In silico validation of RNA-Seq results can identify gene fusions with oncogenic potential in glioblastoma

Ainhoa Hernandez, Ana Maria Muñoz-Mármo, Anna Esteve-Codina, Francesc Alameda, Cristina Carrato, Estela Pineda, Oriol Arpi-Lluciá, Maria Martínez-García, Mar Mallo, Marta Gut, Sonia del Barco, Oscar Gallego, Marc Dabad, Carlos Mesia, Beatriz Bellosillo, Marta Domenech, Noemí Vidal, Iban Aldecoa, Nuria de la Iglesia & Carmen Balana*

RNA-sequencing (RNA-Seq) can identify gene fusions in tumors, but not all these fusions have functional consequences. Using multiple data bases, we have performed an in silico analysis of fusions detected by RNA-Seq in tumor samples from 139 newly diagnosed glioblastoma patients to identify in-frame fusions with predictable oncogenic potential. Among 61 samples with fusions, there were 103 different fusions, involving 167 different genes, including 20 known oncogenes or tumor suppressor genes (TSGs), 16 associated with cancer but not oncogenes or TSGs, and 32 not associated with cancer but previously shown to be involved in fusions in gliomas. After selecting in-frame fusions able to produce a protein product and running Oncofuse, we identified 30 fusions with predictable oncogenic potential and classified them into four non-overlapping categories: six previously described in cancer; six involving an oncogene or TSG; four predicted by Oncofuse to have oncogenic potential; and 14 other in-frame fusions. Only 24 patients harbored one or more of these 30 fusions, and only two fusions were present in more than one patient: FGFR3::TACC3 and EGFR::SEPTIN14. This in silico study provides a good starting point for the identification of gene fusions with functional consequences in the pathogenesis or treatment of glioblastoma.

Glioblastoma is the most aggressive primary brain tumor. Standard therapy is surgery followed by radiation therapy with concomitant and adjuvant temozolomide, but median survival remains around 14–16 months. Except for the prolonged progression-free—but not overall – survival afforded by bevacizumab, no pharmacological intervention has been able to alter the course of the disease. Considering this poor prognosis and lack of effective therapies, it is clearly important to develop novel treatment strategies based on molecular data.
Gene fusions are chimeras of two coding or regulatory DNA sequences. Some result from genomic rearrangements that give rise to a single transcription unit, while others originate by trans-splicing and are only present at the transcript level. Several biological processes contribute to the formation of gene fusions and there are multiple computational tools to analyze them. The increasing importance of gene fusions in solid tumors has recently been recognized due to the emergence of high-throughput technologies, such as RNA-Seq.

Gene fusions have been described in different tumor types, but most appear not to have functional consequences, although some are involved in the initial steps of tumor development and progression.

The first fusion to be identified in glioblastoma was FIG::ROS1, in which an intrachromosomal deletion leads to a constitutively active kinase with oncogenic activity. Since then, multiple studies and case reports have described different low-frequency fusions in 30–50% of glioblastomas. The genes most involved in fusions in IDH wild-type glioblastoma are EGRF (6–13%), FGFR3 (3%), MET (1–4%) and the NTRK gene family (1–2%). All of these genes codify for receptor tyrosine kinases, whose rearrangement leads to oncogenic kinase activation. Several drugs have been approved by the FDA as standard therapy for tumor patients harboring specific gene fusions, but glioblastomas are underrepresented because of the low frequency of recurrent fusions.

Traditional methods for detecting fusions include Southern blotting, fluorescent in situ hybridization (FISH), and RT-PCR. Next generation sequencing (NGS), including RNA-Seq, can provide a wealth of information on gene expression and chromosomal rearrangements. However, data interpretation is hindered by several constraints: false positives and negatives can confound results; fusions known to be present in healthy tissue must be ruled out as they do not have oncogenic potential; and the translational function of the fusion (driver vs passenger) needs to be identified. Moreover, not all fusions involve genes with a potential or demonstrated role in cancer and not all of them generate in-frame gene fusions, with transcripts that could produce a protein with functional biological effects. Hence, not all fusions are optimal candidates for further validation.

We have examined gene fusion detected by RNA-Seq in a series of newly diagnosed glioblastomas and performed an in silico study to predict their oncogenic potential. We selected fusions with demonstrated or possible oncogenic potential and examined their frequency and their correlation with patient characteristics and outcome, with the aim of identifying frequently recurrent fusions that warranted validation.

Results

Tumor tissue samples were obtained from 329 of the 432 glioblastoma patients registered in the GLIOCAT project. After multiple RNA extractions from each sample, 357 RNA libraries were prepared and RNA-Seq results were obtained for 151 patient tumor samples. Fusions were assessed with STAR fusion in 139 formalin-fixed, paraffin-embedded (FFPE) samples obtained at first surgery. Four of these samples had paired fresh-frozen (FF) samples obtained at first surgery and four had FFPE samples obtained at a second surgery performed at the time of recurrence. Data on molecular subtypes according to the Gene Set Enrichment Analysis (ssGSEA) and Intrinsic Glioma Subtypes (IGS) algorithms were available for 124 samples obtained at first surgery using the GlioVIS and the R clusterRepro packages.

Gene fusions by initial RNA-Seq in glioblastoma samples. Among the 139 patients with FFPE samples obtained at first surgery, RNA-Seq detected one or more fusions in 61 (43.9%). Table 1 displays the patient characteristics according to the presence or absence of fusions. Fusions were more prevalent in the classical TCGA and the IGS-18 subtypes. Tumors with MGMT methylation had more fusions than those without methylation. Of the 68 tumors with MGMT methylation, 36 (59%) had fusions, while of the 63 tumors without MGMT methylation, only 20 (32.8%) had fusions (P = 0.01).

Among the 61 tumor samples with fusions, there were a total of 263 fusions, corresponding to 103 different fusions, with a median of two fusions per sample (range, 1–13). Nine fusions were detected in more than one sample and 101 were detected more than once in the same sample at different breakpoints (Supplementary Table S1 and Supplementary Dataset 1).

Of the 103 different fusions detected, 79 were intrachromosomal and 24 interchromosomal. The majority of fusions were located at chromosome 12, where there were 40 fusions, 34 of which were intrachromosomal (33% of all fusions). Chromosome 7 had 22 fusions, 15 of which were intrachromosomal (14.6% of all fusions). Chromosomes 3 and 9 had 16 and 8 fusions, respectively (Fig. 1).

Using FusionHub, we classified fusions according to the type of genes they included. The 103 fusions involved 167 different genes (Supplementary Tables S2A and S2B and Supplementary Dataset 1) that can be classified as follows: 1) known oncogenes or tumor suppressor genes (TSGs) (n = 20, 11.9%); 2) genes associated with cancer but not oncogenes or TSGs (n = 16, 9.6%); 3) genes not associated with cancer but involved in fusions in gliomas (n = 32, 19.2%); 4) not associated with cancer and not involved in fusions in gliomas (n = 99, 59.3%) (Supplementary Tables S3A–D, respectively, and Supplementary Dataset 1).

Selection of fusions with oncogenic potential. As shown in Fig. 2, we first eliminated six fusions (detected in 45 samples) because they had previously been detected in healthy tissue (Table 2), which would indicate no relevant role in cancer. Of the remaining 97 fusions, ten (detected in 14 samples) had previously been detected in cancers, including gliomas (Table 2) and 21 (detected in 33 samples) had not previously been identified in cancer but included an oncogene or TSG (Table 3). The remaining 66 fusions have not been described in healthy tissue or in cancer and did not include an oncogene or TSG.

After eliminating the frame-shifted fusions, verifying the breakpoints with the Integrative Genomics Viewer and running Oncofuse, we classified the remaining 30 fusions in the four previously established categories: (1) six were previously described in cancer; (2) six were not previously described in cancer but involved an oncogene or TSG; (3) four were predicted by Oncofuse to have oncogenic potential; and (4) 14 were other in-frame fusions that gave rise to a constitutively active kinase with oncogenic activity.
that can produce a protein but that have not previously been described in cancer, do not involve an oncogene or TSG, and were not predicted to have oncogenic potential by Oncofuse (Supplementary Dataset 2). These 30 different fusions were considered to have oncogenic potential (Table 4).

Clinical characteristics of patients with gene fusions with oncogenic potential. Twenty-four patient samples harbored one or more of the 30 fusions categorized as having oncogenic potential (Table 4). Two tumor samples had two fusions, two had three fusions, and one had four fusions. When we compared the clinical characteristics of the patients whose tumors had one or more of these fusions, there was no correlation...
with patient age or MGMT methylation status. All patients except one were IDH wild-type. Two patients were secondary glioblastomas with a history of previous low-grade glioma that had been treated with surgery alone. Three patients were classified as TCGA mesenchymal subtype, one of whom was classified as IGS-23 subtype, while the remaining patients were TCGA classical or proneural.

Two fusions had previously been associated with glioblastoma: \textit{FGFR3::TACC3} and \textit{EGFR::SEPTIN14}. Three patients had the \textit{FGFR3::TACC3} fusion, all of whom were men older than 63 years and one of whom had MGMT methylation. Two patients had the \textit{EGFR::SEPTIN14} fusion, both of whom were women with MGMT methylation.

**Figure 1.** Circos plot showing the chromosomes involved in the fusions detected in this study.

**Figure 2.** Procedures and data bases used in the present study to select the fusions with oncogenic potential. From the long list of fusions detected by RNA-Seq, we used STAR-Fusion to detect fusion genes and FusionInspector to validate predicted fusions. We then used FusionHub to eliminate fusions previously described in healthy tissue, identify fusions previously described in cancers, and explore whether either gene had been identified as an oncogene or tumor suppressor gene (TSG) or had been associated with cancer. We next used FusionValidate to select only in-frame fusions and finally ran Oncofuse to predict the oncogenic potential of each fusion.
Table 2. Classification of fusions detected in 61 glioblastoma samples: fusions previously detected in healthy tissue (N = 6) or in cancers (N = 10). LG low-grade glioma, GB glioblastoma.

| Fusion               | Detected in no. samples | Previously described in cancer | Previously described in gliomas | % of all patients included | Type of fusion |
|----------------------|-------------------------|--------------------------------|---------------------------------|---------------------------|----------------|
| KCNB4::CNOT2         | 1                       | No                             | No                             | 0.72                      | In-frame       |
| NUP214::TMOD1        | 1                       | No                             | No                             | 0.72                      | Frame-shifted  |
| PFKFB3::RP11-365J2.2 | 3                       | No                             | No                             | 2.16                      | Unknown        |
| PID1::DNER           | 1                       | No                             | No                             | 0.72                      | In-frame       |
| RP1-34H18.1::NAV3     | 10                      | Yes                            | No                             | 7.2                       | Unknown        |
| RP11-444D3.1::SOX5   | 29                      | No                             | No                             | 21                        | Unknown        |

Detected in cancers (N = 14 samples)

| Fusion               | Detected in no. samples | Oncogene or TSG? | In fusions with other genes in glioma? | Oncogene or TSG? | In fusions with other genes in glioma? | Type of fusion    |
|----------------------|-------------------------|------------------|----------------------------------------|------------------|----------------------------------------|-------------------|
| FRS2::KIF5A          | 1                       | Yes              | No                                     | No               | No                                     | Unknown           |
| EGFR::SEPTIN14       | 2                       | Yes              | No                                     | Yes              | No                                     | Unknown           |
| PFKFB3::TACC3        | 3                       | Yes              | No                                     | Yes              | No                                     | In-frame          |
| CAPZA2::MET          | 1                       | Yes              | No                                     | No               | No                                     | Frame-shifted     |
| CLIC4::SRM1          | 2                       | Yes              | No                                     | Yes              | No                                     | Frame-shifted     |
| DPF5L3::JAKMIP2      | 1                       | Yes              | No                                     | No               | No                                     | Unknown           |
| LANCL2::VOPP1        | 1                       | Yes              | No                                     | Yes              | No                                     | Unknown           |
| R3HDM2::AVIL         | 1                       | Yes              | No                                     | No               | No                                     | Unknown           |
| RAB3P::BEST3         | 1                       | Yes              | No                                     | Yes              | No                                     | Frame-shifted     |
| SEC61G::EGFR         | 1                       | Yes              | No                                     | Yes              | No                                     | In-frame / Frame-shifted |

Table 3. Classification of fusions detected in 61 glioblastoma samples: fusions not previously detected in healthy tissue or cancer but involving an oncogene or TSG (N = 21). TSG tumor suppressor gene, LG low-grade glioma, HG high-grade glioma, GB glioblastoma.

| Fusion               | Detected in no. samples | Oncogene or TSG? | In fusions with other genes in glioma? | Oncogene or TSG? | In fusions with other genes in glioma? | Type of fusion    |
|----------------------|-------------------------|------------------|----------------------------------------|------------------|----------------------------------------|-------------------|
| RP11-384J7.2::LSAMP  | 13                      | No               | No                                     | No               | No                                     | LG                |
| GNAQ::CEP78          | 1                       | Oncogene         | Yes                                    | No               | No                                     | Frame-shifted     |
| MALAT1::EGFR         | 1                       | Oncogene         | Yes                                    | No               | No                                     | Unknown           |
| MITF::ST18           | 1                       | Oncogene         | Yes                                    | No               | No                                     | Unknown           |
| RERE::PSMD6          | 1                       | Oncogene         | Yes                                    | LG               | No                                     | Frame-shifted     |
| VPS53::VWDE          | 1                       | Possible TSG     | No                                     | LG               | No                                     | Frame-shifted     |
| XRCC5::LINC01614     | 1                       | Possible TSG     | No                                     | LG               | No                                     | Unknown           |
| ABL1::SZKD1          | 1                       | Oncogene or TSG  | Yes                                    | No               | No                                     | In-frame / Frame-shifted |
| AGAP2::KIF5A         | 1                       | Oncogene or TSG  | Yes                                    | No               | No                                     | LG                |
| CDK6::RP11-745C15.2  | 1                       | Oncogene         | Yes                                    | No               | No                                     | Unknown           |
| EGFGR::BP3HDM2       | 1                       | Oncogene         | Yes                                    | HG & LG          | No                                     | In-frame          |
| EGFGR::BP11-745C15.2 | 1                       | Oncogene         | Yes                                    | HG & LG          | No                                     | Unknown           |
| HMGA2::LPH           | 1                       | Oncogene         | Yes                                    | GB               | No                                     | Frame-shifted     |
| IAZF1::SEPTP5        | 1                       | Oncogene         | Yes                                    | HG & LG          | No                                     | Unknown           |
| STAT3::CPAP61        | 1                       | Oncogene         | Yes                                    | No               | No                                     | Unknown           |
| USP22::TMC3          | 1                       | Oncogene         | No                                     | No               | No                                     | LG                |
| BEST3::EGFR          | 1                       | No               | No                                     | Yes              | GB                                     | In-frame          |
| C3orf62::FBXN1       | 1                       | No               | No                                     | No               | No                                     | LG                |
| CEP78::GNAQ          | 1                       | No               | No                                     | Yes              | LG                                     | Frame-shifted     |
| CTDSP2::GLI1         | 1                       | No               | No                                     | No               | No                                     | In-frame          |
| SLC35E3::EGFR        | 1                       | No               | No                                     | No               | No                                     | Frame-shifted     |
Table 4. Characteristics of patients with tumors harboring one or more of 30 gene fusions with oncogenic potential. IHC immunohistochemistry, Pro proneural, Cla classical, Mes mesenchymal, NA not available.

| Tumor samples N = 24 | No. fusions detected per sample | Fusions detected in each sample | Patient characteristics |
|----------------------|--------------------------------|-------------------------------|-------------------------|
|                      |                                | Age  | Sex  | MGMT promoter methylation? | Type of glioma | Survival (months) | G-CIMP? | IDH1 mutations? (by IHC) | TCGA subtype | IGS subtype |
| AC0340               | 1                              | PID1::DNER<sup>a</sup>         | 67   | Man  | No  | Primary    | 7.62    | No  | No  | Pro  | 9    |
| AC0346               | 1                              | ACVR1B::SCAF11<sup>a</sup>     | 50   | Man  | Yes | Secondary | 30.49   | Yes | Yes | Pro  | 9    |
| AC0365               | 1                              | CN0T2::RBMS2<sup>a</sup>       | 71   | Man  | Yes | Primary   | 10.61   | No  | No  | Pro  | 17   |
| AC6287               | 1                              | NUDT3::MAP4<sup>a</sup>        | 48   | Man  | Yes | Primary   | 27.86   | No  | NA  | Cla  | 18   |
| AA6367               | 2                              | ZMPSTE24::CACNA1D ADD2::C2orf42 | 53   | Woman | No  | Primary   | 4.53    | No  | No  | Cla  | 18   |
| AA6255               | 1                              | AVL::CMAP<sup>a</sup>          | 77   | Man  | Yes | Primary   | 21.36   | No  | No  | Mes  | 18   |
| AC6246               | 1                              | TSF4::KIF5A<sup>a</sup>        | 65   | Woman | No  | Primary   | 12.98   | No  | No  | Pro  | 0    |
| AC6253               | 1                              | PDIA5::JQCB1<sup>a</sup>       | 73   | Man  | Yes | Primary   | 9.69    | No  | No  | Mes  | 23   |
| AC6237               | 1                              | LAMAS::PSMD3<sup>a</sup>       | 71   | Woman | Yes | Primary   | 4.30    | No  | No  | Cla  | 18   |
| AC6281               | 1                              | KIF5A::AVL<sup>a</sup>         | 79   | Man  | Yes | Primary   | 2.79    | No  | No  | Pro  | 9    |
| AC6282               | 1                              | WSB1::SEZ6<sup>a</sup>         | 64   | Woman | No  | Primary   | 8.74    | No  | No  | Pro  | 18   |
| AA6373               | 1                              | PIK3CB::EPHB1<sup>a</sup>      | 54   | Man  | No  | Primary   | 24.15   | No  | No  | Pro  | 18   |
| AA6366               | 1                              | TRK1::TMPRSS12<sup>a</sup>     | 79   | Man  | No  | Primary   | 16.30   | No  | No  | Mes  | 18   |
| AC6276               | 1                              | CREB5::ARCA1<sup>a</sup>       | 78   | Man  | Yes | Primary   | 8.15    | No  | No  | Cla  | 18   |
| AA6380               | 4                              | AGAP2::KIF5A EGFR::R3HDM2 USP22::TMCS3 KCMN8::C<sup>a</sup> | 61   | Man  | Yes | Primary   | 36.50   | No  | No  | Pro  | 22   |
| AC0344               | 1                              | CEP78::GNAQ<sup>a</sup>        | 57   | Man  | Yes | Primary   | 26.18   | No  | No  | Cla  | 18   |
| AA6364               | 1                              | EGFR::SEPTIN1<sup>a</sup>      | 55   | Woman | Yes | Primary   | 42.55   | NA  | NA  | NA   | NA   |
| AC0438               | 1                              | EGFR::SEPTIN1<sup>a</sup>      | 62   | Woman | Yes | Primary   | 21.13   | No  | No  | Cla  | 18   |
| AC0364               | 3                              | CLIC4::SRRM1<sup>a</sup>       | 62   | Woman | No  | Secondary | 12.65   | No  | No  | Cla  | 18   |
| AC6283               | 1                              | CAPZA2::MET<sup>a</sup>        | 80   | Man  | No  | Primary   | 1.51    | No  | No  | Pro  | 9    |
| AA6397               | 1                              | FGFR3::TACC3<sup>a</sup>       | 63   | Man  | No  | Primary   | 32.89   | No  | No  | Cla  | 18   |
| AC6281               | 1                              | FGFR3::TACC3<sup>a</sup>       | 70   | Man  | No  | Primary   | 9.76    | No  | NA  | Cla  | 18   |
| AC2104               | 3                              | FGFR3::TACC3<sup>a</sup>       | 75   | Man  | Yes | Primary   | 30.82   | No  | No  | Cla  | 18   |
| AA6386               | 2                              | SEC61G::EGFR<sup>a</sup> CALD1::ADAM2<sup>a</sup> | 70   | Woman | No  | Primary   | 10.81   | No  | No  | Pro  | 18   |

(Table 4). The remaining fusions with oncogenic potential were each found in only one patient; this low frequency precluded a validation by RT-PCR of these fusions.

There were no differences in overall survival between patients with no fusions, those with fusions with oncogenic potential, and those with non-oncogenic fusions (p = 0.59).

### Comparison of fusions in FFPE vs FF tumor tissue.

Four patients had paired FFPE and FF tissue from the first surgery and four others had paired FFPE tissue from both the first and second surgery. More fusions were detected in FF than in FFPE tissue, but fusions with oncogenic potential were detected in both types of samples. EGFR::SEPTIN1<sup>a</sup> was detected in both FFPE and FF samples from one patient; CLIC4::SRRM1, ZMPSTE24::CACNA1D and ADD2::C2orf42 were detected in samples from another patient; and TBK1::TMPRSS12 was detected in samples from a third patient. FGFR3::TACC3 was detected in the FFPE sample from the first surgery but not in the sample from the second surgery.

### Discussion

In order to explore the role of gene fusions in glioblastoma, we have analyzed fusions by RNA-Seq in tumor samples from 139 newly diagnosed, uniformly treated glioblastoma patients. Since our RNA-Seq results provided a long list of gene fusions, we performed an in silico study to predict their oncogenic potential. We first eliminated the fusions previously described in healthy tissue and then selected those previously described in cancer, those involving oncogenes or TSGs, and those identified by Oncosure<sup>a</sup> as having oncogenic potential. We limited our
selection to in-frame fusions that could produce a protein with a biological effect. We identified a final list of 30 gene fusions with oncogenic potential, which were present in glioblastoma samples from 24 of the 139 patients included in the study. We then examined the frequency of these fusions in our series of glioblastoma patients and their potential correlation with patient characteristics and outcome.

RNA-Seq is useful in the assessment of tumors as a method to detect druggable fusions. However, it provides a multitude of data that do not necessarily have biological significance. Moreover, several of the fusions have not been properly validated individually in the tissue in question, probably because of the large amount of data obtained and the difficulty of identifying the fusions that are biologically meaningful.

Methods for the detection of gene fusions are constantly evolving and it is certain that new methods will become available in the future. We used several methods in our analyses. For example, we used STAR fusion but later ARRIBA became available. Nevertheless, a recent study has shown that both of these methods outperformed others and were equally accurate at detecting fusions. In addition, we ran DEEPrior in parallel with Oncofuse but chose Oncofuse as the final method since we found that Oncofuse results were more reliable. We also considered using PEGASUS but at the time of our study, it used the old version of the human genome (hg19), which dates from 2014. Finally, another method, ChimDriver, has just been reported this year. This diversity of currently available and newly emerging platforms means that it will be necessary to carefully determine the best method to use in the future to detect gene fusions and to identify those with oncologic potential. Two fusions identified as having oncogenic potential in our study had previously been associated with glioblastoma: FGFR3::TACC3 and EGFR::SEPTIN14. The FGFR-TACC fusion has been reported in 1.2–8.3% of glioblastomas. The latest WHO classification of gliomas describes fusions that occur in IDH-wild-type glioblastoma at an estimated frequency of > 1% . EGFR, one of the most frequent genes involved in recurrent in-frame fusions, is commonly found fused to SEPTIN14 or to PSIP, with a frequency of 4% and 2.2%, respectively, in glioblastoma. In our series, EGFR was involved in fusions in 5% of patients but not all the fusions involving EGFR had oncogenic potential. In fact, in-frame fusions involving EGFR with oncogenic potential were only detected in four patients (2.8%): EGFR::SEPTIN14 in two samples, SEC61G::EGFR in one, and EGFR::R3HD2 in one. Of these, only the EGFR::SEPTIN14 fusion was a bona fide driver, as the SEC61G::EGFR and EGFR::R3HD2 fusion proteins would lack the EGFR tyrosine kinase domain. The remaining EGFR fusions detected would produce either a frameshift transcript or no transcript at all.

Other EGFR alterations are also frequent in glioblastoma, including EGFR amplification, the EGFRvIII mutation, and altered splicing and rearrangements. In our study, EGFR amplification was detected by FISH in all cases with EGFR fusions except one (with the EGFR::SEPTIN14 fusion where there was insufficient available tissue for FISH analysis) (data not shown). The co-occurrence of EGFR fusions with EGFR amplification and EGFR vIII (exon 2–7 deletion) has also been previously reported. This “two-hit” alteration has been described for several oncogenes in different tumor types, and it has been suggested that these oncogenes would be dosage-sensitive, with amplification of a mutated copy further increasing tumor fitness. This could be the case in our specimen with the co-occurrence of the EGFR::SEPTIN14 fusion and EGFR amplification, but it would not explain the existence of putatively non-functional EGFR fusions in cases with EGFR amplification. However, previous studies in glioblastoma have reported an increase in DNA breaks near genes targeted by copy number gains, including EGFR. Taking this into account, we can speculate that the non-functional EGFR fusions could be the by-product of localized genome instability and would thus have no significance in the biology of the tumor. Along these same lines, in our study, we have detected several non-productive gene fusions in the 12q region, another breakpoint-enriched region in glioblastoma.

Unfortunately, therapies targeting different alterations of EGFR have failed to confer survival benefit, although these studies did not include EGFR fusions. Fusions involving the NTRK genes have also been reported in glioblastoma, but they are more common in pediatric populations and were not detected in our samples. Other fusions reported in glioblastoma, including LANCL2::RP11-745C15.2, LANCL2::SEPTIN14, and PTPRZ1::MT, were not detected in our samples, although the CAPZA2::MET fusion was detected in one sample (0.7%). At present, only some fusions previously detected in glioblastoma are potentially druggable: ROS1 fusions, FGFR::TACC, NTRK fusions, and MET fusions.

In the present study, we have identified 30 fusions with oncogenic potential. Each of these fusions was detected in < 1% of cases. Therefore, although our intention was to validate by RT-PCR the fusions identified in our in silico study, their low frequency made it unreasonable to do so in our series of patients. Such a low frequency of potentially oncogenic gene fusions suggests that the detection of individual fusions by RT-PCR would be neither reasonable nor cost-effective and that RNA-Seq would thus be the best procedure for searching for targetable fusions. Moreover, we found no correlation with patient characteristics that could identify a patient as a potential holder of any specific fusion.

Nonetheless, although many of the fusions identified in our study have not yet been described in glioblastoma, several of them involve actionable gene alterations that have been successfully targeted in other cancers. Considering the rarity of specific gene fusions in glioblastoma, it is not feasible to conduct a clinical trial limited to this subset of patients. However, the implementation of NGS in the molecular characterization of tumors is helping to identify a constantly increasing number of molecular alterations that are present in small subsets of a plethora of tumor types. We therefore believe that the NGS analysis of glioblastoma may allow the inclusion of glioblastoma patients in exploratory basket trials of specific tumor-agnostic biomarkers, as has been done with other rare gene alterations. Our in silico study to detect in-frame fusions with oncogenic potential thus provides a good starting point for the identification of fusions that may be relevant to the pathogenesis or treatment of glioblastoma.
Methods

Patients and study design. From 2004 to 2014, the GLIOCAT project\(^{49,50}\) collected clinical data from 432 consecutive glioblastoma patients from six institutions, all of whom had received the standard first-line treatment (surgery followed by radiotherapy with concurrent and adjuvant temozolomide). The pathological diagnosis was confirmed by pathologists according to WHO 2007 classification guidelines\(^{51}\) before patients were included in the project. Once selected for inclusion, each case was anonymized and given a number to identify it across all data.

The following data were recorded: age, sex, symptoms, tumor characteristics, radiological characteristics, type of surgery, post-surgical performance, Mini Mental Status Examination (MMSE) score, details of radiotherapy and temozolomide treatments and treatment at relapse, date of progression, subsequent treatments, date and status at last control, and date and status of death or last control alive. Once patients were included in the study, MGMT methylation status was determined if it had not previously been assessed.

This study was approved by the Institutional Review Board of the Hospital Germans Trias i Pujol (PI-14-016) and by the Ethics Committees of all the participating institutions and their biobanks and was conducted in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments. All patients or their representatives gave their informed consent.

Tissue microarray (TMA) construction and immunohistochemical analyses (IHC). TMAs were constructed using a Veridiam Tissue Array Instrument (El Cajon, CA, USA), model VTA-100, using a 1-mm diameter needle. Consecutive 4-µm-thick sections were obtained and hematoxylin–eosin staining was done in sections 1, 20, and 40 in order to evaluate the persistence of the tumor at each spot.

IDH1-R132H analysis was done with the Dianova Cat# DIA-H09, RRID:AB_2335716, antibody. Four cases with doubtful IHC were sequenced to assess IDH status.

DNA extraction and assessment of MGMT methylation. DNA was extracted from two 15-µm sections of FFPE tissue using the QIAamp DNA Mini Kit (QIAGEN GmbH, Hilden, Germany), following the manufacturer’s protocol. In cases with less than 50% of tumor cells, the tumor tissue was macrodissected manually. Then 500 ng of extracted DNA was subjected to bisulfite treatment using the EZ DNA Methylation-Gold Kit (Zymo Research Corporation, Irvine, CA). MGMT promoter methylation status was determined by methylation-specific PCR (MSP) as previously described\(^{52}\).

RNA-Seq assessments. RNA extraction from FFPE and FF samples was performed on five 15-µm deep tissue sections using the RNeasy FFPE Kit (Qiagen, Hilden, Germany) according to the manufacturer’s recommendations. RNA quantity and purity were measured with the NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and the Qubit RNA HS Assay Kit (Invitrogen, Eugene, OR, USA). The highest-quality RNA samples were sent to the Centro Nacional de Análisis Genómico (CNAG-CRG, Barcelona, Spain) for analysis by RNA-Seq. Methods for assessing quantification, purity and quality of samples have been previously described\(^{16,49}\).

The libraries were sequenced on HiSeq2000 (Illumina) in paired-end mode with a read length of 2 × 76 bp using TruSeq SBS Kit v4. Each sample was sequenced in a fraction of a sequencing v4 flow cell lane, following the manufacturer’s protocol. Image analysis, base calling and quality scoring of the run were processed using Real Time Analysis (RTA 1.18.66.3) software and followed by the generation of FASTQ sequence files by CASAVA.

Classification of glioblastoma molecular subtypes. The TCGA classification of glioblastoma molecular subtypes\(^{17,33,53}\) was performed with the GlioVis portal\(^{18}\). The GlioVis glioblastoma TCGA cohort according to the ssGSEA method was selected as training dataset for both glioblastoma molecular subtype and Glioma CpG island methylator phenotype (G-CIMP) predictions. The IGS classification of glioblastoma molecular subtypes was done with the RclusterRepro package\(^{17,19}\), using the centroids for IGS0, IGS9, IGS16, IGS17, IGS18, IGS22 and IGS23 subtypes, as described by Gravendeel\(^{19}\) and used in several European series\(^{54,55}\).

Identification of candidate gene fusions. Figure 2 shows the procedures and data bases used in the present study to select the fusions with oncogenic potential. STAR-Fusion (https://github.com/STAR-Fusion/STAR-Fusion/tree/STAR-Fusion-v1.9.0)\(^{9}\) was used to detect fusion genes based on discordant read alignments. Predicted fusions were further validated with FusionInspector in “validate” mode, which re-aligns the reads to a reference containing the genome and the fusion-gene contigs identified in the former step. Candidate fusions were annotated according to prior knowledge of fusion transcripts relevant to cancer biology (or previously observed in normal samples and thus less likely to have oncogenic potential) and assessed for the impact of the predicted fusion event on coding regions, indicating whether the fusion was in-frame or frame-shifted, along with combinations of domains expected to exist in the putative chimeric protein.

Selection of fusions with oncogenic potential. We then used FusionHub\(^{30}\) (https://fusionhub.demopersistent.com/), which provides information from 28 public fusion and gene databases, and other data bases in the literature\(^{56}\) (Fig. 2). We first eliminated fusions previously described in healthy tissue. We then identified fusions previously described in cancers, including gliomas, and looked at whether any of the genes in the fusions was known to be fused with other genes in cancers. Finally, we explored whether either gene had been identified as an oncogene or TSG or had been associated with cancer.
Next, we selected only in-frame fusions, which could produce a protein with biological effect, and manually verified the break-points with Integrative Genomics Viewer (IGV, version 2.9.4) using the reference sequence hg38\(^2\). We reviewed the exons of each gene involved in the fusion as well as the amino acids with respect to the reference sequence.

To predict the oncogenic potential of each fusion, we ran DEEPrior\(^2\) and Oncofuse\(^2\). We found that DEEPrior did not predict an oncogenic role for the known oncogenic fusion FGFR3-TACC3, which led us to choose Oncofuse (www.unav.es/genetica/oncofuse.html) for our analysis. Oncofuse provides information on the Bayesian probability of a fusion being a driver (or class 1), with a higher value indicating a higher probability, or a passenger (or class 0) by giving a Bonferroni-corrected P-value that does not take into account whether the fusion is in-frame when calculating the P-value. We set the probability of a fusion being a driver at \(P > 0.75\) and the P-value for it being a passenger at \(P < 0.05\).

These procedures provided us with a final list of fusions with probable oncogenic potential in glioblastoma and allowed us to classify them into four categories: (1) fusions that had previously been described in cancer; (2) fusions that had not been described in cancer but that involve genes previously described as oncogenes or TSGs; (3) fusions that did not meet the above conditions but had a high Oncofuse probability of having oncogenic potential; and (4) fusions that did not meet any of these conditions but produce a protein product. We then looked at the incidence of the selected fusions in our sample set. We compared the results obtained in FFPE and paired FF tissue from the same patient and compared the results found in samples obtained at initial surgery and those obtained at relapse.

Data availability
The datasets generated during the current study are available from the corresponding author on reasonable request. Molecular data underlying the findings described in the manuscript are fully available without restriction from the Bioproject Sequence Read Archive: https://www.ncbi.nlm.nih.gov/sra/PRJNA833243 and http://www.ncbi.nlm.nih.gov/bioproject/613395.

Received: 1 June 2022; Accepted: 16 August 2022
Published online: 24 August 2022

References
1. Stupp, R. et al. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. Lancet Oncol. 10, 459–466. https://doi.org/10.1016/s1470-2045(09)70025-7 (2009).
2. Wilks, P. Y. et al. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. Neuro Oncol. 22, 1073–1113. https://doi.org/10.1093/neuonc/noaa106 (2020).
3. Chinot, O. L. et al. Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. N. Engl. J. Med. 370, 709–722. https://doi.org/10.1056/NEJMoa1308345 (2014).
4. Gilbert, M. R. et al. A randomized trial of bevacizumab for newly diagnosed glioblastoma. N. Engl. J. Med. 370, 699–708. https://doi.org/10.1056/NEJMoa1308573 (2014).
5. Jia, X., Xie, Z. & Li, H. Intergenically spliced chimeric RNAs in cancer. Trends Cancer 2, 475–484. https://doi.org/10.1016/j.trecan.2016.07.006 (2016).
6. Haas, B. J. et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. Genome Biol. 20, 213. https://doi.org/10.1186/s13059-019-1842-9 (2019).
7. Mertens, F., Johannsson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. Nat Rev Cancer 15, 371–381. https://doi.org/10.1038/nrc3947 (2015).
8. Mitelman, F., Johannson, B. & Mertens, F. The impact of translocations and gene fusions on cancer causation. Nat Rev Cancer 7, 233–245. https://doi.org/10.1038/nrc2091 (2007).
9. Charest, A. et al. Fusion of FIG to the receptor tyrosine kinase ROS in a glioblastoma with an interstitial del(6)(q21q21). Genes Chromosomes Cancer 37, 58–71. https://doi.org/10.1002/gcc.10207 (2003).
10. You, G., Fan, X., Hu, H., Jiang, T. & Chen, C. C. Fusion genes altered in adult malignant gliomas. Front Neural 12, 715206. https://doi.org/10.3389/fnene.2021.715206 (2021).
11. WHO Classification of Tumours Editorial Board. Central nervous system tumours. Lyon (France): International Agency for Research on Cancer; 2021. (WHO classification of tumours series, 5th ed.; vol. 6). https://publications.iarc.fr/601 (2021).
12. Hong, D. S. et al. Larotrectinib in patients with TRK fusion-positive solid tumours: a pooled analysis of three phase 1/2 clinical trials. Lancet Oncol 21, 531–540. https://doi.org/10.1016/S1470-2045(19)30856-3 (2020).
13. Shaw, A. T., Hsu, P. P., Awad, M. M. & Engelman, J. A. Tyrosine kinase gene rearrangements in epithelial malignancies. Nat Rev Cancer 13, 772–787. https://doi.org/10.1038/nrc3612 (2013).
14. Doebele, R. C. et al. Entrectinib in patients with advanced or metastatic NTRK fusion-positive solid tumours: integrated analysis of three phase 1–2 trials. Lancet Oncol. 20, 2045–2053. https://doi.org/10.1016/j.lucancer.2019.10.006 (2019).
15. Ah-Pine, F. et al. RNA-sequencing of IDH-wild-type glioblastoma with chromothripsis identifies novel gene fusions with potential oncogenic properties. Transl Oncol 14, 100884. https://doi.org/10.1016/j.tranon.2020.100884 (2021).
16. Esteve-Codina, A. et al. A comparison of RNA-Seq results from paired formalin-fixed paraffin-embedded and fresh-frozen glioblastoma tissue samples. PLoS ONE 12, e0170632. https://doi.org/10.1371/journal.pone.0170632 (2017).
17. Wang, Q. et al. Tumor evolution of glioma-intrinsinc gene expression subtypes associates with immunological changes in the microenvironment. Cancer Cell 32, 42–56. https://doi.org/10.1016/j.ccell.2017.06.003 (2017).
18. Bowman, R. L., Wang, Q., Carro, A., Verhaak, R. G. & Squatrito, M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. Neuro Oncol 19, 139–141. https://doi.org/10.1093/neuonc/nov247 (2017).
19. Gravendeel, L. A. et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. Cancer Res 69, 9065–9072 (2009).
20. Panigrahi, P., Jere, A. & Anamika, K. FusionHub: a unified web platform for annotation and visualization of gene fusion events in human cancer. PLoS ONE 13, e0196588. https://doi.org/10.1371/journal.pone.0196588 (2018).
21. Robinson, J. T. et al. Integrative genomics viewer. Nat Biotechnol 29, 24–26. https://doi.org/10.1038/nbt.1754 (2011).
22. Shugay, M., Ortiz de Mendilib, I., Vizmanos, J. L. & Novo, F. J. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. Bioinformatics 29, 2539–2546. https://doi.org/10.1093/bioinformatics/btt445 (2013).
Acknowledgements

We thank the Fundació la Marató de TV3 for their grant, which enabled us to undertake this project, and our project coordinator, Natalia Garcia Balaña, for her support in coordinating the data, the centers, and the transfer of samples. We also thank the biobanks of the participating institutions: Fundació Institut Mar d’Investigacions Mèdiques, Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Institut d’Investigació en Patologia de la Infantiatge (IDP), Institut d’Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), and the Barcelona Institute of Biomedical Research (IDIBAPS).
Ciències de la Salut Germans Trias i Pujol (IGTP), Fundació Institut de Recerca de l’Hospital de la Santa Creu i St. Pau (IIB Sant Pau), and the Xarxa de Bancs de Tumors sponsored by Pla Director d’Oncologia de Catalunya (XBTO).

Author contributions
C.B. and A.H. conceived and designed the study. C.B., A.H., A.M.M.-M., and A.E.-C. developed the methodology. C.B., F.A., E.P., O.A.-L., M.M.-G., M.G., S.D.B., O.G., B.B., M.D., N.d.L., C.C., C.M., N.V., and I.A. acquired the data. A.E.-C., M.D., A.H., A.M.M.-M., and C.B. analyzed and interpreted the data, technical, or material support. C.B., A.H., A.E.-C., F.A., M.M., and O.A.-L. provided administrative, technical, and material support. C.B., F.A., E.P., M.M.-G., and M.G. supervised the study. C.B. and A.H. wrote the first draft of the manuscript. All authors critically reviewed and accepted the final manuscript.

Funding
This study was funded by the Fundació La Marató TV3 (665/C/2013) (http://www.ccma.cat/tv3/marato/proyecto-financiados/2012/231/). Marta Domenech is funded by ISCIII/AES (CM19/00068). Carmen Balana is funded by ISCIII/AES (PI18_01062) and ISCIII/AES (INT19/0032). Anna Esteve-Codina is funded by ISCIII/MINECO (PT17/0009/0019) and co-funded by FEDER.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-18608-8.

Correspondence and requests for materials should be addressed to C.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022