Title: Comprehensive molecular characterization of multifocal glioblastoma proves their monoclonal origin and reveals novel insights into clonal evolution and heterogeneity of glioblastomas

Journal: *Neuro-Oncology*

Authors: Khalil Abou-El-Ardat, Michael Seifert, Kerstin Becker, Sophie Eisenreich, Matthias Lehmann, Karl Hackmann, Andreas Rump, Gerrit Meijer, Beatriz Carvalho, Achim Temme, Gabriele Schackert, Evelin Schröck, Dietmar Krex, Barbara Klink

Contact author address: Institute for Clinical Genetics, Faculty of Medicine Carl Gustav Carus, TU Dresden, Fetscherstrasse 74, 01307 Dresden, Germany, E-mail: barbara.klink@tu-dresden.de

Supplementary Materials and Methods:

Patients

Four tumor pairs from four patients, three foci from one, and one tumor focus from an additional patient (patient 6, of whom no sufficient tumor material was available from the second focus) were included in this study. None of the patients had a previous history of glioma. The patients were mostly male (5/6) and had an average age of onset of 67 years. The tumors appeared in the same cerebral hemisphere except in patients 1 and 6 (see Supplementary Table S1). Patient 5 was previously published by Krex et al[1]. In addition to fresh frozen tumor samples, primary cell cultures (between passages 6 to 10) were available from four foci of two patients (patients 4 and 5). Tumor foci were considered separated when there was no contrast enhancing tissue between both foci in contrast (Gadolinium™)-enriched T1-weigthed MRI (1.5 Tesla) and no alterations in T2 weighted images suspicious for low-grade tumor.

DNA and RNA extraction from fresh frozen tumor material and primary tumor cells

Tumor content was evaluated on hematoxylin-eosin (H&E)-stained sections of fresh-frozen tumor material by an experienced neuropathologist as described previously and only samples with a tumor content of at least 80 % were included[2]. Also, pieces with extensive tumor necrosis were excluded. DNA was extracted from fresh frozen tumor material and primary cell cultures by phenol:chloroform extraction using standard protocols. RNA was extracted with the QIAGEN miRNeasy Mini Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. RNA quality was assessed using the Agilent RNA 6000 Nano chip on a 2100 Bioanalyzer (Agilent, Santa Clara, CA, USA) and RNAs with an RNA Integrity Number (RIN) below 8 were excluded.

Molecular karyotyping using array comparative genome hybridization (array CGH)

Arrays were scanned using an Agilent microarray scanner. Agilent's CytoGenomics Editions 2.7 and 2.9 were used for extraction and processing of Raw data (using the integrated Featur Extraction software) and to determine deleted and amplified regions based on the draft of the reference human genome (GRCh37/hg19) using the Default Analysis Method - CGH v2. All results were additionally checked by eye to confirm results and were evaluated by a board-certified medical geneticist (Barbara Klink). CNVs that were commonly found in the Database of Genomic Variants (http://dgv.tcag.ca/) and therefore can be considered as polymorphisms that are most likely germline were excluded from further analysis.

Spectral Karyotyping (SKY)

SKY analysis was performed as described previously[3]. SKY images of about 20 metaphase chromosomes per cell line stained with a mixture of five fluorochromes (green, orange, red, far-red, and near-infrared) were captured using a DMRXA epifluorescence microscope (Leica Mikrosysteme Vertrieb GmbH, Wetzlar, Germany) with an HCX PL SAPO 63x/1.30 oil

objective (Leica) with the SpectraCube® system (Applied Spectral Imaging, Migdal HaEmek, Israel) and the SKYView® imaging software (Applied Spectral Imaging).

RT-PCR for EGFRvIII detection

Total RNA was extracted using Qiagen's miRNeasy Mini Kit (Qiagen GmbH, Hilden, Germany) and reverse transcribed to cDNA using SuperScript VILO cDNA Synthesis Kit (Invitrogen, Carlsbad, USA). PCR was performed using the *EGFR* primers published by Yoshimoto et al.[4] (Ex1-F/EGFRvIII-R) and those designed by us (Ex1-F/Ex8-R) using the following program: an initial incubation step of 95 degrees for 15 minutes followed by 34 cycles of denaturation at 95 degrees for 30 seconds, annealing at 60 degrees for 30 seconds and elongation for 1.5 minutes at 72 degrees (primer sequences are in Supplementary Table S2). A final elongation step at 72 degrees for 10 minutes was added. The products were run on a 1.5% agarose gel stained with GelRed (Thermo Scientific, Fisher Scientific, Schwerte, Germany) at 110V for 1 hour in 1x Tris-Borate-EDTA (TBE) buffer. The gel was photographed under UV light.

Microarray data analysis and identification of differentially regulated genes

Measured gene-specific probe hybridization intensities of all samples were transformed into log2-intensities and quantile normalized [5]. Mapping of Agilent probe identifiers to genes was done using the Ensembl Biomart (GRCh37.p12). For genes with more than one corresponding probe, gene-specific log2-intensities were computed by averaging across all gene-specific probe intensities. This resulted in a normalized data set containing gene-specific log2-intensities for 24,933 genes for each sample. Finally, log-ratio gene expression profiles of each tumor sample were generated with respect to each reference normal brain sample [Stratagene Human Brain Total RNA (Agilent, Santa Clara, CA, USA), Clonetech Human Brain Total RNA (Takara Bio Europe SAS, Saint-Germain-En-Laye, France), BioChain Total RNA Human Adult Normal Tissue: Brain: Frontal Lobe (BioCat GmbH, Heidelberg, Germany].

Tumor samples were further associated with clinically relevant subtypes (Proneural, G-CIMP, Neural, Classical, Mesenchymal) described by Verhaak et al.[6] and Brennan et al. [7]. Therefore, the corresponding data of the Verhaak classifier (ClaNC 840 gene list) was downloaded from https://tcga-data.nci.nih.gov/docs/publications/gbm_exp/ and for the G-CIMP phenotype taken from Tab. 2 in Noushmehr et al. [8]. The relative expression level (log2-ratio) of each gene in the tumor compared to normal brain tissue was calculated. These were then used to calculate the correlations between the molecular subtypes, and to test whether the correlation of an individual sample from a specific subtype was significantly greater than zero (Pearson's product moment correlation test) as described previously [9]. Our subtype classification is limited to the set of genes that were overlapping between the Verhaak signatures and the microarray used in this study, while the Verhaak subtypes of GBMs were revealed by combining gene expression data from different platforms. They selected for each subtype a set of representative genes with characteristic subtype-specific expression behavior. As it is shown in Fig2 of the Verhaak paper, the vast majority of these subtype-specific genes have very similar expression patterns. In other words, they have highly correlated gene expression

profiles. Thus, even if genes from this multi-platform approach are not present on our microarray, the other signature genes with very similar expression levels that are present on our microarray still provide sufficient information for correct subtype classifications, as shown previously [9].

Further, a Hidden Markov Model (HMM) with second-order state-transitions and second-order autoregressive emissions introduced in Seifert et al. was used to identify differentially expressed genes in the log-ratio profile of each tumor sample [10]. Since the normal brain reference samples were very similar, we computed an average normal brain reference across all samples and determined the corresponding log-ratio gene expression profile for each tumor sample. Standard settings were used to adapt the HMM to the tumor log-ratio profiles. The resulting HMM was used to classify each gene in the tumor profile either as underexpressed, unchanged, or overexpressed in tumor compared to the normal reference. We further analyzed the identified underexpressed and overexpressed genes of each tumor sample in the context of known cancer signaling pathways. We therefore made use of the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways in cancer overview (path:hsa05200) for which we determined corresponding pathway genes based on ConsensusPathDB [11]. We further extended this KEGG cancer pathway view by systematically adding other known cancer-relevant pathways for DNA repair, telomere maintenance, DNA replication and Hedgehog signaling. A detailed summary of included cancer signaling pathways with functional annotations is given in Table S4 of Text S1 in Seifert et al. [10] An additional pathway analysis was performed for highly differentially expressed genes compared to normal brain samples (strongly underexpressed: log2-fold-change < -3; strongly overexpressed: log2-fold-change > 3) using DAVID's (ver. 6.7; http://david.abcc.ncifcrf.gov/) functional annotation tool on KEGG pathways and functional annotation clustering on Biological Biochemical Image Database (BBID), Biocarta, and KEGG pathways.

Panel Next Generation Sequencing

Fifty ng of DNA from all twelve tumors and from blood available from patients 1-4 was enriched for 1737 exons of 94 genes using Illumina's TruSight Cancer Panel (Illumina Inc., San Diego, CA, USA) and was sequenced on an Illumina MiSeq Desktop sequencer (Illumina Inc., San Diego, CA, USA) using a 100-bp paired-end approach and Illumina's V2 chemistry (for a list of genes please see http://www.illumina.com/products/trusight_cancer.html). The high quality reads in the FastQ files were mapped to the hg19 (GRCh37) draft of the human genome using the CLC Genomic Workbench's proprietary mapping algorithm (ver. 7.4; Qiagen, Aarhus, Denmark). Variant calling was done using the Fixed Ploidy Variant Caller for variants appearing in more than 1% of the reads and all variants that were called in the blood were filtered out as being germline where the blood was available. The remaining variants were filtered for nonsynonymous exonic and splice site mutations.

Evaluation of mutations identified using NGS and Sanger-Sequencing

The potential impact of missense mutations was assessed using five mutation impact prediction tools: SIFT (http://sift.jcvi.org/)[12], PolyPhen2 (http://genetics.bwh.harvard.edu/pph2/)[13], Mutation Taster (http://www.mutationtaster.org/)[14], Mutation Assessor (http://mutationassessor.org/)[15], and Combined Annotation Dependent Depletion (CADD; http://cadd.gs.washington.edu/)[16]. Furthermore, the frequency of all variants in large datasets of normal human genomes and exomes was cross-checked: The

1000 Genome Project (http://www.1000genomes.org), The HapMap Project (http://hapmap.ncbi.nlm.nih.gov), The Exome Variant Server (EVS; http://evs.gs.washington.edu/EVS/), and the Exome Aggregation Consortium (http://exac.broadinstitute.org/), and only variants with a frequency below 0.01 were considered. Finally, the variants were looked up in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/) and in the single nucleotide polymorphism database (dbSNP138; http://www.ncbi.nlm.nih.gov/snp/).

Only variants that were rare (< 1% of the population as reported in the HapMap, EVS and Exome Aggregation Consortium) and either clearly protein disrupting (e.g. splice site mutations, stop mutation, frameshift mutation) or missense variants predicted to be protein damaging according to three of the five mutation impact tools are reported in our manuscript.

All mutations identified that were different between foci in one patient were manually checked in the reads of the other foci to exclude false negative callings due to filtering.

References

1. Krex D, Mohr B, Appelt H, Schackert HK, Schackert G. Genetic analysis of a multifocal glioblastoma multiforme: a suitable tool to gain new aspects in glioma development. *Neurosurgery.* 2003; 53(6):1377-1384; discussion 1384.
2. Eisenreich S, Abou-El-Ardat K, Szafranski K, et al. Novel CIC point mutations and an exon-spanning, homozygous deletion identified in oligodendroglial tumors by a comprehensive genomic approach including transcriptome sequencing. *PLoS One.* 2013; 8(9):e76623.
3. Schrock E, du Manoir S, Veldman T, et al. Multicolor spectral karyotyping of human chromosomes. *Science.* 1996; 273(5274):494-497.
4. Yoshimoto K, Dang J, Zhu S, et al. Development of a real-time RT-PCR assay for detecting EGFRvIII in glioblastoma samples. *Clin Cancer Res.* 2008; 14(2):488-493.
5. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics.* 2003; 19(2):185-193.
6. 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(1):98-110.
7. Brennan C, Verhaak R, McKenna A, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013; 155:462-477.
8. Noushmehr H, Weisenberger DJ, Diefes K, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*. 2010; 17(5):510–522.
9. Seifert M, Garbe M, Friedrich B, Mittelbronn M, Klink B. Comparative transcriptomics reveals similarities and differences between astrocytoma grades. *BMC Cancer.* 2015; 15(1):952.
10. Seifert M, Abou-El-Ardat K, Friedrich B, Klink B, Deutsch A. Autoregressive higher-order hidden markov models: exploiting local chromosomal dependencies in the analysis of tumor expression profiles. *PLoS One.* 2014; 9(6):e100295.
11. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: toward a more complete picture of cell biology. *Nucleic Acids Res.* 2011; 39(Database issue):D712-717.
12. Choi Y, Sims GE, Murphy S, Miller JR, Chan AP. Predicting the functional effect of amino acid substitutions and indels. *PLoS One.* 2012; 7(10):e46688.
13. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. *Nature methods.* 2010; 7(4):248-249.
14. Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nature methods.* 2014; 11(4):361-362.
15. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. *Nucleic acids research.* 2011; 39(17):e118.
16. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature genetics.* 2014; 46(3):310-315.