request number DR-093 for the current study. Both WGS and clinical data are freely available for academic use from the Hartwig Medical Foundation through standardized procedures and request forms can be found at https://www.hartwigmedicalfoundation.nl. All tools and scripts used for processing of the WGS data are available at https://github.com/hartwigmedical/ and/or can be provided by authors upon request.
Tables

Table 1. Median mutational frequencies of the CPCT-02 glioblastoma cohort (N=42) and PCAWG cohort (N=39). Quartiles 1 and 3 are indicated in brackets and the Mann-Whitney U test was applied as statistical test. Single Nucleotide Variants (SNVs), Multi-Nucleotide Variants (MNVs), Structural Variants (SVs).

<table>
<thead>
<tr>
<th>Median [Q1, Q3]</th>
<th>CPCT-02</th>
<th>PCAWG</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding variants</td>
<td>63.5 [58-82.5]</td>
<td>61 [51.5-80.5]</td>
<td>0.4134</td>
</tr>
<tr>
<td>SNVs</td>
<td>60.5 [54-78.5]</td>
<td>59 [48-76]</td>
<td>0.4989</td>
</tr>
<tr>
<td>InDels</td>
<td>4 [2-6.75]</td>
<td>2 [1-3.5]</td>
<td>0.004669</td>
</tr>
<tr>
<td>MNVs</td>
<td>0 [0-0.75]</td>
<td>N.A.</td>
<td>N.A.</td>
</tr>
<tr>
<td>Noncoding variants</td>
<td>7957 [7016-8749]</td>
<td>7322 [5590-8959]</td>
<td>0.1549</td>
</tr>
<tr>
<td>SVs</td>
<td>98 [69.75-180.25]</td>
<td>100 [77.25-146]</td>
<td>0.8661</td>
</tr>
</tbody>
</table>


Figure 2: Arm-level copy-number alterations of Glioblastoma cohort (N=42)
Figure 3: Circular overview of the copy-number alterations in the CPCT-02 glioblastoma whole-genome-sequencing cohort (N=42)
Figure 4: Driver genes in Glioblastoma cohort (N=42)

A. Mutation per Mb (Genome-wide NGS, Agilent)

B. MSI Status

C. Chromothripsis

D. Mutational Categories

Driver genes

- TERT
- CDKN2A
- EGFR
- PTEN
- ERRF
- CDK4
- TP53
- CCND2
- PDGFR

Mutational categories:
- SNV
- Insertions
- MNV

MSI Status
- MSI
- Stable

Mutational Categories:
- Deep Deletion
- Shallow Deletion
- Neutral
- Shallow Gain
- Deep Gain

- Missense Variant
- Frameshift Variant
- Structural Variant
- Promoter Variant
- Multiple Mutations
- Inframe Insertion/Deletion
- Stop Gained
- Splice Region Variant
- Splice Site Acceptor Variant
- Splice Site Donor Variant
- Synonymous Variant
Figure 5: Overview of potential targets and biomarkers in glioblastoma patients

A

Number of samples

On-Label

Label

Off-Label

B

Actionable change or biomarker

- Missense mutation
- Amplification
- Deletion
- Stop Gain
- Frameshift mutation
- Fusion
- Microsatellite Instability
- InDel
- >100 Nonsynonymous Mutations

C

Potentially actionable genes

Label

- On-Label
- Off-Label

Samples