



ARTICLE

<https://doi.org/10.1038/s42003-019-0369-7>

OPEN

# The combination of neoantigen quality and T lymphocyte infiltrates identifies glioblastomas with the longest survival

Jing Zhang <sup>1</sup>, Francesca P. Caruso<sup>2,3</sup>, Jason K. Sa <sup>4</sup>, Sune Justesen<sup>5</sup>, Do-Hyun Nam<sup>4,6,7</sup>, Peter Sims<sup>8</sup>, Michele Ceccarelli<sup>2,9</sup>, Anna Lasorella<sup>1,10,11</sup> & Antonio Iavarone<sup>1,11,12</sup>

Glioblastoma (GBM) is resistant to multimodality therapeutic approaches. A high burden of tumor-specific mutant peptides (neoantigens) correlates with better survival and response to immunotherapies in selected solid tumors but how neoantigens impact clinical outcome in GBM remains unclear. Here, we exploit the similarity between tumor neoantigens and infectious disease-derived immune epitopes and apply a neoantigen fitness model for identifying high-quality neoantigens in a human pan-glioma dataset. We find that the neoantigen quality fitness model stratifies GBM patients with more favorable clinical outcome and, together with CD8<sup>+</sup> T lymphocytes tumor infiltration, identifies a GBM subgroup with the longest survival, which displays distinct genomic and transcriptomic features. Conversely, neither tumor neoantigen burden from a quantitative model nor the isolated enrichment of CD8<sup>+</sup> T lymphocytes were able to predict survival of GBM patients. This approach may guide optimal stratification of GBM patients for maximum response to immunotherapy.

<sup>1</sup>Institute for Cancer Genetics, Columbia University Medical Center, New York, NY 10032, USA. <sup>2</sup>Department of Science and Technology, Università degli Studi del Sannio, 82100 Benevento, Italy. <sup>3</sup>BIOGEM Istituto di Ricerche Genetiche 'G. Salvatore', Campo Reale, 83031 Ariano Irpino, Italy. <sup>4</sup>Institute for Refractory Cancer Research, Samsung Medical Center, Seoul, Republic of Korea. <sup>5</sup>Immunitrack Aps, Rønnegade 4, 2100 Copenhagen East, Denmark. <sup>6</sup>Department of Health Sciences and Technology, SAIHST, Sungkyunkwan University, Seoul, Republic of Korea. <sup>7</sup>Department of Neurosurgery, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Republic of Korea. <sup>8</sup>Department of Systems Biology, Columbia University Medical Center, New York, NY 10032, USA. <sup>9</sup>ABBVIE, Redwood City (CA), Redwood City, CA 94063, USA. <sup>10</sup>Department of Pediatrics, Columbia University Medical Center, New York, NY 10032, USA. <sup>11</sup>Department of Pathology and Cell Biology, Columbia University Medical Center, New York, NY 10032, USA. <sup>12</sup>Department of Neurology, Columbia University Medical Center, New York, NY 10032, USA. Correspondence and requests for materials should be addressed to M.C. (email: [m.ceccarelli@gmail.com](mailto:m.ceccarelli@gmail.com)) or to A.L. (email: [al2179@cumc.columbia.edu](mailto:al2179@cumc.columbia.edu)) or to A.I. (email: [ai2102@cumc.columbia.edu](mailto:ai2102@cumc.columbia.edu))

Recent reports have shown that nonsynonymous coding mutations may increase tumor immunogenicity. In selected tumor types such as melanoma, lung cancer, and colorectal tumors, the somatic mutational burden correlates with the probability to generate immunogenic peptides that are presented to CD8<sup>+</sup> T cells on restricted HLA-I subtypes<sup>1–4</sup>. However, in most solid tumors, T-cell immunity and a productive response to immune therapies have been reported only in a minority of patients and the identity of immunogenic tumor antigens, also referred to as neoantigens, remains unknown.

High-grade glioma is the most frequent type of primary brain tumor, with grade IV glioma (glioblastoma, GBM) being an invariably lethal tumor type with median survival below 15 months<sup>5,6</sup>. In the case of glioma higher mutational load is associated with increased tumor aggressiveness<sup>7</sup>. Consequently, the role, if any, of mutation-generated neoantigens as inducers of immunogenic responses in GBM has remained elusive. A further element that limits a productive anti-tumor immunity in GBM is the tumor microenvironment, which is dominated by myeloid-derived cells, mostly blood-derived macrophages and resident microglia<sup>8–11</sup> actively operating to exclude T lymphocytes and undermine their function<sup>12</sup>. Accordingly, GBM typically lack significant number of T lymphocyte infiltrates<sup>13</sup>. The recognized unique genetic landscape and the biological features of the GBM microenvironment led to the exclusion of high-grade glioma patients from several multi-cancer studies that have characterized tumor immunity and reinforced the notion that a lymphocyte depleted and immunosuppressive microenvironment is a distinctive feature of malignant gliomas<sup>14</sup>.

In this manuscript, we present the application of a neoantigen quality model for the accurate prediction of immunogenic neoantigens in IDH wild-type GBM, the largest and most aggressive group of high-grade gliomas. We found that in IDH wild-type GBM the production of high-quality neoantigens and infiltration of T lymphocytes are distinctive features of patients with the longest survival. The unique immunogenic attributes of this GBM subgroup informs on a cohort of patients who are optimally outfitted to mount the most effective responses following immunotherapy treatments.

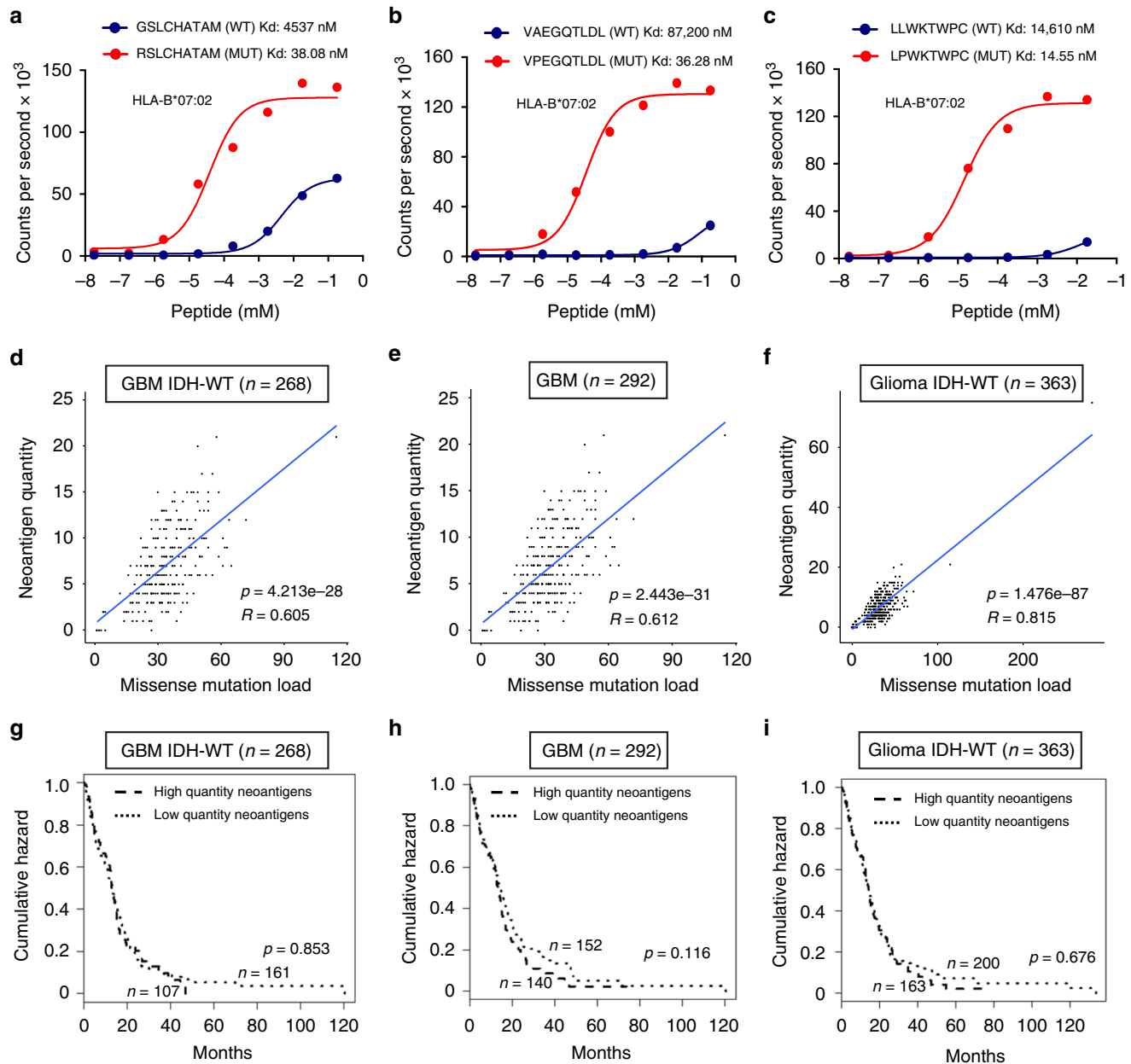
## Results

**Neoantigen quantity fails to predict survival of gliomas.** To define the importance of neoantigens in human glioma, we designed a stringent neoantigen prediction algorithm that considers the differential binding affinity of mutant and wild-type 9-mer peptides to HLA-I (neoantigen quantity model, Supplementary Fig. 1). Binding affinity was determined by netMHCpan-4.0<sup>15</sup>. Only mutant peptides with binding affinity IC<sub>50</sub> < 500 nM for the restricted HLA-I subtype and binding affinity of the corresponding wild-type peptide IC<sub>50</sub> > 500 nM for all HLA-I subtypes were retained as neoantigens (Supplementary Figs. 1–3). HLA-I subtyping for each glioma patient was done through the implementation of four different algorithms (PolySolver<sup>16</sup>, OptiType<sup>17</sup>, PHLAT<sup>18</sup>, and seq2HLA<sup>19</sup>). We applied the neoantigen prediction algorithm to the ATLAS-TCGA pan-glioma cohort that includes 303 GBM and 509 lower-grade glioma (LGG) profiled by whole exome sequence (WES), RNA-seq, and Agilent transcriptomic array. Clinical annotations were also available for this cohort (Supplementary Data 1). The pan-glioma cohort had been previously stratified on the basis of histology, genetic alterations, DNA methylation, and transcriptome clustering resulting in 19 glioma subgroups with distinct overlaps (Supplementary Data 1)<sup>20</sup>. First, we calculated the HLA-I allele frequency in each glioma subtype and found that different glioma subtypes have similar HLA-I allele frequency (Benjamini–Hochberg corrected Fisher exact test *p* value

> 0.05). We characterized neoantigens and immune landscape for each glioma subgroup, stratified tumors into high and low-neoantigen groups on the basis of the mean value of the neoantigen load and compared survival by Kaplan–Meier analysis. As experimental validation of the neoantigen prediction, we used a homogenous, proximity-based luminescent oxygen channeling immunoassay to determine the affinity kinetics of the predicted glioma neoantigens for binding to HLA-I subtypes<sup>21</sup>. This analysis including 14 matched glioma neoantigens and corresponding wild-type peptides revealed that each mutant peptide bound with higher affinity to HLA-I than the wild-type counterpart, thus validating the stringency of our approach (Fig. 1a–c, Supplementary Fig. 4). However, neoantigen load, which correlated with mutational load across glioma subtypes (Fig. 1d–f), did not distinguish patients according to clinical outcome in the cohort of GBM IDH wild-type, GBM, glioma IDH wild-type or the most aggressive form of glioma (mesenchymal and classical), but a higher neoantigen load was associated with worse prognosis in lower grade gliomas (with or without co-deletion of chromosome 1p and chromosome 19q and regardless of histology) and in glioma of the proneural and neural subtype (Fig. 1g–i, Supplementary Fig. 5). Recently, it has been proposed that the difference in binding affinity between any wild-type and mutant peptide (termed differential agretopicity index, DAI<sup>22,23</sup>) is a more accurate indicator of peptide immunogenicity than the binding affinity of the mutant peptide and it has been shown that the mean DAI of all tumor peptide pairs was a predictor of survival in melanoma and non-small cell lung cancer<sup>24</sup>. We calculated mean DAI for each glioma in the TCGA cohort and determined that similar to the results obtained from the quantity model, patients in different glioma sub-groups with high (above mean) or low (below mean) DAI had similar prognosis (Supplementary Fig. 6). In some glioma sub-types and the aggregated cohort of all gliomas, high DAI was associated with a worse clinical outcome (Supplementary Fig. 6). Together, these findings suggest that in contrast to other cancer types<sup>25,26</sup>, the neoantigen load is only a representation of the tumor mutation burden and is not associated with better survival.

## High-quality neoantigens predict better survival in GBM.

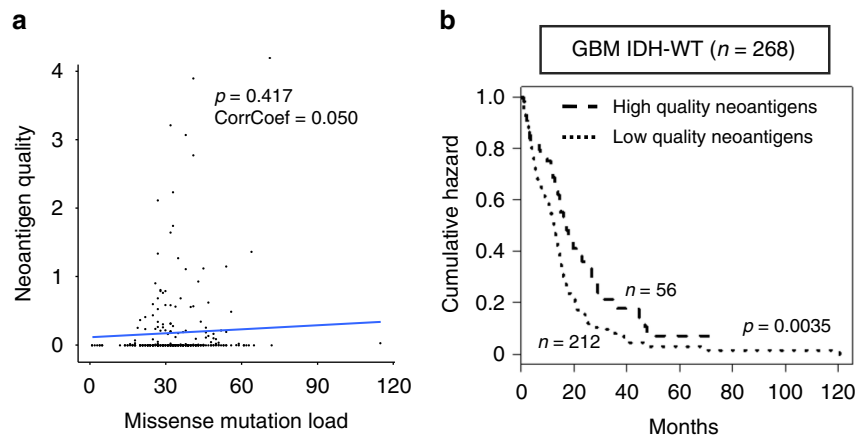
Similarity to known pathogen immunogens has been used to define neoantigen fitness (neoantigen quality model) and appears to provide a more accurate prediction of anti-tumor immunity and patients survival than the neoantigen load<sup>27</sup>. We postulated that T cell receptors (TCRs) that can recognize pathogenic antigens can also recognize similar non-pathogenic neoantigens generated by glioma cells and applied a neoantigen fitness model to glioma. In this model the candidate neoantigens are used to compute the Neoantigen Recognition Potential (NRP) which is the product of two terms (Supplementary Fig. 7 and Methods). The first term approximates the probability that a presented neoantigen will be recognized by the TCR repertoire and it depends on the similarity of the mutant peptide to human infectious disease-derived peptide sequences with positive immune assays in the Immune Epitope Database (IEDB)<sup>28</sup>. The second term, defined as the amplitude and similar to the DAI, accounts for the ratio between the binding probabilities of wild-type and mutant peptides. This model was successfully applied to predict response to checkpoint blockade immune-therapy in melanoma and lung cancer<sup>29</sup>. The first term of NRP depends on the parameters of a logistic model that were optimized for best stability and robustness across the major subtypes of high-grade gliomas (Supplementary Fig. 8). We discovered that NRP is significantly higher in IDH wild-type gliomas when compared with IDH mutant gliomas (Mann–Whitney *U* test *p*-value = 2.82e–13, Supplementary Fig. 9a).



**Fig. 1** Neantigen quantity is not prognostic of survival in glioma. **a–c** In vitro binding affinity kinetics of neoantigens and corresponding wild-type peptides for their restricted HLA class I allele. Representative results for **a**, GBM IDH wild-type; **b**, GBM; **c**, glioma IDH wild-type. Data are shown as counts per second with increasing peptide concentration ( $\log_{10}$  mM). Data are mean of  $n = 4$  technical replicates from two independent experiments for wild-type peptide;  $n = 2$  technical replicates for mutant peptide (**a, c**);  $n = 4$  technical replicates from two independent experiments for wild-type peptide;  $n = 6$  technical replicates from two independent experiments for mutant peptide (**b**). **d–f** Analysis of the correlation between neoantigen quantity and missense mutation load; **d**, GBM IDH wild-type; **e**, GBM; **f**, glioma IDH wild-type. **g–i** Stratification of survival according to neoantigen quantity score; dashed line: high-quantity neoantigens; dotted line: low-quantity neoantigens. **g**, GBM IDH wild-type; **h**, GBM; **i**, glioma IDH wild-type.  $n$  is the number of patients.  $p$ -value was determined using a log-rank test

The application of the quality model analysis, which is independent of mutation burden (Fig. 2a), to 268 IDH wild-type GBM patients, revealed that high-quality neoantigen score (above the mean values) was associated with a significantly longer survival (Fig. 2b). This effect was independent of age, gender, and mutation load (Supplementary Table 1). Similarly, stratification by the mean value of total neoantigen quantity separated the cohorts of GBM (including 292 patients) and IDH wild-type gliomas (including 363 patients) into two distinct prognostic subgroups (Supplementary Fig. 10a, b). A non-statistically significant trend to better survival for high-quality neoantigens was also

observed in classical, classic-like, and mesenchymal glioma (Supplementary Fig. 10c–e). We also determined that neoantigens with high-quality score were not restricted to specific HLA-I subtypes in IDH wild-type GBM patients (Benjamini–Hochberg corrected proportion test,  $p$  value  $> 0.05$ ). As an independent validation of the quality model, we used WES from 46 primary GBMs from a recently published cohort for which we obtained most updated survival data<sup>30</sup>. We confirmed that the 15 patients with tumors predicted to contain high-quality neoantigens had a significantly better survival (log rank  $p = 0.0339$ , Supplementary Fig. 11).



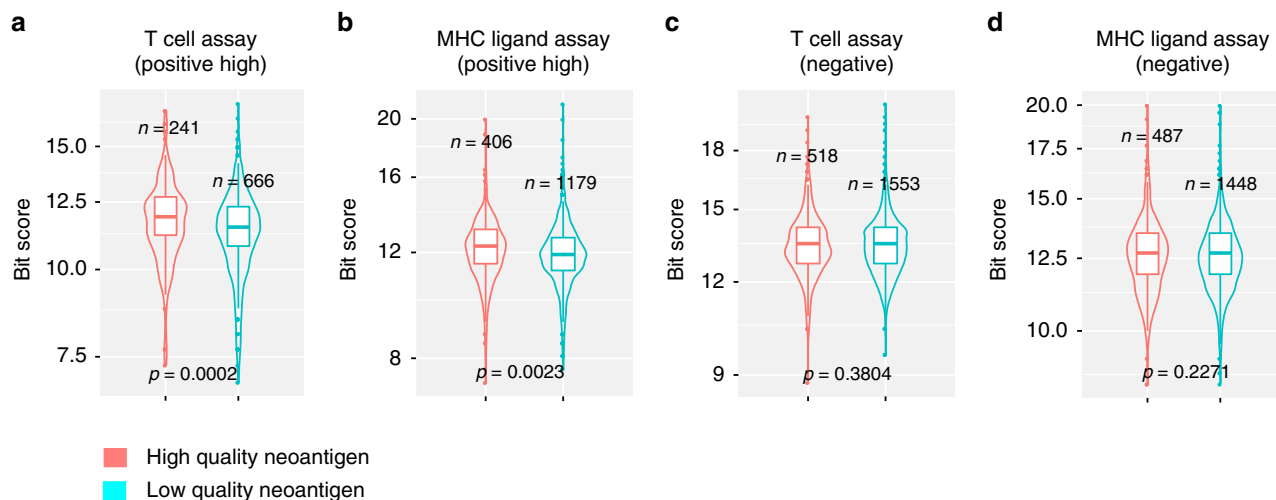
**Fig. 2** High-quality neoantigens are prognostic of better survival in IDH wild-type GBM. **a** Lack of correlation between neoantigen quality and missense mutation load ( $R = 0.050$ ,  $p = 0.417$ ). **b** Stratification of survival according to neoantigen quality score (log rank test  $p = 0.0035$ )

As IDH wild-type tumors compose the largest and most homogeneous subgroup of GBM and exhibit optimal performance in the glioma fitness model analysis, we focused our subsequent analyses on this group of tumors. HLA class I molecules are highly polymorphic with variation located in the peptide-binding region, with each variant binding only to a highly restricted set of peptide ligands<sup>31</sup>. Compared to heterozygous HLA-I carriers, homozygosity for HLA-I loci is predicted to present a smaller and less diverse repertoire of tumor-derived neoantigens to cytotoxic T lymphocytes (CTLs)<sup>32</sup>. We therefore asked whether greater diversity (heterozygosity) in the repertoire of antigen-presenting HLA-I molecules is associated with more efficient recognition of high-quality neoantigens. The analysis of the variations at each of the HLA-I genes (HLA-A, HLA-B, and HLA-C) in the high and low-quality neoantigen cohorts revealed that at least one HLA-I allele was homozygous in 25.0% (53 of 212) of IDH wild-type GBM containing only low-quality neoantigens but the frequency of HLA-I homozygosity was significantly lower in the high-quality neoantigen group (10.7% or 6 of 56, One-sided Fisher exact test  $p = 0.0136$ , Supplementary Table 2). Although not reaching statistical significance, also the amplitude term of neoantigens in HLA-I homozygosity was inferior to the amplitude term of neoantigens in HLA-I heterozygosity (median amplitude term: 1.583 and 1.785 for HLA-I homozygosity and HLA-I heterozygosity, respectively; Mann-Whitney  $U$  test  $p$  value = 0.176, Supplementary Fig. 12a). We also determined that the number of neoantigens in HLA-I homozygous IDH wild-type GBM is significantly lower than HLA-I heterozygous patients (Mann-Whitney  $U$  test  $p$  value = 0.0001, Supplementary Fig. 12b). Together, these results indicate that homozygosity for HLA-I alleles hinders the ability to generate and recognize high-quality neoantigens.

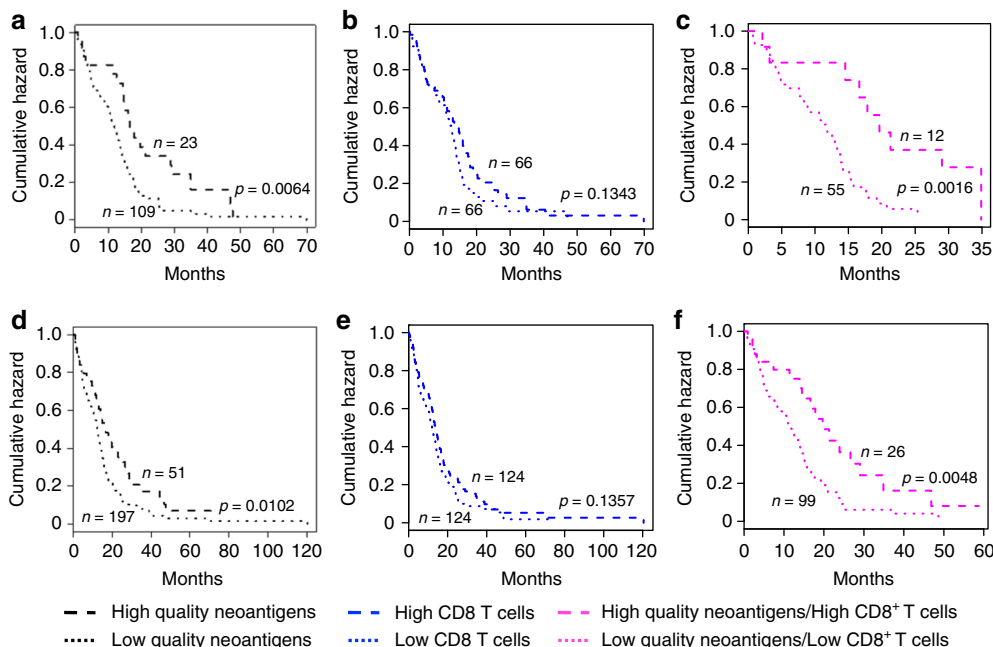
As the IEDB database also contains a collection of human immune epitopes tested with experimental assays and annotated for their ability to trigger an immune response, we sought to provide an independent validation of the quality of the computationally identified neoantigens in human gliomas. We aligned the sequences of the neoantigens detected in the high and low-quality neoantigen group of IDH wild-type GBM with experimentally validated human epitopes (allergy and autoimmune-derived) in IEDB. The analysis showed that, in comparison with low-quality neoantigens, high-quality neoantigens exhibited greater similarity to human peptides in IEDB that score as highly immunogenic (“positive high”) in T cell and MHC ligand assays ( $p = 0.0002$  and  $p = 0.0023$ , respectively; Fig. 3a, b). Conversely, we found no difference between low and high-quality neoantigens

in the alignment with human peptides in IEDB that score negative in validation immune assays ( $p = 0.3804$  and  $0.2271$ , respectively, Fig. 3c, d). Taken together, neoantigen quality rather than neoantigen load correlates with immunogenicity and predicts survival in IDH wild-type GBM.

**High-quality neoantigens and CD8<sup>+</sup> T cells identify the longest survivors.** Having established that high-quality neoantigens improve survival of IDH wild-type GBM patients, we turned our attention to the role of the non-tumor cells that compose the GBM microenvironment. In particular, we sought to determine whether the presence of any of the non-tumor cell populations (immune and non-immune) that infiltrate human GBM impact survival and/or cooperate with high-quality neoantigens to determine a more favorable outcome of GBM patients. The GBM tumor microenvironment is dominated by myeloid-derived cells that can be separated into blood-derived macrophages and microglia. We recently reported the single cell transcriptome of 8 GBM<sup>33</sup>. The analysis of cells of the tumor microenvironment from this study using single cell MWW-Gene Set Test (MWW-GST)<sup>34,35</sup> led to the identification of 6 distinct cell populations (blood-derived macrophages, microglia, CD8<sup>+</sup> T lymphocytes, oligodendrocytes, endothelial cells and pericytes). For each cell type, we selected a specific signature of 30 genes and calculated the median enrichment score in order to separate IDH wild-type GBM patients into groups with high or low infiltration of individual cell populations and asked whether the enrichment of specific cell types was associated with changes in survival. We also computed the individual signature of 12 previously identified tumor infiltrating immune cells, thus including a total of 18 cell type-specific signatures plus two signatures for the interferon gamma response pathway (Supplementary Data 2)<sup>33,36–39</sup>. The analysis was performed on 132 and 248 TCGA-derived GBM that had been analyzed by RNA-seq or Agilent expression arrays, respectively. Whereas we detected a significant effect on survival of some cell populations in one of the cohorts, none of the 20 signatures concordantly differentiated the survival of GBM patients in both cohorts (Supplementary Table 3). Infiltration of CD8<sup>+</sup> T lymphocytes that has been shown to predict a favorable prognosis in several tumor types<sup>40–43</sup> was associated only with a non-statistically significant trend to a better clinical outcome in the RNAseq and Agilent microarray cohorts ( $p = 0.134$  and  $p = 0.136$ , respectively; Fig. 4b, e). Confirming the lack of significance of CD8<sup>+</sup> T cell enrichment score for survival, we found a lower CD8<sup>+</sup> T cell enrichment score in the more favorable IDH mutant



**Fig. 3** High-quality neoantigens are associated with immunogenicity in IDH wild-type GBM. Comparison of the similarity between neoantigens and human immune epitopes scored as immunogenic (positive high) or non-immunogenic (negative) in high and low-quality neoantigen groups of IDH wild-type GBM. **a, c** T cell assays; **b, d** MHC ligand assays. *n* number of neoantigens. *p*-value was determined using two-sided Mann-Whitney *U* test



**Fig. 4** Synergistic effect of CD8<sup>+</sup> T cells and high-quality neoantigens on survival of IDH wild-type GBM. **a–c** Analysis of the cohort for which WES and RNAseq were available. **a** Survival of patients stratified according to neoantigen quality score. **b** Survival of patients stratified according to CD8<sup>+</sup> T lymphocyte enrichment score. **c** Survival of patients stratified by neoantigen quality and CD8<sup>+</sup> T lymphocytes infiltration score. **d–f** Analysis of the cohort for which WES and Agilent microarray were available. **d** Survival of patients stratified according to neoantigen quality score. **e** Survival of patient stratified according to CD8<sup>+</sup> T lymphocyte enrichment score. **f** Survival of patients stratified by neoantigen quality and CD8<sup>+</sup> T lymphocytes infiltration score. Black dashed and dotted lines represent samples with high and low-quality neoantigens, respectively. Blue dashed and dotted lines represent patients with high and low CD8<sup>+</sup> T lymphocytes, respectively. Pink dashed and dotted lines represent patients with high-quality neoantigens and high CD8<sup>+</sup> T lymphocytes and low-quality neoantigens and low CD8<sup>+</sup> T cells, respectively. *n* number of patients. *p*-value was determined by the log-rank test

when compared with IDH wild-type gliomas (Wilcoxon *p*-value 2.26E-16, Supplementary Fig. 9b). When IDH wild-type GBM patients were stratified according to the enrichment of each non-tumor cell type in combination with high or low-quality neoantigens, elevated CD8<sup>+</sup> T lymphocytes and high-quality neoantigens emerged as the only features that induced a synergistic effect on survival and classified a subgroup of approximately 10% of IDH wild-type GBM patients with the longest survival in both

RNAseq and Agilent microarray datasets ( $p = 0.0016$  and  $p = 0.0048$ , respectively; Fig. 4a–f and Supplementary Table 3). The positive role of CD8<sup>+</sup> T cells in the synergy with high-quality neoantigens was recapitulated by the trend for a similar cooperation towards a better clinical outcome between high-quality neoantigens and the interferon gamma signature, which is used as a surrogate of CD8<sup>+</sup> T cell function in transcriptomic analyses<sup>38,39,44</sup> (Supplementary Table 3). The synergistic effect of

high-quality neoantigens and CD8<sup>+</sup> T lymphocyte activation was still significant in a multivariate regression model including age, gender and mutational load as additional covariates (Supplementary Table 4). To determine the relationship between T cell infiltration, high-quality neoantigens and tumor purity, we used the tumor purity values calculated by ABSOLUTE<sup>45</sup>, a validated computational approach for the inference of the fraction of stromal/immune cells and consequently tumor cell purity. As expected, we found a negative correlation between CD8<sup>+</sup> T cell enrichment score and tumor purity (correlation coefficient =  $-0.457$ ,  $p = 3.695e-8$ , Supplementary Fig. 13a). However, there was no correlation between neoantigen quality score and tumor purity (correlation coefficient =  $0.0839$ ,  $p = 0.186$ , Supplementary Fig. 13b), indicating that high-quality neoantigens are independent of broad immune cell infiltration. Aberrant DNA methylation of genes expressed by immune cells were reported to regulate the extent of immune infiltration in solid tumors<sup>46–48</sup>. Therefore, we examined whether the enrichment of CD8<sup>+</sup> T cells in IDH wild-type GBM is associated with differential DNA methylation. Towards this aim, we performed an integrated analysis of gene expression and DNA methylation. From this analysis, diverse immune response categories emerged as significantly hypomethylated and upregulated in IDH wild-type GBM patients with high CD8<sup>+</sup> T cells for both RNAseq (Supplementary Fig. 14a, b; Supplementary Data 3, 5) and Agilent expression data (Supplementary Fig. 14c, d; Supplementary Data 4, 6). Taken together, the combination of productive neoantigen T cell recognition and epigenetically directed intra-tumor infiltration of CD8<sup>+</sup> T lymphocytes characterizes a sizable group of IDH wild-type GBM patients who experience a more favorable prognosis.

**High-quality neoantigens and CD8<sup>+</sup> T cells activate the immune response.** To identify the transcriptomic features of IDH wild-type GBM with high-quality neoantigens and high CD8<sup>+</sup> T cells, we used the combination of the easy ensemble (ee) undersampling technique and Mann–Whitney–Wilcoxon (MWW) test statistics (ee-MWW) we recently developed<sup>34</sup> and generated a ranked list of genes from RNAseq and Agilent microarray datasets of IDH wild-type GBM discriminating tumors with high-quality neoantigens/high CD8<sup>+</sup> T cells from those with low-quality neoantigens/low CD8<sup>+</sup> T cells. The gene ontology enrichment map network ( $Q < 0.00001$ , normalized enrichment score,  $NES > 0.6$ ) revealed that the most significant biological processes enriched in IDH wild-type GBM from both RNAseq and Agilent microarray datasets were immune response categories, thus providing additional evidence for the specific immune functions implemented within the tumors that contain high-quality neoantigens and high CD8<sup>+</sup> T cells (Fig. 5a, b).

**GBM with high-quality neoantigens and CD8<sup>+</sup> T cells harbor distinct genetic lesions.** Next, we sought to identify the genetic features (mutations and copy number variations, CNVs) that distinguish GBM that generate high-quality neoantigens and high CD8<sup>+</sup> T lymphocyte signatures. We failed to find recurrent genes harboring mutations that produce neoantigens in the high-quality neoantigens/high CD8<sup>+</sup> T cell group. Similarly, this group lacked specific recurrent mutations. In contrast, we found that tumors without high-quality neoantigens and CD8<sup>+</sup> T cells harbored a number of recurrently mutated genes (somatic mutations and CNVs; Fig. 5c, d), some of which are important cancer drivers. In particular, we found that genetic alterations of *PIK3CA*, *RB1* and *MDM2* were present in 24–28% of GBM unable to generate high-quality neoantigens and attract CD8<sup>+</sup> T lymphocytes but only in 0–8% of GBM with high-quality neoantigens and high CD8<sup>+</sup> T lymphocytes (RNA-seq,  $p = 0.06$ ; Agilent,  $p = 0.02$ ;

Supplementary Table 5). Together with the other genes that were exclusively mutated in the low-quality neoantigens/low CD8<sup>+</sup> T cell group, these findings point to the set of genetic determinants that should support the prospective exclusion of patients with GBM from the high-quality neoantigens/high CD8<sup>+</sup> T lymphocytes group.

## Discussion

With neoantigens emerging as attractive targets in the development of personalized immunotherapies, strategies for the rapid identification of relevant neoantigens have become a major priority<sup>24,27,29</sup>. This study describes such strategy using a computational approach for the identification of GBM patients harboring high-quality neoantigens that, together with CD8<sup>+</sup> T lymphocyte infiltrates, perform optimally in identifying patients with the longest survival and a functionally activated tumor immune microenvironment. This information might be of clinical importance for the accurate stratification of the subgroup of GBM patients having the best probability to benefit from immunotherapies. GBM are classified as lymphocyte-depleted tumors<sup>49</sup> lacking naive T cells that are instead found sequestered in large numbers in the bone marrow<sup>50</sup>. Nevertheless, recent studies based on single cell profiling have shown that a small proportion of GBM show CD8<sup>+</sup> T lymphocyte infiltrates<sup>33</sup>.

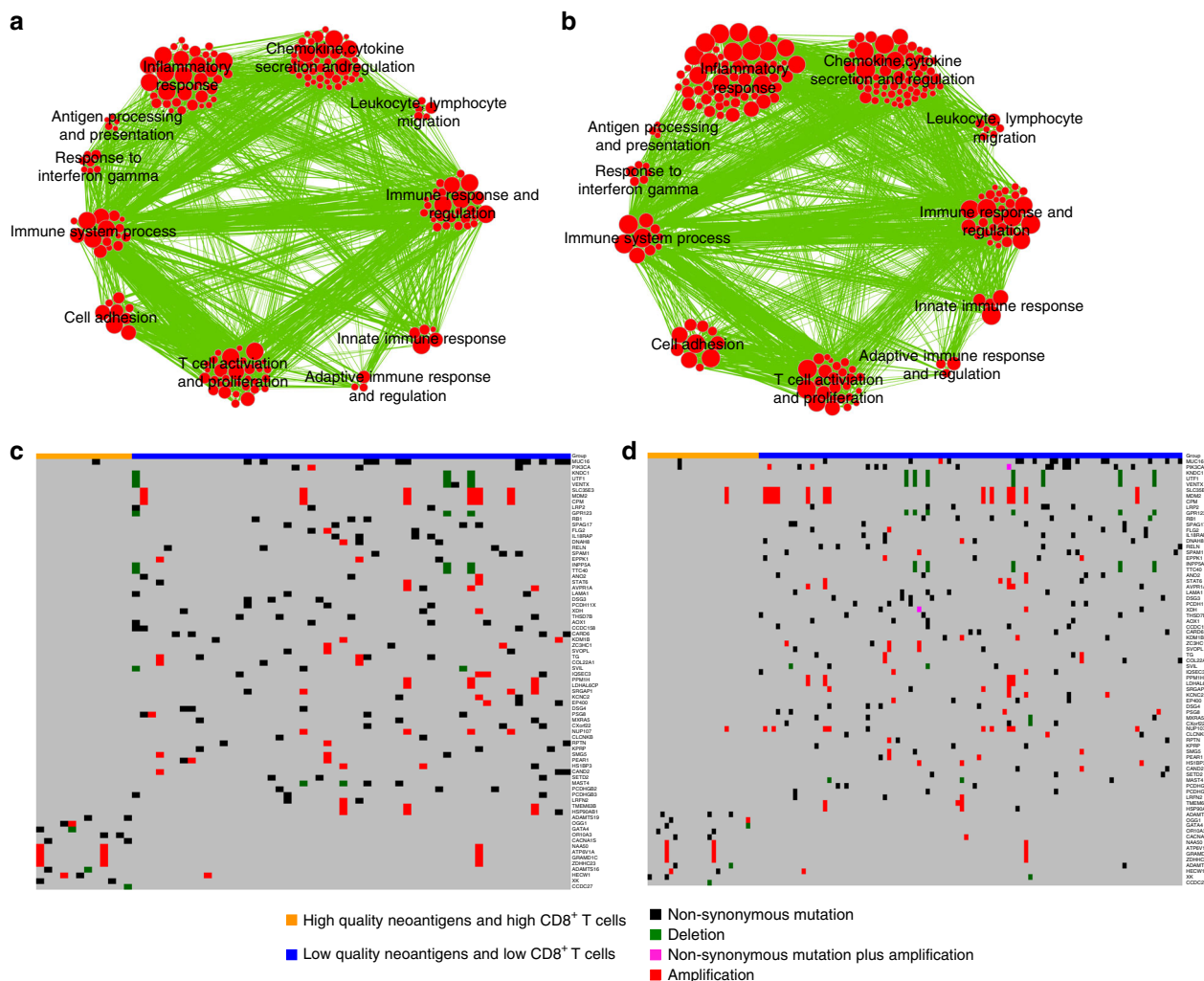
Progression from low-grade to high-grade glioma evolves through increasing mutational burden<sup>7</sup>. Even among GBM patients, a higher number of mutations is associated with a more aggressive disease and worse survival<sup>7</sup>. Therefore, whereas in several tumor types the tumor mutational burden is associated with activation of an immune response and better survival, GBM displays an opposite behavior<sup>1,3,26,51–53</sup>. Consistent with this notion, we found that the simple estimate of the mutation-derived neoantigen load using a quantity model and the DAI index failed to segregate sub-groups of patients with distinct clinical outcome. Conversely, a high-quality neoantigen model that evaluates the similarity of tumor antigens with highly immunogenic pathogen-derived antigens emerged from our work as the exclusive parameter that was able to identify patients with IDH wild-type GBM who display a more favorable clinical outcome.

Results from the previous studies that have analyzed the role of CD8<sup>+</sup> T lymphocytes for GBM survival have been conflicting<sup>54,55</sup>. Some of the studies reported that the presence of T lymphocyte infiltrates was associated with a more favorable clinical course of the disease<sup>56–58</sup>, but others reached opposite conclusions<sup>54,59</sup>. In our analysis, which is based on mRNA expression profile from GBM-derived single T lymphocytes, enrichment of the T cell-specific signature was associated with a weak positive effect on survival. However, when combined with the presence of high-quality neoantigens, CD8<sup>+</sup> T lymphocyte infiltrates provided the best predictive model for the identification of the longest survivors among IDH wild-type GBM.

We suggest that the combination of high-quality neoantigen fitness model and elevated T lymphocyte-specific gene signature together with histopathological verification of tumor infiltration by CD8<sup>+</sup> T cells should be used in current clinical trials of IDH wild-type GBM for the identification of those patients who have the highest likelihood of clinical response to immune therapy.

## Methods

**Data preparation and preprocessing.** The patient cohort is from the ATLAS-TCGA pan-glioma study<sup>20</sup>, which includes 1122 glioma patients with clinical information. For 812 glioma patients with tumor and matched normal samples 30,729 somatic mutations were called using exome-seq data bam files from TCGA Data Portal (<http://tcga-data.nci.nih.gov/tcga/>)<sup>20</sup> using at least two of three methods, MuTect<sup>60</sup>, VarScan<sup>61</sup>, and RADIA<sup>62</sup>.



**Fig. 5** Gene ontology enrichment networks and genetic characteristics of IDH wild-type GBM with high-quality neoantigens and high CD8<sup>+</sup> T cells. **a** Enrichment map network of statistically significant GO categories in the patient cohort analyzed by WES and RNAseq. **b** Enrichment map network of statistically significant GO categories in the patient cohort with WES and Agilent data available (normalized enrichment score, NES > 0.6, and  $q$ -value < 0.00001). Nodes represent GO terms and lines their connectivity. Node size is proportional to the number of genes in the GO category and line thickness indicates the fraction of genes shared between groups. **c** Landscape of somatic genomic alterations (non-synonymous mutations, copy number alterations) in IDH wild-type GBM (GBM cohort analyzed by WES and RNAseq). **d** Landscape of somatic genomic alterations (non-synonymous mutations, CNVs) in IDH wild-type GBM (GBM cohort analyzed by WES and Agilent microarrays). Rows and columns represent genes and tumor samples, respectively. Genomic alterations are indicated. Genes are sorted according to frequency (% patients) in patients having both high-quality neoantigens and high CD8<sup>+</sup> T lymphocytes or patients having both low-quality neoantigens and low CD8<sup>+</sup> T lymphocytes, respectively

RNA-seq raw counts of 667 cases (513 LGG and 154 GBM) were downloaded, normalized, and filtered using the Bioconductor package TCGAbiolinks<sup>63</sup> including TCGAquery, TCGAdownload, and TCGAprepare for level 3 data from platform “IlluminaHiSeq\_RNASeqV2”. The union of the two matrices (LGG and GBM) was then normalized using within-lane normalization to adjust for GC-content effect on read counts and upper-quantile between-lane normalization for distributional differences between lanes by applying the TCGAanalyze\_Normalization function encompassing EDASeq protocol<sup>20</sup>.

Gene expression microarray data with Agilent chip (G4502A) at level 3 were downloaded from TCGA Data Portal. The gene expression data matrix includes 583 samples (573 GBM and 10 normal brains) and 17,814 genes.

Data from 1084 glioma patients with tumor and normal samples that had been profiled on Affymetrix SNP6.0 GeneChip arrays, processed into genome segmentation files<sup>64</sup> and analyzed by GISTIC2.0 to identify focal copy number changes<sup>20,65</sup> were downloaded from TCGA Data Portal.

We used TCGAbiolinks<sup>63</sup> using TCGAquery, TCGAdownload to obtain data from 140 GBM samples that had been profiled using Illumina platform HumanMethylation450, which interrogates 485,421 CpG sites (data level = 3, platform type = “HumanMethylation450”). We removed from the analysis data point with a corresponding  $p$ -value greater than 0.01, which were deemed not to carry statistically significant difference from background and defined as “NA” in TCGA level 3 data.

**HLA-I typing.** The four-digit resolution HLA-I type of 812 patients (including LGG and GBM) with exome sequence bam files available in the TCGA cohort was determined using POLYSOLVER (POLYmorphic loci reSOLVER)<sup>16</sup>. For 647 out of 812 patients (including LGG and GBM) having RNAseq bam files available, OptiType<sup>17</sup>, PHLAT<sup>18</sup>, and seq2HLA<sup>19</sup> were further applied to infer their four-digit resolution HLA-I types. The four-digit HLA-I type was determined if the predictions were consistent in any one of the following analyses: POLYSOLVER and OptiType; POLYSOLVER and PHLAT; POLYSOLVER and seq2HLA; OptiType and PHLAT; OptiType and seq2HLA.

**Neoantigen prediction.** Missense mutations were used to generate a list of all possible 9-mer peptides. Binding affinities of mutant and corresponding wild-type 9-mer peptides, relevant to the patient’s HLA-I alleles, were predicted using netMHCpan-4.0<sup>15</sup>. High-affinity binders were defined as those with IC<sub>50</sub> equal or less than 500 nM. Low-affinity wild-type binders were defined as having IC<sub>50</sub> greater than 500 nM. More stringent criteria were used to infer neoantigens. A mutant-specific binder, relevant to the restricted HLA-I allele, was referred to as neoantigen when the mutant IC<sub>50</sub> was less than 500 nM and IC<sub>50</sub> of the corresponding wild-type binder, relevant to all HLA-I alleles of the patient, more than 500 nM. All the downstream analyses were based on the inferred neoantigens (the mutant peptides) and their corresponding wild-type peptides.

In the neoantigen fitness model, we calculated the neoantigen recognition potential (NRP) for each neoantigen using a recently developed method<sup>27,29</sup>. Briefly, each neoantigen was associated with a fitness cost designated as recognition potential, which is the likelihood that it is effectively recognized by the TCR repertoire. Given a neoantigen, the recognition potential was calculated as  $A \times R$ .  $A$ , the amplitude, is the ratio of the relative probability that a neoantigen is bound to a class I MHC times the relative probability that the wild-type counterpart of the neoantigen is not bound

$$A = \frac{1}{K_d^{MT}} \cdot \frac{K_d^{WT}/[L] + \epsilon(1 + K_d^{WT}/[L])}{1 + \epsilon(1 + K_d^{WT}/[L])} \approx \frac{K_d^{WT}}{K_d^{MT}} \cdot \frac{1}{1 + (\epsilon/[L])K_d^{WT}}$$

$\epsilon$  is a pseudo-count,  $K_d$  is the original dissociation constant, and  $[L]$  the peptide concentration<sup>29</sup>. We set  $\epsilon/[L]$  to be 0.0003<sup>29</sup>.

$R$  is the probability that a presented neoantigen will be recognized by the TCR repertoire. We estimated  $R$  using a sigmoid function applied to the score of the local alignment between the peptide sequences and the set of 2552 unique epitopes in the IEDB database<sup>29</sup>. The sigmoid function has two parameters  $a$  and  $k$ . These parameters define the shape of the sigmoid function and are optimized as explained below. In particular,  $a$  representing the horizontal displacement of the binding curve, and  $k$  is the steepness of the curve at  $a$ .  $A \times R$  represents the neoantigen recognition potential (NRP).

For each patient, NRP was first calculated for each neoantigen. The total neoantigen quality of a specific patient is equal to the mean value of NRPs of all neoantigens inferred for this patient.

**Parameter training.** To choose the optimal model parameters  $a$  and  $k$ , we generated 4000 different  $a$  and  $k$  settings with  $a$  increasing from 1 to 40 at the incremental step of 1 and  $k$  increasing from 0.1 to 10 at the incremental step of 0.1. We selected the  $a$  and  $k$  that maximize the log-rank test scores of the survival analysis of a given patient cohort.

**Leave one out cross validation (LOOCV).** Given a patient cohort of  $n$  samples, each sample was sequentially removed from the set and the remaining samples ( $n-1$ ) were used as training set on which the quality model was reoptimized. The excluded sample sequentially became the test set and was classified as high or low-quality neoantigen group. After  $n$  samples were sequentially classified, Kaplan–Meier analysis was applied to determine whether there was a statistically significant survival difference between high and low-quality neoantigen subgroups.

**Random subsampling validation.** Given a patient cohort, we randomly split samples into two groups in ratio of 4 vs. 1 in 100 runs. The larger group were used as training set and the smaller group as test set. The parameter settings ( $a, k$ ) of the quality model were trained on the training set and tested on the test set. If samples in the test sets were separated into high and low-quality neoantigens groups with significant survival difference, the parameter setting ( $a, k$ ) was successfully retained. The process was repeated 100 times to calculate the percentage of success for each parameter setting ( $a, k$ ).

To calculate the significance for the success rate of each parameter setting ( $a, k$ ), we performed 10,000 permutations of neoantigen qualities of patients across samples in the cohort under each parameter setting ( $a, k$ ) with the random split of 4 versus 1. For each parameter setting ( $a, k$ ), the success rate under each permutation was then compared with the success rate for the actual patient data.  $p$ -value was calculated as the percentage of permutations under which success rate was equal to or larger than the success rate of actual data.

**Neoantigen quantity model.** In neoantigen quantity model, the number of inferred neoantigens was counted for each patient. Patients of a given cohort were stratified according to the mean value of the number of neoantigens into high and low neoantigen quantity groups.

**Gene signatures of immune cells.** Gene signatures and relative sources of the distinct cell populations are reported in Supplementary Data 2.

**Stratification of patients based on immune cell NES and NRP.** To evaluate the enrichment of each immune cell type in TCGA glioma samples we used Normalized Enrichment Score (NES) of the MWW Gene Set test<sup>34</sup>. NES is an estimate of the probability that the expression of a gene in the gene set is greater than the expression of a gene outside this set. Specifically,

$$NES = 1 - \frac{U}{mn}$$

where  $m$  is the number of genes in a gene set,  $n$  the number of those outside the gene set,

$U = nm + \frac{m(m+1)}{2} - T$ , and  $T$  is the sum of the ranks of the genes in the gene set.

For each immune cell signature, the survival analysis was performed by dividing the patient cohort by the median value of NES score into low and high-immune

cell groups. By intersecting high and low-quality neoantigen groups with low and high-immune cell groups information, patients were stratified into four groups (high-quality neoantigens and high-immune cells; high-quality neoantigens and low-immune cells; low-quality neoantigens and high-immune cells; low-quality neoantigens and low immune cells) and Kaplan–Meier analysis was performed to evaluate their relationship with survival.

**ee-MWW based GO enrichment analysis.** To extract enriched GO categories in high-quality neoantigens and high CD8<sup>+</sup> T cells compared with low-quality neoantigens and low CD8<sup>+</sup> T cells, we used our recently developed ee-MWW method<sup>34</sup> for comparing unbalanced datasets. The group with higher number of samples (majority class) is subsampled in order to have the same size of the subgroup with less samples (minority class), this process is repeated several times ( $K = 10,000$ ) and the MWW test statistics is averaged across the samplings. In particular, the MWW gene-wise was applied to the two class of samples. For each sample subset  $k$  and the gene  $j$ , the  $U_{jk}$  value of the test statistic was retained. The value associate with each gene is the mean  $U_{jk}$  values across the  $K$  random subsets,  $\bar{U}_j = \frac{1}{K} \sum_{k=1}^K U_{jk}$ . The collection of  $\bar{U}_j$  values is a set of values that are the input of enrichment method NES<sup>34</sup>.

**Identification of differentially expressed genes.** To identify genes differentially expressed between tumors in the group with high-quality neoantigens and high CD8<sup>+</sup> T cells and those in the group with low-quality neoantigens and low CD8<sup>+</sup> T cells from RNAseq data, we performed TCGAanalyze\_DEA implementing the EdgeR protocol<sup>66</sup>. Multiple testing using the Benjamini–Hochberg procedure was applied to generate FDR. Genes with fold change >1.5 and FDR < 0.05 were considered as differentially expressed genes. To identify genes differentially expressed between tumors in the group with high-quality neoantigens and high CD8<sup>+</sup> T cells and those in the group with low-quality neoantigens and low CD8<sup>+</sup> T cells from Agilent microarray data, we used the Wilcoxon test followed by multiple testing using the Benjamini–Hochberg method for FDR estimation. Genes with fold change >1.5 and FDR < 0.05 were considered as differentially expressed genes.

**Analysis of mutations and DNA copy number changes.** To define the mutations and DNA copy number changes that differentiate tumors in the group with high-quality neoantigens and high CD8<sup>+</sup> T cells from tumors in the group with low-quality neoantigens and low CD8<sup>+</sup> T cells, we restricted the analysis to genes with no more than one alteration in the low-quality neoantigens and low CD8<sup>+</sup> T cells group. One-sided proportion test (“greater”) was adopted to identify genes with different frequencies of mutation or copy number changes between the two groups. Genes were first ordered based on the  $p$ -value of the proportion test. We then selected the minimum number of genes on the basis of increasing  $p$ -value, so that all samples were covered by at least one alteration. The same process was used to select the genes with mutation or copy number changes specifically occurring in low-quality neoantigens and low CD8<sup>+</sup> T cells group. The gene list derived from patients with both WES and RNAseq was highly overlapping with the gene list derived from patients with both WES and Agilent.

**DNA methylation analysis.** Wilcoxon test followed by multiple testing using the Benjamini–Hochberg method for FDR estimation were used to identify DNA probes differentially methylated between the high-quality neoantigens and high CD8<sup>+</sup> T cell group of glioma and the low-quality neoantigens and low CD8<sup>+</sup> T cell group of tumors. The probes with FDR < 0.05 and absolute difference in mean methylation beta-value > 0.2 were defined as differentially methylated probes. The annotations of each Illumina platform HumanMethylation450 probe were downloaded from TCGA Data Portal website.

**Integrative expression and DNA methylation analysis.** We analyzed differences in DNA methylation level between IDH wild-type GBMs groups with high CD8<sup>+</sup> T cell and low CD8<sup>+</sup> T cell. After removing probes not associated with promoters, the final methylation data matrix was composed of 42 IDH wild-type GBMs (17 high CD8<sup>+</sup> T cell, 25 low CD8<sup>+</sup> T cell samples and 85,421 probes) in the RNAseq cohort or 97 IDH wild-type GBMs with Agilent microarray data available (42 high CD8<sup>+</sup> T cell, 55 low CD8<sup>+</sup> T cell and 78,747 probes), respectively. Differentially DNA methylation analysis was then performed between samples with high CD8<sup>+</sup> T cells versus samples with low CD8<sup>+</sup> T cells using the two-sided MWW test. Differential expression analysis was performed between high versus low CD8<sup>+</sup> T cell samples using edgeR and two-sided MWW test for samples with RNAseq and Agilent microarray data, respectively. Starburst plot for comparison of DNA methylation and gene expression data was constructed using the absolute value of  $\log_{10}(p\text{-value})$  for differential DNA methylation ( $x$  axis) and gene expression ( $y$  axis) multiplied by the sign of the difference in methylation ( $x$  axis) or gene expression ( $y$  axis). A  $p$ -value less than 0.05 was considered as significant. Gene Ontology (GO) enrichment was then computed using two-sided Fisher’s exact test (FET) for a list of significant genes (hypo-methylated and upregulated genes, hyper-methylated and down-regulated genes in GBM having high CD8<sup>+</sup> T cells compared with those with low CD8<sup>+</sup> T cells). The significant GO terms from FET ( $p$ -value < 0.05;  $q$ -value < 0.25) were further analyzed using the



Enrichment Map<sup>67</sup> application of Cytoscape<sup>68</sup>. In the network, nodes represent the terms and edges represent known term interactions and are defined by the number of shared genes between the pair of terms. Size of the nodes is proportional to the number of genes in the category. The overlap between gene sets is computed according to the overlap coefficient (OC), defined as:

$$OC = \frac{|A \cap B|}{\min(|A|, |B|)}$$

where **A** and **B** are two gene sets, and  $|X|$  equals to the number of elements within set **X**. We set a cutoff of  $OC > 0.5$  to select the overlapping gene sets.

**In vitro peptide-HLA I binding assay.** Peptide-HLA class I in vitro binding affinities were determined as described previously<sup>21,69,70</sup>. Purified recombinant HLA class I heavy chains were diluted into a refolding buffer (tris-maleate buffer, pH 6.6) containing  $\beta$ 2m and serial 10-fold dilutions (0.018 nM to 180  $\mu$ M) of the test peptide, and incubated for 48 h at 18 °C to allow for equilibrium to be reached in phosphate-buffered saline (PBS). The HLA concentration was 1.25 nM, and  $\beta$ 2m concentration was 10 nM. Complex formation was detected using a proximity-based luminescence oxygen channeling immunoassay. Donor beads were obtained pre-conjugated with streptavidin from Perkin Elmer; acceptor beads were conjugated in house with W6/22, a pan-specific anti-HLA class I mouse monoclonal antibody (Sigma-Aldrich) using standard procedures as described by the manufacturer. Binding affinity (Kd) was determined as described previously<sup>21,69,70</sup> using the GraphPad Prism software 7.0. Data are means of counts per second. Amino acid abbreviations: A Ala; C Cys; D Asp; E Glu; F Phe; G Gly; H His; I Ile; K Lys; L Leu; M Met; N Asn; P Pro; Q Gln; R Arg; S Ser; T Thr; V Val; W Trp; Y Tyr.

**Quality model validation using an independent GBM dataset.** We used whole exome sequencing data, processed mutations and updated survival information of 46 primary GBMs<sup>30</sup>. HLA-I subtype for each patient was obtained by mapping raw data to human reference genome hg19 using BWA aligner and applying POLYSOLVER. Neoantigen quality model with  $a = 21$ ,  $k = 1.6$  was then applied to calculate NRP for each neoantigen.

**Statistics.** Comparisons between two groups were performed using an unpaired two-tailed Mann-Whitney *U*-test. Survival curves were compared using a log-rank test (Mantel-Cox). Categorical variables were compared using one-sided or two-sided Fisher exact test as indicated in figure legends. Multivariate survival analysis was performed using Cox regression model.

**Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

All data supporting the findings of this study are available within the published article and its supplementary information files. Figure 1 a–c have associated source data (Supplementary Table 6). All materials and other data supporting this study are available from the authors upon reasonable request.

## Code availability

We used the original codes of the authors for the computation of neoantigen quality<sup>27</sup>. A collection of the R procedures to perform ee-MWW and MWW-GST is available at <http://github.com/miccec/yaGST><sup>34</sup>.

Received: 29 October 2018 Accepted: 6 March 2019

Published online: 23 April 2019

## References

- Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
- Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
- Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
- Snyder, A. et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N. Engl. J. Med.* **371**, 2189–2199 (2014).
- Porter, K. R., McCarthy, B. J., Freels, S., Kim, Y. & Davis, F. G. Prevalence estimates for primary brain tumors in the United States by age, gender, behavior, and histology. *Neuro Oncol.* **12**, 520–527 (2010).
- Stupp, R. et al. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* **352**, 987–996 (2005).
- Draaisma, K. et al. PI3 kinase mutations and mutational load as poor prognostic markers in diffuse glioma patients. *Acta Neuropathol. Commun.* **3**, 88 (2015).
- Quail, D. F. & Joyce, J. A. The microenvironmental landscape of brain tumors. *Cancer Cell* **31**, 326–341 (2017).
- Rossi, M. L., Hughes, J. T., Esiri, M. M., Coakham, H. B. & Brownell, D. B. Immunohistological study of mononuclear cell infiltrate in malignant gliomas. *Acta Neuropathol.* **74**, 269–277 (1987).
- Morantz, R. A., Wood, G. W., Foster, M., Clark, M. & Gollahon, K. Macrophages in experimental and human brain tumors. Part 1: Studies of the macrophage content of experimental rat brain tumors of varying immunogenicity. *J. Neurosurg.* **50**, 298–304 (1979).
- Hambardzumyan, D., Gutmann, D. H. & Kettenmann, H. The role of microglia and macrophages in glioma maintenance and progression. *Nat. Neurosci.* **19**, 20–27 (2016).
- Hussain, S. F. et al. The role of human glioma-infiltrating microglia/macrophages in mediating antitumor immune responses. *Neuro Oncol.* **8**, 261–279 (2006).
- Weller, M. & Fontana, A. The failure of current immunotherapy for malignant glioma. Tumor-derived TGF- $\beta$ , T-cell apoptosis, and the immune privilege of the brain. *Brain Res. Rev.* **21**, 128–151 (1995).
- Woroniciecka, K. I., Rhodin, K. E., Chongsathidkiet, P., Keith, K. A. & Fecci, P. E. T-cell dysfunction in glioblastoma: applying a new framework. *Clin. Cancer Res.* **24**, 3792–3802 (2018).
- Jurtz, V. et al. NetMHCpan-4.0: improved peptide-MHC Class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J. Immunol.* **199**, 3360–3368 (2017).
- Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).
- Szolek, A. et al. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics* **30**, 3310–3316 (2014).
- Bai, Y., Ni, M., Cooper, B., Wei, Y. & Fury, W. Inference of high resolution HLA types using genome-wide RNA or DNA sequencing reads. *BMC Genom.* **15**, 325 (2014).
- Boegel, S. et al. HLA typing from RNA-Seq sequence reads. *Genome Med.* **4**, 102 (2012).
- Ceccarelli, M. et al. Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
- Harndahl, M. et al. Peptide binding to HLA class I molecules: homogenous, high-throughput screening, and affinity assays. *J. Biomol. Screen.* **14**, 173–180 (2009).
- Duan, F. et al. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J. Exp. Med.* **211**, 2231–2248 (2014).
- Wood, M. A. et al. Population-level distribution and putative immunogenicity of cancer neoepitopes. *BMC Cancer* **18**, 414 (2018).
- Ghorani, E. et al. Differential binding affinity of mutated peptides for MHC class I is a predictor of survival in advanced lung cancer and melanoma. *Ann. Oncol.* **29**, 271–279 (2018).
- Thomas, A. et al. Tumor mutational burden is a determinant of immune-mediated survival in breast cancer. *Oncotimmunology* **7**, e1490854 (2018).
- Samstein, R. M. et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat. Genet.* **51**, 202–206 (2019).
- Balachandran, V. P. et al. Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
- Vita, R. et al. The immune epitope database (IEDB) 3.0. *Nucleic Acids Res.* **43**, D405–D412 (2015).
- Luksza, M. et al. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. *Nature* **551**, 517–520 (2017).
- Wang, J. et al. Clonal evolution of glioblastoma under therapy. *Nat. Genet.* **48**, 768–776 (2016).
- Doherty, P. C. & Zinkernagel, R. M. A biological role for the major histocompatibility antigens. *Lancet* **1**, 1406–1409 (1975).
- Chowell, D. et al. Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* **359**, 582–587 (2018).
- Yuan, J. et al. Single-cell transcriptome analysis of lineage diversity in high-grade glioma. *Genome Med.* **10**, 57 (2018).
- Frattini, V. et al. A metabolic function of FGFR3-TACC3 gene fusions in cancer. *Nature* **553**, 222–227 (2018).
- D'Angelo, F. et al. The molecular landscape of glioma in patients with Neurofibromatosis 1. *Nat. Med.* **25**, 176–187 (2019).
- Savas, P. et al. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nat. Med.* **24**, 986–993 (2018).
- Bengsch, B. et al. Epigenomic-guided mass cytometry profiling reveals disease-specific features of exhausted CD8 T cells. *Immunity* **48**, 1029–1045 e1025 (2018).

38. Jamieson, N. B. & Maker, A. V. Gene-expression profiling to predict responsiveness to immunotherapy. *Cancer Gene Ther.* **24**, 134–140 (2017).
39. Ayers, M. et al. IFN-gamma-related mRNA profile predicts clinical response to PD-1 blockade. *J. Clin. Invest.* **127**, 2930–2940 (2017).
40. Sharma, P. et al. CD8 tumor-infiltrating lymphocytes are predictive of survival in muscle-invasive urothelial carcinoma. *Proc. Natl Acad. Sci. USA* **104**, 3967–3972 (2007).
41. Zhang, L. et al. Intratumoral T cells, recurrence, and survival in epithelial ovarian cancer. *N. Engl. J. Med.* **348**, 203–213 (2003).
42. Pages, F. et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N. Engl. J. Med.* **353**, 2654–2666 (2005).
43. Mahmoud, S. M. et al. Tumor-infiltrating CD8+ lymphocytes predict clinical outcome in breast cancer. *J. Clin. Oncol.* **29**, 1949–1955 (2011).
44. Hendrickx, W. et al. Identification of genetic determinants of breast cancer immune phenotypes by integrative genome-scale analysis. *Oncoimmunology* **6**, e1253654 (2017).
45. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).
46. Dedeurwaerder, S. et al. DNA methylation profiling reveals a predominant immune component in breast cancers. *EMBO Mol. Med.* **3**, 726–741 (2011).
47. Jeschke, J. et al. DNA methylation-based immune response signature improves patient diagnosis in multiple cancers. *J. Clin. Invest.* **127**, 3090–3102 (2017).
48. Chakravarthy, A. et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat. Commun.* **9**, 3220 (2018).
49. Thorsson, V. et al. The immune landscape of cancer. *Immunity* **48**, 812–830 e814 (2018).
50. Chongsathidkiet, P. et al. Sequestration of T cells in bone marrow in the setting of glioblastoma and other intracranial tumors. *Nat. Med.* **24**, 1459–1468 (2018).
51. Lyu, G. Y., Yeh, Y. H., Yeh, Y. C. & Wang, Y. C. Mutation load estimation model as a predictor of the response to cancer immunotherapy. *NPJ Genom. Med.* **3**, 12 (2018).
52. Rajasagi, M. et al. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* **124**, 453–462 (2014).
53. Hodges, T. R. et al. Mutational burden, immune checkpoint expression, and mismatch repair in glioma: implications for immune checkpoint immunotherapy. *Neuro Oncol.* **19**, 1047–1057 (2017).
54. Han, S. et al. Tumour-infiltrating CD4(+) and CD8(+) lymphocytes as predictors of clinical outcome in glioma. *Br. J. Cancer* **110**, 2560–2568 (2014).
55. Preusser, M., Lim, M., Hafner, D. A., Reardon, D. A. & Sampson, J. H. Prospects of immune checkpoint modulators in the treatment of glioblastoma. *Nat. Rev. Neurol.* **11**, 504–514 (2015).
56. Kmiecik, J. et al. Elevated CD3+ and CD8+ tumor-infiltrating immune cells correlate with prolonged survival in glioblastoma patients despite integrated immunosuppressive mechanisms in the tumor microenvironment and at the systemic level. *J. Neuroimmunol.* **264**, 71–83 (2013).
57. Mostafa, H. et al. Immune phenotypes predict survival in patients with glioblastoma multiforme. *J. Hematol. Oncol.* **9**, 77 (2016).
58. Yang, I. et al. CD8+ T-cell infiltrate in newly diagnosed glioblastoma is associated with long-term survival. *J. Clin. Neurosci.* **17**, 1381–1385 (2010).
59. Yue, Q. et al. The prognostic value of Foxp3 + tumor-infiltrating lymphocytes in patients with glioblastoma. *J. Neurooncol.* **116**, 251–259 (2014).
60. Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
61. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).
62. Radenbaugh, A. J. et al. RADIA: RNA and DNA integrated analysis for somatic mutation detection. *PLoS ONE* **9**, e111516 (2014).
63. Colaprico, A. et al. TCGAAbiolinks: an R/bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).
64. McCarroll, S. A. et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.* **40**, 1166–1174 (2008).
65. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
66. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
67. Isserlin, R., Merico, D., Voisin, V. & Bader, G. D. Enrichment map—a cytoscape app to visualize and explore OMICs pathway enrichment results. *F1000Res.* **3**, 141 (2014).
68. Smoot, M. E., Ono, K., Ruschekinski, J., Wang, P. L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432 (2011).
69. Braendstrup, P. et al. Identification and HLA-tetramer-validation of human CD4+ and CD8+ T cell responses against HCMV proteins IE1 and IE2. *PLoS ONE* **9**, e94892 (2014).
70. Hong, E. et al. Configuration-dependent presentation of multivalent IL-15:IL-15 ralpha enhances the antigen-specific T cell response and anti-tumor immunity. *J. Biol. Chem.* **291**, 8931–8950 (2016).

## Acknowledgements

We are grateful to Luciano Garofano and Fulvio D'Angelo for sharing computational methods. This work was supported by NIH U54CA193313 and R01CA131126 to A.L.; R01CA178546, U54CA193313, R01CA179044, R01CA190891, R01NS061776 and The Chemotherapy Foundation to A.I. and AIRC IG 2018-ID 21846 to M.C.

## Author contributions

A.I., A.L. and M.C. conceived and coordinated the studies and provided overall supervision. J.Z. developed and performed bioinformatics analyses and wrote the computational sections. F.P.C. and P.S. identified signatures for immune cells. S.J. performed in vitro peptide-HLA I binding assay. J.K.S. and D.H.N. provided WES data of GBM and clinical information. A.I., A.L., and M.C. wrote the manuscript with input from all authors.

## Additional information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s42003-019-0369-7>.

**Competing interests:** The authors declare no competing interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019