

**TITLE:** Computational characterization of suppressive immune microenvironments in glioblastoma

**Authors:** Suvi Luoto<sup>1</sup>, Ismaïl Hermelo<sup>1</sup>, Elisa M. Vuorinen<sup>1</sup>, Paavo Hannus<sup>1</sup>, Juha Kesseli<sup>1</sup>, Matti Nykter<sup>1,2</sup>, Kirsi J. Granberg<sup>1,2</sup>

1 BioMediTech Institute and Faculty of Medicine and Life Sciences, University of Tampere, Tampere, Finland. 2 Science Center, Tampere University Hospital, Tampere, Finland.

**Running title:** Computational analysis of immune responses in glioblastoma

**Abbreviations:**

CNV	Copy number variation
DAC	5-aza-2'-deoxycytidine
GBM	Glioblastoma
GEO	Gene Expression Omnibus
HLA	Human leukocyte antigen
IPA	Ingenuity Pathway Analysis
NSC	Neural stem cell
TCGA	The Cancer Genome Atlas
TIL	Tumor-infiltrating lymphocyte
Treg	Regulatory T cell

**Additional information:**

Correspondence

Matti Nykter, University of Tampere, Faculty of Medicine and Life Sciences, P.O. Box 100, 33014 University of Tampere, Finland, Tel. +358 50 318 6869, [matti.nykter@uta.fi](mailto:matti.nykter@uta.fi) and Kirsi J. Granberg, University of Tampere, Faculty of Medicine and Life Sciences, P.O. Box 100, 33014 University of Tampere, Finland, +358 50 318 5819, [kirsi.granberg@uta.fi](mailto:kirsi.granberg@uta.fi)

Disclosure of Potential Conflicts of Interest

The authors declare no potential conflicts of interest.

## **Abstract**

The immunosuppressive microenvironment in glioblastoma (GBM) prevents an efficient antitumoral immune response and enables tumor formation and growth. Although an understanding of the nature of immunosuppression is still largely lacking, it is important for successful cancer treatment through immune system modulation. To gain insight into immunosuppression in GBM, we performed a computational analysis to model relative immune cell content and type of immune response in each GBM tumor sample from The Cancer Genome Atlas RNA-seq dataset. We uncovered high variability in immune system-related responses and in the composition of the microenvironment across the cohort, suggesting immunological diversity. Immune cell compositions were associated with typical alterations such as IDH mutation or inactivating NF1 mutation/deletion. Furthermore, our analysis identified three GBM subgroups presenting different adaptive immune responses: Negative, Humoral, and Cellular-like. These subgroups were linked to transcriptional GBM subtypes and typical genetic alterations. All G-CIMP and IDH-mutated samples were in the Negative group, which was also enriched by cases with focal amplification of CDK4 and MARCH9. IDH1-mutated samples showed lower expression and higher DNA methylation of MHC-I -type HLA genes. Overall, our analysis reveals heterogeneity in the immune microenvironment of GBM and identifies new markers for immunosuppression. Characterization of diverse immune responses will facilitate patient stratification and improve personalized immunotherapy in the future.

## **Statement of Significance**

This study utilizes a computational approach to characterize the immune environments in glioblastoma and shows that glioblastoma immune microenvironments can be classified into three major subgroups, which are linked to typical glioblastoma alterations such as IDH mutation, NF1 inactivation, and CDK4-MARCH9 locus amplification.

## **Background**

Glioblastoma (GBM) is the most common malignant brain tumor in adults. Despite improvements in treatment, the median survival time of GBM patients remains approximately only 15 months after diagnosis (1,2). Poor prognosis is mostly due to the high proliferation rate,

treatment resistance against chemotherapy and all tested targeted therapies, and aggressive infiltration of the cancer cells into surrounding non-malignant brain tissue. Immunotherapies, which have delivered encouraging and long-lasting responses in many human cancers (3), have become a topic of great interest in GBM as well (4–6). Their general scope is to overcome the immunosuppression in the tumor microenvironment to activate the patient's own immune system to fight against the tumor. Immunosuppression can be generated via different mechanisms, such as regulatory T cells, checkpoint inhibitors and by secreting cytokines that inhibit the function of the effector cells (7). Understanding the mechanisms through which immunosuppression is established in GBM tumors is the key for successful personalized immunotherapies in the near future. The immune microenvironment of GBM has been characterized by the presence of specific immune cell types, but the implications of these cell types to the disease state are not well understood. Some studies associate the presence of tumor-infiltrating lymphocytes (TILs) with improved patient overall survival in GBM (8,9) while others have not observed such a correlation (10,11). Likewise, the total number of macrophages has been reported to correlate positively with patient survival (12) but opposite results have also been reported (13). Thus, the clinical relevance of TIL and macrophage infiltration remains unclear.

Both microglia and peripherally recruited macrophages can act as tumor-associated macrophages in the GBM microenvironment (14–16). Infiltration of either or both is a common feature in GBM (4,17–19), while the lymphocyte infiltration rate is generally low (19,20). The number of immune cells in the GBM microenvironment has been associated with specific alterations (21–24). For example, *IDH* mutation has recently been shown to associate with decreased immune cell infiltration (22,23), whereas inactivated *NFI* has been associated with increased macrophage infiltration (24).

We developed a computational analysis framework for modeling the GBM immune microenvironment to better understand the function and role of the immune system in GBM. Our approach builds on the regression analysis-based gene expression deconvolution that has been successfully used to estimate relative proportions of selected cell types from RNA expression data (25–27). We used regression analysis to computationally estimate the proportions of immune cell types and other normal cell components in the GBM microenvironment. The

regression analysis was combined with a data driven-analysis of immune system and immune response-related gene sets. This approach enabled us to study different prevailing immunological states in the GBM microenvironment and facilitated the analysis of both structural and functional aspects of the tumor-immune system interaction.

## **Materials and Methods**

### *RNA-seq data processing*

Raw sequencing reads from RNA-seq experiments of 156 primary GBM samples generated by the Cancer Genome Atlas (TCGA) were downloaded from the NCI Genomic Data Commons. Raw sequencing reads from RNA-seq data of normal cell types (granulocyte, macrophage M1, macrophage M2, neuron, fibroblast, CD4<sup>+</sup> T cell, CD8<sup>+</sup> T cell, endothelial cell, neutrophil, B cell and neural stem cell (NSC)) and RNA-seq of brain tissue were obtained from the NCBI Gene Expression Omnibus (GEO; Supplementary Table S1). Sequencing reads were aligned using STAR aligner version 2.4.0 (28) and Ensembl reference genome GRC37. Expression levels were quantified as RPKM based on Gencode annotations release 19 using bedtools version 2.19.0. Data were quantile normalized and log<sub>2</sub> transformed. In the case of multiple GBM samples from the same patient (patients TCGA-06-0211 and TCGA-06-0156) in the TCGA dataset, samples were combined by taking the mean. The final processed TCGA dataset consisted of 154 unique GBM samples. Likewise, the mean expressions of macrophage M1 and macrophage M2 were used as a macrophage sample and the mean expression of neutrophil and granulocyte was used as a granulocyte sample in the analysis. For validation, an independent RNA-seq dataset including of 59 primary GBM samples and two RNA-seq datasets including whole blood samples containing 5 (dataset 1) and 2 (dataset 2) samples with observed cell proportions (27), were downloaded from GEO and processed as described above.

### *Microarray analysis*

Raw microarray data from mixtures of four transformed cell lines and data from the individual cell lines were downloaded from the GEO (Accession number GSE11058). Mixtures of cell lines contain Raji (from B cells), IM-9 (from B cells), Jurkat (from T cells) and THP-1 (from monocytes) cell lines in four different known proportions. R packages ‘affy’ and ‘annotate’ were

used to process the raw data. The data were background corrected using RMA and the probe sets were annotated using the ‘hgu133plus2.db’ (version 2.1) array annotation data.

#### *Clinical data, genomic alterations, and methylation data*

Clinical data for GBM samples were downloaded from TCGA data portal. Mutation and copy number variation (CNV, determined using GISTIC 2.0), data were downloaded from cBioPortal (29,30) for typically mutated and altered genes in GBM (Supplementary Table S1). Beta values from Illumina Infinium Human DNA Methylation 450 array of 155 GBM samples generated by TCGA were downloaded from the NCI Genomic Data Commons.

#### *Statistical analyses*

Statistical analyses were performed using R version 3.2.2. All analyses testing for differences in cluster activities were performed using Wilcoxon’s rank-sum test, and all analyses testing for differences in immune cell proportions were performed using Fisher’s exact test. In Fisher’s exact test samples were divided into groups based on, e.g., alteration and subgroup, and based on whether the coefficient of a specific immune cell type was negative, zero or positive. Associations with patient survival were tested using the log-rank test.

#### *Identification of the immune system related gene clusters*

Clustering analysis was performed on gene expression data of 154 GBM samples by using the Markov Cluster Algorithm (31) and absolute values of Pearson correlation as a similarity metric. Before clustering analysis genes were filtered based on their variance and the genes with low variance, below 0.035, were omitted. Filtering resulted in expression profiles for 27172 genes. Clusters containing fewer than 10 genes were excluded from the analysis. For the remaining gene clusters, gene set enrichment for Gene Ontology categories and KEGG pathways was tested using Fisher’s exact test. Categories with p-values smaller than 0.05 were considered statistically significant. After the enrichment analysis the clusters that showed an enrichment of immune response related Gene Ontology and KEGG pathway terms were chosen for further analyses and named based on the enrichments and Ingenuity Pathway Analysis (IPA) (QIAGEN, CA, USA) results.

### *Regression analysis*

Regression analysis was performed for RNA-seq and microarray data as follows; in each sample, the expression profile of each gene was modeled as the sum of the expression profiles of that gene times the proportion of each cell type in the sample (principle outlined in Supplementary Fig. S1). For each sample separately, this can be written as an equation:

$$y = X\beta + x_{ref}\beta_{ref} + \beta_0 + \varepsilon, \text{ where } X \in \mathbb{R}^{m \times n}, \beta \in \mathbb{R}^n, y \in \mathbb{R}^m,$$

where  $X$  is the matrix of the expression levels of  $m$  genes (rows) in  $n$  cell types (columns),  $y$  is the vector of expression levels of all genes in one sample,  $\beta$  is the vector of relative levels of cell types present in the sample,  $x_{ref}$  is a median sample composed from a separate sample group,  $\beta_{ref}$  is the relative level of median sample present in the sample,  $\beta_0$  is constant and  $\varepsilon$  is the residual. We use logarithmic data for expression levels in  $X$  and  $y$ , so that the equation is a locally linear approximation of the variability in the cell type content between samples around the reference expression profile. The logarithmic data is also a better fit to our model, and involves the effects of genes with low expression in the model fit. To estimate  $\beta$ ,  $\beta_{ref}$ , and  $\beta_0$  in the equation, linear regression with elastic net regularization (32) was used, minimizing the error criterion

$$\frac{1}{2m} \left\| y - \beta_0 - X\beta - x_{ref}\beta_{ref} \right\|_2^2 + \lambda \left[ \frac{(1-\alpha) \left\| [\beta, \beta_{ref}] \right\|_2^2}{2} + \alpha \left\| [\beta, \beta_{ref}] \right\|_1 \right], \text{ where } \lambda \geq 0, 0 \leq \alpha \leq 1.$$

Elastic-net mixing parameter  $\alpha$  mixes ridge ( $\alpha = 0$ ) and lasso ( $\alpha = 1$ ) regression. The regularization parameter  $\lambda$ , which gives the most regularized model such that the error is within one standard error of the minimum, was chosen using 10-fold cross-validation, and a value of 0.5 was used for the elastic-net mixing parameter  $\alpha$ .

Code for the regression analysis is available at [http://github.com/NykterLab/GBM\\_immune](http://github.com/NykterLab/GBM_immune).

*Validation of the regression analysis using simulated and real measurement data*

For generation of validation data, RNA-seq data from four normal cell types (B cell, granulocyte, macrophage and neuron) and from a randomly chosen GBM sample (TCGA-27-1834) were used as a starting point. First, 20 different proportions of GBM sample were used from range [0, 1]. The proportions of four normal cell types were randomly chosen from the range [0, 1] so that the proportions of cell types and the GBM sample added up to one for each sample. The expression profiles were weighted with the corresponding proportions and added together. Noise was added to the simulated samples (0-100% of the variance of the data). One hundred samples were simulated using this process in each case.

Next, to test the effect of the missing reference cell type, the proportion of the GBM sample was set to 0.8 and the proportion of macrophage was set to 16 different proportions from the range [0, 0.15]. Again, the proportions of other cell types were randomly chosen from the range [0,1] so that the total proportion of all cell types and GBM sample was one for each sample and weighted expression profiles were added together. Noise was also added to these simulated samples (0-100% of the variance of the data). As a third validation dataset, microarray data from mixtures of four transformed cell lines and individual cell lines was used.

The regression analysis was performed for all simulated, cell line data, dataset 1, and dataset 2 as follows. Regression analysis was performed for all five datasets using all the genes in the identified 8 immune system related gene clusters. For the first set of simulated samples, all four normal cell types and median GBM reference were used as reference cell types. The GBM reference was composed by taking the median from the expression of 12 randomly chosen GBM samples. For sets of simulated samples with varying macrophage proportion, the regression analysis was performed without the expression profile of macrophage as reference. For mixtures of four cell lines, all cell lines were used as a reference together with a median mixture sample in the regression analysis. One replicate of each mixture was randomly chosen and the resulting four samples were used to generate the median sample, whereas the regression analysis was performed for the remaining replicates. For dataset 1 median of the B cell and T cell profiles, neutrophil, and macrophage were used as reference cell types together with the median of four whole blood samples. For dataset 2 median of the T cell profiles, B cell, and macrophage were used as reference cell types. No median sample was used due to the low number of samples.

### *Regression analysis for glioblastoma samples*

The regression analysis for the GBM samples was performed as described for 142 GBM samples not used for the GBM reference sample. Data from normal cell types (granulocyte, macrophage, neuron, fibroblast, endothelial cell, CD4<sup>+</sup> T cell, CD8<sup>+</sup> T cell, B cell, neural stem cell) and normal brain tissue were used as explanatory variables. In addition, a GBM reference was used as an explanatory variable.

### *Cluster activity*

For each identified immune related cluster, cluster activity was calculated as follows. For each gene, expression values were scaled to the interval [0,1] after truncating values below the 3rd and above the 97th percentiles to corresponding percentiles to remove outliers from the data. As some of the clusters contained a smaller subset of genes that negatively correlate with other genes in the cluster, cluster activity was calculated as a median of the majority of genes having a positive pairwise correlation with each other.

### *Clustering*

Data in heat maps were clustered using Euclidean distance and complete linkage if not otherwise specified.

A consensus clustering of samples was performed using cluster activities from all 8 immune system related clusters, with Pearson's correlation as a similarity metric and k-means clustering.

### *Demethylation experiments*

BT142mut glioma cells were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA), authenticated with targeted sequencing and grown under recommended culture conditions with regular Mycoplasma testing. Cells were seeded onto 24-well plates (120,000 cells/well) and treated with 1 $\mu$ M 5-aza-2'-deoxycytidine (abbreviated as DAC, Sigma Aldrich, St Louis, MO, USA) or corresponding DMSO control for seven days. The drug was replenished every other day.



### *Quantitative real-time PCR analysis*

Total RNA was extracted using RNeasy Mini kit (Qiagen, Hilden, Germany). HLA and GAPDH (normalization control) expression (primer sequences in Supplementary Table S1) was measured with CFX96 Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) using Maxima SYBR Green qPCR master mix (ThermoFisher Scientific, Waltham, MA, USA) without ROX.

## **Results**

### *Patterns of immune response related gene signatures in glioblastoma*

The proportion of nonmalignant cells in the TCGA GBM tumor samples can be up to 40% (33). This normal cell contamination includes stromal cells but also accumulated immune cells that are likely to induce immune response related signatures into the expression data. To extract these signatures from 154 of the TCGA RNA-seq samples generated from primary GBMs (see Methods, Supplementary Table S2), we organized all the genes into gene sets based on co-expression using the Markov Cluster Algorithm (31) and chose immune system related gene sets from them. Next, gene set enrichment analysis was used to identify 8 clusters that showed a statistical enrichment of immune responses related to Gene Ontology or KEGG pathway terms (Fig. 1A, Supplementary Table S3). These clusters, containing 17 to 1436 genes, were considered as signatures for immune system activity in the tumor microenvironment. For each cluster, cluster activity was quantified (see Methods). Different clusters had distinct cluster activity profiles across the patient cohort (Fig. 1B), revealing the heterogeneity in immune system-related responses in GBM.

We ran the IPA upstream analysis for the genes in each cluster (Supplementary Table S4) and used this information together with Gene Ontology and KEGG enrichments when naming the gene clusters. Four clusters ('Macrophages and T cell response', 'Humoral response and lymphocytes', 'Antigen presentation and interferon response', and 'Gamma delta T cells') are strongly associated with immune responses whereas three clusters ('Negative regulation of lymphocyte response', 'Leukocyte migration' and 'Leukocyte differentiation and chemotaxis') contain a smaller proportion of immune response related terms. Lower proportions of immune-

related associations may result from correlation between immune response and some other cellular processes in the tumor.

The largest cluster, called ‘Macrophages and T cell response’, contains several macrophage and lymphocyte-related genes, and the cluster has many inflammation-related upstream regulators, e.g., *IL10*, *IL6*, *STAT3*, *NFkB*, *TLR4*, and *MYD88* (Supplementary Tables S2 and S4). Several genes in this cluster, such as *FOXP3*, *IL2*, *TGFB1*, and *IL6*, are linked to the regulation and function of regulatory T (Treg) cells and Th17 cells, but the majority of lymphocyte-associated genes are general CD4<sup>+</sup> T cell genes, such as CDs and interleukin receptors. On the other hand, the ‘Humoral response and lymphocytes’ cluster was associated with many Th2- and B-cell related regulators, such as *IL4*, *CD3*, *EBF1*, *CD40*, and *CD79A*. The ‘Gamma delta T cells’ cluster includes several gamma delta T cell TCR subunits as well as other genes involved in cellular and T cell immune responses. Both clusters ‘Antigen presentation and interferon response’ and ‘Negative regulation of T-cell activation, PD-L1’ show association with interferon responses but with slight differences: the former is associated with type I interferons, e.g., different *IFNAs* and *IFNBs* as upstream regulators, whereas the latter is associated with type II interferon *IFN-* response and *IFNG* is an upstream regulator of this cluster. *IFN-* is known to induce PD-L1 expression (34,35). The ‘Negative regulation of lymphocyte response’ cluster shows positive associations with many central nervous system-related terms. It was named due to the associations with negative regulation of lymphocyte chemotaxis and with the negative regulation of antigen processing and presentation.

#### *Development and validation of the model for estimation of immune cell composition*

We developed a regression model (see Methods) that utilizes the expression of genes from the selected immune system-related clusters to understand the composition of the microenvironment in more detail. The gene expression pattern in each sample was modeled as a combination of the reference samples representing cell and tissue types that can be present in the microenvironment. For the initial method validation, we computationally mixed measurement data from immune cell types to obtain simulated datasets of varying quality (see Methods). We ran the regression analysis for these simulated datasets in two different conditions: (1) with varying proportions of GBM and (2) with the lack of one reference cell type as an explanatory component. In all the

cases, our model was able to infer the coefficients with high accuracy (Supplementary Fig. S2). Missing a reference cell type has the largest effect on the accuracy of the model (Supplementary Fig. S2).

To validate the ability of our model to estimate the relative composition of the microenvironment from real measurement data, we applied the model to expression data from mixtures of four cell lines of immune origin and for two different RNA-seq datasets from whole blood samples (see Materials and Methods). With these real data, our model was able to infer the relative cell compositions with high accuracy (Supplementary Fig. S2 and Supplementary Fig. S3).

#### *GBM samples differ in their estimated immune cell composition*

Next, we modeled the composition of the microenvironment in GBM samples. Expression profiles from 9 normal cell types and normal brain tissue were used to infer the regression coefficients using genes from immune system related clusters. The GBM reference was included in the regression model type to improve its performance. As a consequence, relative cell components should be interpreted as differences from the median GBM reference. The resulting coefficients for immune cells and all the cell types and samples are shown in Figure 2A and Supplementary Figure S4, respectively. The regression coefficients are referred to as relative immune cell components in later analyses. The obtained results reveal a high degree of diversity with interesting patterns of contributions of immune cell types to the expression profiles across the samples. The sum of our immune cell type estimates correlated with the leukocyte component estimated from DNA methylation data from same GBM tumors, and our results were also consistent with results obtained with CIBERSORT method (25) (Supplementary Fig. S5).

We wanted to determine the associations between the estimated immune cell type components and cluster activities to identify the most informative clusters in the context of immune cell type compositions. This was done using Pearson's correlation. The strongest positive correlations were observed between the activity of the 'Humoral response and lymphocytes' cluster and components of B cells and CD4<sup>+</sup> T cells (Fig. 2B). Macrophage components had the strongest association with activity of the 'Macrophage and T cell response' cluster. All the immune cell components except the CD8<sup>+</sup> T cell component had a negative correlation with the

activity of the ‘Negative regulation of lymphocyte response’ cluster. The CD8<sup>+</sup> T cell component had a negative correlation with clusters ‘Macrophage and T cell response’, ‘Leukocyte migration’ and ‘Leukocyte differentiation and chemotaxis’. As the ‘Macrophage and T cell response’ cluster was positively associated with CD4<sup>+</sup> T cell accumulation and the cluster includes several typical CD4<sup>+</sup> T cell genes, negative association with the CD8<sup>+</sup> T cell component suggests that CD4<sup>+</sup> and CD8<sup>+</sup> T cells do not really co-accumulate in the GBM microenvironment. The lack of co-accumulation of CD4<sup>+</sup> and CD8<sup>+</sup> T cells can also be observed in Figure 2A, and it can be considered one of the failures impairing a successful anti-tumoral response.

*Immune system related responses are associated with genetic alterations and patient survival*

When we compared the estimated immune cell components and cluster activities to typical genetic alterations in GBM, several associations were observed (Fig. 2C, Supplementary Fig. S6, Supplementary Table S5). Samples with an *IDH1* mutation or amplification of the genetic locus containing *CDK4* were associated with a lower macrophage component ( $p < 0.05$  and  $p < 0.01$ , respectively, Fisher’s exact test) whereas *NFI* inactivation was associated with higher macrophage component ( $p < 0.001$ , Fisher’s exact test) (Fig. 2C). Likewise, the activity of the ‘Macrophage and T cell response’ cluster was decreased in samples with *CDK4* locus amplification ( $p < 0.05$ , Wilcoxon rank-sum test) and increased in samples with *NFI* inactivation ( $p < 0.05$ , Wilcoxon rank-sum test) (Fig. 2C). *IDH1* mutated samples were also associated with a lower CD4<sup>+</sup> T cell component ( $p < 0.05$ , Fisher’s exact test) and *CDK4* locus amplified samples had a lower CD4<sup>+</sup> T cell component compared to the samples with *CDK4* locus gain ( $p < 0.05$ , Fisher’s exact test) (Fig. 2C).

The association of *CDK4* amplification with a low CD4<sup>+</sup> T cell component, low macrophage component, high activity of the ‘Negative Regulation of Lymphocyte Response’ cluster, and low activity of the ‘Macrophage and T cell response’ cluster was somewhat surprising, as *CDK4* is a known cell cycle regulator. As genes that are co-amplified with *CDK4* might generate the association, we screened the adjacent genomic neighborhood for genes with a potential to dysregulate the immune system. The adjacent *AGAP2*, *TSPAN31* and *MARCH9* genes were focally co-amplified with *CDK4* in TCGA GBM data (Supplementary Table S6). Among them,

an E3 ubiquitin-protein ligase *MARCH9* was the most prominent candidate as it is known to mediate MHC-I ubiquitination, which targets MHC-I to lysosomal degradation. This decreases MCH-I levels on the cell surface (36), leading to impaired antigen presentation by *MARCH9* overexpressing cells. The locus containing the focal amplification of *CDK4* and *MARCH9* will be referred to as *CDK4-MARCH9* locus in this article. Despite the *CDK4-MARCH9* locus amplification and *IDH* mutation being similarly associated with a shortage of immune responses, *CDK4* amplification was not associated with patient survival in the cohort.

When we analyzed how cluster activities and immune cell type components are associated with GBM transcriptional subclassification (37), several significant associations were discovered (Supplementary Fig. S7, Supplementary Fig. S8). High cluster activities were observed in mesenchymal samples; this was most evidently the case for the ‘Macrophage and T cell response’ cluster ( $p < 0.01$ , Wilcoxon rank-sum test) (Supplementary Fig. S7). Similarly, the macrophage component was also high in the mesenchymal subtype ( $p < 0.001$ , Fisher’s exact test) (Supplementary Fig. S8). Furthermore, high B cell and  $CD4^+$  T components were commonly observed in the mesenchymal samples, consistent with a previous report (21) (Supplementary Fig. S8). On the other hand, cluster activities and immune cell components tend to be low in proneural samples, except for the cluster ‘Negative regulation of lymphocyte response’, which had a significantly higher activity in proneural samples than in all other subgroups ( $p < 0.001$ , Wilcoxon’s rank-sum test) (Supplementary Fig. S7).

We performed survival analysis to determine whether the prevailing immune response affects the patient survival. None of the immune cell components or the immune cluster activities was directly associated with overall patient survival. However, among samples with a significantly higher macrophage component than our median GBM reference, high activity of the ‘Antigen presentation and interferon response’ cluster was associated with prolonged overall patient survival ( $p = 0.0038$ , log-rank test) (Fig. 2D) when median cluster activity in the analyzed cohort was used as threshold. In the same cohort, a trend towards worse survival was seen with high activity of the ‘Humoral response and lymphocytes’ cluster. The difference was statistically significant when the cluster activity value 0.060 was used as a threshold ( $p = 0.035$ , log-rank test) (Fig. 2D) instead of the median cluster activity (0.099).

### *GBM cases can be computationally grouped into three immune response related subgroups*

We performed a k-means based consensus clustering analysis for cluster activities to determine whether prevailing immune responses could identify patient subgroups (Supplementary Fig. S9). This analysis identified three sample groups that were associated with TCGA transcriptional subtypes (classification from year 2010, based on whole sample expression (37), and from year 2017, based on the expression in malignant cells (24)), G-CIMP status, genetic alterations, cluster activities, and estimated immune cell components (Fig. 3). The patterns of cluster activities and estimated immune cell type components indicate that the sample subgroups present distinct prevailing adaptive immune responses in the tumor microenvironment. Sample subgroups were named as Negative (54 samples), Humoral (14 samples), and Cellular-like (74 samples) to represent these different responses (Supplementary Table S7). As an independent validation, similar subgroups were also obtained when consensus clustering was performed for cluster activities of primary GBM samples published by Bao et al. (38) (Supplementary Fig. S10).

As illustrated in Fig. 3A, the Negative subgroup is enriched by the proneural subtype and by amplification of *CDK4-MARCH9* locus. All G-CIMP positive and *IDH1* mutated samples belong to this group. High activity of the ‘Negative regulation of lymphocyte response’ cluster is associated with the Negative group as well. The Negative group has no other positive associations to cluster activities or immune cell components (Figs. 3A-B). The Humoral subgroup is associated with higher activities of the ‘Humoral response and lymphocytes’ and ‘Macrophages and T cell response’ clusters as well as high B cell and CD4<sup>+</sup> T cell components. (Figs. 3A-B). This group consists mostly of samples of the mesenchymal subtype (Fig. 3A). The Cellular-like subgroup has a higher activity of the ‘Negative regulation of T-cell activation, PD-L1’ cluster than other subgroups, and it is also positively associated with activity of the ‘Gamma delta T cells’ cluster. The Cellular-like group is enriched by the classical subtype and by samples with a high macrophage component. It also has more samples with *EGFR* amplification than the other two groups. Interestingly, some of the alterations, such as inactivating *NF1* mutations/deletions, which were associated with estimated immune cell proportions or cluster activities (Fig. 2C, Supplementary Figs. S6-S7, Supplementary Table S5), did not show any

association with the immune subgroups, suggesting a linkage to other aspects of immune microenvironment. Altogether, the consensus clustering provided a framework for categorizing GBM tumors based on immune status in the tumor microenvironment and revealed genetic alterations that are associated with this status.

CD8<sup>+</sup> cytotoxic T cells recognize and target malignant cells via MHC-I mediated antigen presentation on the surface of the malignant cells (39). As our analysis indicates low adaptive immune response in the Negative subgroup cases, we analyzed the MHC-I expression in these tumors. All the potential MHC-I subunit human leukocyte antigens (*HLAs*), namely *HLA-A*, *HLA-B*, and *HLA-C*, had significantly lower expression in the Negative subgroup than in other subgroups (Supplementary Fig. S11). A clear difference was also observed, when *IDH1*-mutated tumors were compared with other tumors (Fig. 4A). As *IDH* mutation is associated with hypermethylator phenotype (40), we postulated that increased DNA methylation might drive decreased *HLA* gene expression in these tumors. Indeed, *HLA* genes were significantly more methylated in *IDH*-mutated than other tumors (Fig. 4B, Supplementary Table S7). *HLA* protein levels were also low in an *IDH1*-mutated glioma cell line BT142mut when compared to *IDH1* wild-type glioma cell lines and patient-derived glioblastoma cultures (Supplementary Fig. S12). Furthermore, the expression of *HLAs* was increased in methyltransferase inhibitor DAC -treated cells when compared to control-treated cells (Fig. 4C). Our data shows that in Negative group expression of the *HLA* components of MHC-I are decreased, especially in *IDH1*-mutated cells, at least partly due to DNA methylation-driven suppression of gene expression. This suggests decrease in MHC-I-mediated antigen presentation in tumors of Negative group.

## Discussion

Immune therapies have become a promising option for cancer treatment, which increases the demand for better stratification of the patients based on the function of the immune system in their tumor microenvironment. When previous computational studies have analyzed immune system function in the tumor microenvironment, they have concentrated on estimating the compositions of immune cells present in the tumor microenvironment (26,41). However, these studies typically lack information on the type of the immune response associated with the accumulation of the analyzed immune cells. We have used computational methods to identify the

immune system-related gene signatures across the dataset and estimated the immune cell compositions in the tumor microenvironment separately for each sample. To our knowledge, we are the first to have integrated the computational estimation of immune cell compositions to the data driven analysis of the immune system related signatures present in the tumor microenvironment. We have shown that the TCGA GBM cohort includes subgroups of samples with different prevailing immune responses. We have also demonstrated that the type of immune response is relevant for the disease and patient survival. These results were largely obtained due to our approach, which allows combining the information on immune cell compositions and immune signatures from the same samples. As TCGA data represents retrospective study, further prospective studies are needed to fully confirm the relevance of our findings for patient care and personalized immunotherapy.

Several studies have shown good performance and utility of computational methods in estimating the cell type proportions from the gene expression profile of cell type mixtures or tissue. Both deterministic and probabilistic modeling approaches have been successfully implemented (25–27,41). We decided to use known reference expression profiles from normal cell types instead of blind deconvolution. In this case, a straightforward way to perform the analysis is to use standard regression models. We ended up using linear regression with elastic net regularization, as the biological data is highly correlated and the elastic net can work with strongly correlated predictors, still giving a sparse solution as an output (32). This regularization helps to prevent overfitting and performs variable selection to output models that show the most significant cell types for each sample.

Recently, two publications reported that the *IDH* mutation is associated with a decreased number of immune cells in the glioma tumor microenvironment (22,23). A decreased number of immune cells might be due to the dysfunction of leukocyte migration, as genes that regulate chemotaxis are down-regulated in *IDH1* mutated tumors (22). Consistently, we showed that *IDH1* mutated GBM samples are characterized by low recruitment of CD4<sup>+</sup> T cells and macrophages, and low activity of most immune system related clusters. We also showed that MHC-I -type *HLA* genes were both less expressed and more methylated in *IDH1*-mutated tumors than in other tumors, and their expression was also increased when methyltransferases were inhibited in *IDH1*-



mutated cells (Fig. 4). This suggests that low immune response in *IDH*-mutated tumors is at least partly due to epigenetically decreased MHC-I expression and thus MHC-I -mediated antigen presentation. Also samples with amplification of the *CDK4-MARCH9* locus were associated with the Negative subgroup and they typically had fewer CD4<sup>+</sup> T cells and macrophages and lower activity of most immune cell clusters compared to the other samples in our data (Fig. 2C and 3A). Unlike in the *IDH1*-mutated tumors, the decreased immune cell proportions might be caused by downregulation or loss of MHC-I proteins from the cell surface as *MARCH9* has been reported to promote this and lysosomal destruction of MHC-I (36). *IDH* mutation is associated with better prognosis, whereas amplification of *CDK4-MARCH9* locus is not associated with prognosis in either the whole GBM cohort or within the Negative subgroup. It is likely that better survival of *IDH* mutated cases is at least partly related to other aspects of these tumors, such as a better response to radiation and chemotherapy (42,43), and not directly caused by a certain immune system status.

Consensus clustering organized the GBM samples into three subgroups (Negative, Humoral, and Cellular-like). Several features, including genetic alterations, cluster activities, estimated immune cell proportions, and DNA methylation, significantly differed between the subgroups (Table 1). The subgroups can be considered to represent different lymphocyte responses: Th2-cell mediated Humoral response, Th1-type Cellular-like response, and the absence of either response in the Negative group. Proper Th1-cell and cytotoxic T cell mediated response is needed for proper antitumoral immune response (44), and the Cellular-like subgroup can thus be considered to include cases where cellular response is partly partly – but not fully – induced. This subgroup was associated with higher activity of ‘Negative regulation of T-cell activation, PD-L1’ and ‘Gamma delta T cells’ clusters. PD-L1 is one of the checkpoint regulators that suppress T cell activation and function (45–47), and PD-1/PD-L1 interaction has been successfully targeted in several malignancies (4,48–50). None of the studied alterations was positively associated with the activity of the ‘Negative regulation of T-cell activation, PD-L1’ cluster. PD-L1 is known to be induced by *IFN-γ* (34,35) and our GBM data supports this, as the cluster also included other *IFN-γ* regulated genes in addition to PD-L1.

Among genetic alterations included in the association analysis, only *EGFR* copy number status was significantly different between the Humoral and Cellular-like subgroups, amplified cases being enriched in the latter group. This might be due partly to the low sample number in the Humoral subgroup, but the result also suggests that the most common GBM alterations are not directing the invoked lymphocyte response to either cellular or humoral types. On the other hand, the Negative subgroup was enriched by *IDH1*-mutated and *CDK4-MARCH9* locus-amplified cases. Still, the immune microenvironment in these tumors is not solely determined by the genetic makeup of malignant cells but also other factors, including the presence of the tumor *per se*, the whole immune system of the individual, and tumor-immune system interaction. Furthermore, received medication might also impact the status of the tumor microenvironment. The status of the immune microenvironment can also be dynamic: it originates at least partly from immunoediting during tumor development and is influenced by changes in the immune system caused e.g. by environmental factors and anticancer treatment. The relative significance of different factors and ways to modify the immune response should be determined in the future, e.g. by analysing the immune cells in tumor microenvironment and their interaction with malignant cells both in *ex vivo* and *in vivo* settings.

The Negative subgroup was characterized by low activity of most immune system related clusters, but, quite surprisingly, only the proportion of granulocytes and CD4<sup>+</sup> T cells was significantly lower in samples in the Negative subgroup than samples in the other two subgroups. On the other hand, high proportion of any cell type was not typical for this subgroup. Future studies are needed to evaluate the relative impact of immune cell recruitment or activation to low immune system response in this subpopulation of GBM cases.

In conclusion, we have analyzed both immune cell compositions and immune responses present in the GBM tumor microenvironments using computational approaches. By combining these two different analyses, we have identified three GBM subgroups with different prevailing immune responses. Interestingly, many of the Negative group samples contain *IDH1* mutation or *MARCH9* amplification, which may cause dysregulation of the immune system in these patients. Future studies are needed to evaluate whether modulation of MHC I -mediated antigen presentation will be a beneficial therapeutic strategy against tumors in the Negative subgroup.

## **Acknowledgments**

We acknowledge Paula Kosonen for the help with sample handling and Sven Nelander (Uppsala University), Karin Forsberg Nilsson (Uppsala University), Joonas Haapasalo (Tampere University Hospital), Hannu Haapasalo (Fimlab Laboratories Ltd), and Kristiina Nordfors (Tampere University Hospital) for the sample material. We thank CSC — IT Center for Science, Ltd. (<https://www.csc.fi/csc>) for providing the computational resources. The results published here are in part based upon data generated by The Cancer Genome Atlas project (dbGaP Study Accession: phs000178.v9.p8) established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov>.

This study was financially supported by grants from the Academy of Finland (grant 312043, 310829 to M. Nykter), Päivikki ja Sakari Sohlberg Foundation (to K.J. Granberg), Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (grants 9T042 and 9UO41 to M. Nykter), Cancer Society of Finland (to K.J. Granberg and M. Nykter), Jenny and Antti Wihuri foundation (to E.M. Vuorinen), and Pirkanmaa Cancer Society (to M. Nykter).

## **References**

1. Stupp R, Hegi ME, Mason WP, van den Bent MJ, Taphoorn MJB, Janzer RC, 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.* 2009;10:459–66.
2. Stupp R, Mason WP, van den Bent MJ. “Radiotherapy plus Concomitant and Adjuvant Temozolomide for Glioblastoma.” *Oncology Times.* 2005;27:15–6.
3. Farkona S, Diamandis EP, Blasutig IM. Cancer immunotherapy: the beginning of the end of cancer? *BMC Med.* 2016;14:73.
4. Tivnan A, Heilinger T, Lavelle EC, Prehn JHM. Advances in immunotherapy for the

- treatment of glioblastoma. *J Neurooncol.* 2017;131:1–9.
5. Kamran N, Calinescu A, Candolfi M, Chandran M, Mineharu Y, Asad AS, et al. Recent advances and future of immunotherapy for glioblastoma. *Expert Opin Biol Ther.* 2016;16:1245–64.
  6. Preusser M, Lim M, Hafler DA, Reardon DA, Sampson JH. Prospects of immune checkpoint modulators in the treatment of glioblastoma. *Nat Rev Neurol.* 2015;11:504–14.
  7. Rabinovich GA, Gabrilovich D, Sotomayor EM. Immunosuppressive Strategies that are Mediated by Tumor Cells. *Annu Rev Immunol.* 2007;25:267–96.
  8. Lohr J, Ratliff T, Huppertz A, Ge Y, Dictus C, Ahmadi R, et al. Effector T-cell infiltration positively impacts survival of glioblastoma patients and is impaired by tumor-derived TGF- $\beta$ . *Clin Cancer Res.* 2011;17:4296–308.
  9. Kmiecik J, Poli A, Brons NHC, Waha A, Eide GE, Enger PØ, 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.* 2013;264:71–83.
  10. Yeung JT, Hamilton RL, Ohnishi K, Ikeura M, Potter DM, Nikiforova MN, et al. LOH in the HLA Class I Region at 6p21 Is Associated with Shorter Survival in Newly Diagnosed Adult Glioblastoma. *Clin Cancer Res.* 2013;19:1816–26.
  11. Berghoff AS, Kiesel B, Widhalm G, Rajky O, Ricken G, Wöhrer A, et al. Programmed death ligand 1 expression and tumor-infiltrating lymphocytes in glioblastoma. *Neuro Oncol.* 2015;17:1064–75.
  12. Donson AM, Birks DK, Schittone SA, Kleinschmidt-DeMasters BK, Sun DY, Hemenway MF, et al. Increased immune gene expression and immune cell infiltration in high-grade astrocytoma distinguish long-term from short-term survivors. *J Immunol.* 2012;189:1920–7.
  13. Lu-Emerson C, Snuderl M, Kirkpatrick ND, Goveia J, Davidson C, Huang Y, et al. Increase in tumor-associated macrophages after antiangiogenic therapy is associated with poor

- survival among patients with recurrent glioblastoma. *Neuro Oncol.* 2013;15:1079–87.
14. Coniglio SJ, Segall JE. Review: molecular mechanism of microglia stimulated glioblastoma invasion. *Matrix Biol.* 2013;32:372–80.
  15. Hambardzumyan D, Gutmann DH, Kettenmann H. The role of microglia and macrophages in glioma maintenance and progression. *Nat Neurosci.* 2016;19:20–7.
  16. Poon CC, Sarkar S, Yong VW, Kelly JJP. Glioblastoma-associated microglia and macrophages: targets for therapies to improve prognosis. *Brain.* 2017;140:1548–60.
  17. Watters JJ, Schartner JM, Badie B. Microglia function in brain tumors. *J Neurosci Res.* 2005;81:447–55.
  18. Atai NA, Bansal M, Lo C, Bosman J, Tigchelaar W, Bosch KS, et al. Osteopontin is up-regulated and associated with neutrophil and macrophage infiltration in glioblastoma. *Immunology.* 2010;132:39–48.
  19. Eder K, Kalman B. The Dynamics of Interactions Among Immune and Glioblastoma Cells. *Neuromolecular Med.* 2015;17:335–52.
  20. Waziri A, Killory B, Ogden AT 3rd, Canoll P, Anderson RCE, Kent SC, et al. Preferential in situ CD4+CD56+ T cell activation and expansion within human glioblastoma. *J Immunol.* 2008;180:7673–80.
  21. Rutledge WC, Kong J, Gao J, Gutman DA, Cooper LAD, Appin C, et al. Tumor-infiltrating lymphocytes in glioblastoma are associated with specific genomic alterations and related to transcriptional class. *Clin Cancer Res.* 2013;19:4951–60.
  22. Amankulor NM, Kim Y, Arora S, Kargl J, Szulzewsky F, Hanke M, et al. Mutant IDH1 regulates the tumor-associated immune system in gliomas. *Genes Dev.* 2017;31:774–86.
  23. Berghoff AS, Kiesel B, Widhalm G, Wilhelm D, Rajky O, Kurscheid S, et al. Correlation of immune phenotype with IDH mutation in diffuse glioma. *Neuro Oncol* [Internet]. 2017; Available from: <http://dx.doi.org/10.1093/neuonc/nox054>

24. Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell*. 2017;32:42–56.e6.
25. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12:453–7.
26. Li B, Severson E, Pignon J-C, Zhao H, Li T, Novak J, et al. Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol*. 2016;17:174.
27. Racle J, de Jonge K, Baumgaertner P, Speiser DE, Gfeller D. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife*. 2017;6:e26476. Available from: <http://dx.doi.org/10.7554/eLife.26476>
28. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15–21.
29. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov*. 2012;2:401–4.
30. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal*. 2013;6:11.
31. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res*. 2002;30:1575–84.
32. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc Series B Stat Methodol*. 2005;67:301–20.
33. Brennan CW, Verhaak RGW, McKenna A, Campos B, Nounshmehr H, Salama SR, et al. The somatic genomic landscape of glioblastoma. *Cell*. 2013;155:462–77.
34. Keir ME, Butte MJ, Freeman GJ, Sharpe AH. PD-1 and Its Ligands in Tolerance and

- Immunity. *Annu Rev Immunol*. 2008;26:677–704.
35. Eppihimer MJ, Gunn J, Freeman GJ, Greenfield EA, Chernova T, Erickson J, et al. Expression and regulation of the PD-L1 immunoinhibitory molecule on microvascular endothelial cells. *Microcirculation*. 2002;9:133–45.
  36. Barte E, Mansouri M, Hovey Nerenberg BT, Gouveia K, Früh K. Downregulation of major histocompatibility complex class I by human ubiquitin ligases related to viral immune evasion proteins. *J Virol*. 2004;78:1109–20.
  37. Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17:98–110.
  38. Bao Z-S, Chen H-M, Yang M-Y, Zhang C-B, Yu K, Ye W-L, et al. RNA-seq of 272 gliomas revealed a novel, recurrent PTPRZ1-MET fusion transcript in secondary glioblastomas. *Genome Res*. 2014;24:1765–73.
  39. Garrido F, Aptsiauri N, Doorduijn EM, Garcia Lora AM, van Hall T. The urgent need to recover MHC class I in cancers for effective immunotherapy. *Curr Opin Immunol*. 2016;39:44–51.
  40. Uhm J. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Yearbook of Neurology and Neurosurgery*. 2012;2012:95–7.
  41. Varn FS, Wang Y, Mullins DW, Fiering S, Cheng C. Systematic Pan-Cancer Analysis Reveals Immune Cell Interactions in the Tumor Microenvironment. *Cancer Res*. 2017;77:1271–82.
  42. Han CH, Batchelor TT. Isocitrate dehydrogenase mutation as a therapeutic target in gliomas. *Chin Clin Oncol*. 2017;6:33.
  43. Lu Y, Kwintkiewicz J, Liu Y, Tech K, Frady LN, Su Y-T, et al. Chemosensitivity of IDH1-Mutated Gliomas Due to an Impairment in PARP1-Mediated DNA Repair. *Cancer Res*. 2017;77:1709–18.

44. Motz GT, Coukos G. Deciphering and reversing tumor immune suppression. *Immunity*. 2013;39:61–73.
45. Carter L, Fouser LA, Jussif J, Fitz L, Deng B, Wood CR, et al. PD-1:PD-L inhibitory pathway affects both CD4(+) and CD8(+) T cells and is overcome by IL-2. *Eur J Immunol*. 2002;32:634–43.
46. Freeman GJ, Long AJ, Iwai Y, Bourque K, Chernova T, Nishimura H, et al. Engagement of the PD-1 immunoinhibitory receptor by a novel B7 family member leads to negative regulation of lymphocyte activation. *J Exp Med*. 2000;192:1027–34.
47. Brown JA, Dorfman DM, Ma F-R, Sullivan EL, Munoz O, Wood CR, et al. Blockade of programmed death-1 ligands on dendritic cells enhances T cell activation and cytokine production. *J Immunol*. 2003;170:1257–66.
48. Zou W, Wolchok JD, Chen L. PD-L1 (B7-H1) and PD-1 pathway blockade for cancer therapy: Mechanisms, response biomarkers, and combinations. *Sci Transl Med*. 2016;8:328rv4.
49. Company B-MS. Opdivo (nivolumab) [prescribing information]. [cited 2017]. Available from: <http://www.opdivo.bmscustomerconnect.com/gateway>
50. Merck & Co. I. Keytruda (pembrolizumab) [prescribing information]. [cited 2017]. Available from: [http://www.merck.com/product/usa/pi\\_circulars/k/keytruda/keytruda\\_pi.pdf](http://www.merck.com/product/usa/pi_circulars/k/keytruda/keytruda_pi.pdf)



**Table 1.** Characteristic features that are enriched in Negative, Humoral or Cellular-like subgroup with different prevailing immune responses. \* Activity of all the other clusters is low in the Negative group, \*\* Activity of the cluster is high only in part of the samples.

	<b>Negative</b>	<b>Humoral</b>	<b>Cellular-like</b>
<b>Mutations</b>	<i>IDH1; ATRX</i>		
<b>Amplifications</b>	<i>CDK4-MARCH9</i> locus	EGFR gain	EGFR amplification
<b>Subtypes</b>	G-CIMP; Proneural	Mesenchymal	Classical
<b>Methylation, high</b>	<i>HLA-A; HLA-B; HLA-C</i>		
<b>Cell types, high</b>		CD4 <sup>+</sup> T cell; B cell	Macrophage
<b>Cell types, low</b>	Macrophage; Granulocyte; CD4 <sup>+</sup> T cell	CD8 <sup>+</sup> T cell	
<b>Immune signature, high</b>	Negative regulation of lymphocyte response*	Humoral response and lymphocytes; Macrophages and T cell response**	Negative regulation of T-cell activation, PD-L1; Antigen presentation and interferon response**; Gamma delta T cells**

Figure 1.

Immune response related gene clusters show variable expression patterns in glioblastoma. A Immune system related Gene ontology and KEGG pathway enrichments in identified immune response related gene clusters. Gene clusters were generated using the Markov Cluster Algorithm. Cluster sizes (number of the genes) are visualized as bar plots on the right side of the

heat map that visualizes enrichment p-values for each term. **B** Immune cluster activities vary between samples. Cluster activity for each immune system related gene cluster was determined by scaling each of the genes to range between 0-1 and then calculating the median expression of the majority of the genes with a positive pairwise correlation to each other. Hierarchical clustering with Pearson's correlation was used to organize samples and clusters in the heat map.

Figure 2.

Immune cell type compositions are associated with cluster activities and typical genetic alterations in glioblastoma. **A** Varying immune cell compositions are observed in glioblastoma. Regression analysis results for five immune cell types are shown. Coefficients are negative or positive depending on whether the estimated cell type composition is lower or higher than in the GBM reference. Compositions of selected cell types were computationally estimated by applying linear regression with elastic net regularization. **B** Compositions of analyzed immune cell types correlate (Pearson's correlation) with specific immune response related clusters. Hierarchical clustering of cell type components and clusters is based on Pearson's correlation. **C** *IDH1* mutation and amplification of the *CDK4-MARCH9* locus (genomic locus containing *CDK4* and *MARCH9* genes) are associated with a lower macrophage component whereas *NFI* inactivation is associated with a higher macrophage component (Fisher's exact test). *CDK4* amplification and *NFI* inactivation are correspondingly associated with the activity of the 'Macrophage and T cell response' cluster (Wilcoxon rank-sum test). *IDH1* mutation and *CDK4* amplification are also associated with lower CD4<sup>+</sup> T cell component (Fisher's exact test). Samples are grouped in figures based on the following genomic alterations: *IDH1* mutation, DNA copy number in *CDK4-MARCH9* locus or inactivating *NFI* mutation/deletion. The colors in the bars describe the relative proportions of CD4<sup>+</sup> T cells and macrophages in the samples in each group. The boxes visualise the activity of the cluster 'Macrophages and T cell response' in each sample group. \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$  **D** Within samples with a high macrophage component, higher activity of the 'Antigen presentation and interferon response' cluster is associated with better overall patient survival and lower activity of the 'Humoral response and lymphocytes' cluster tends to predict better overall patient survival in the same cohort (Log rank-test). Distributions of the cluster activities in the analyzed cohort are illustrated in the histograms. Thresholds used in the survival analysis (median for 'Antigen presentation and interferon response' cluster and

0.060 for ‘Humoral response and lymphocytes’ cluster) are marked with a red line to the histograms.

Figure 3.

Glioblastoma samples can be divided into three subgroups that present distinct immune responses. **A** Consensus clustering revealed three subgroups of samples which were then associated with genetic alterations, transcriptional class, G-CIMP status, and cluster activities. The Negative group (n = 54) includes all the G-CIMP positive and *IDH1* mutated samples. *CDK4* amplification, proneural subtype and higher activity of the ‘Negative regulation of lymphocyte response’ cluster is associated with the Negative group as well. The Humoral group (n = 14) is associated with higher cluster activity of the ‘Humoral response and lymphocytes’ and ‘Macrophages and T-cell response’ clusters. Most samples in the Humoral group are of the mesenchymal subtype. The Cellular-like group (n = 74) is characterized by a higher activity of the ‘Negative regulation of T-cell activation, PD-L1’ and ‘Gamma delta T cells’ clusters and by enrichment of *EGFR* amplified samples. *CDK4-MARCH9* locus: genomic locus containing *CDK4* and *MARCH9* genes. The Wilcoxon rank-sum test was used to estimate the association with immune cluster activities and Fisher’s exact test for all the other association analyses. P-values for the differences between the subgroups are visualised next to the heat map with a greyscale. **B** Tumor samples in three subgroups have different immune cell type compositions. B cell and CD4<sup>+</sup> T cell compositions are high and the CD8<sup>+</sup> T cell composition is low in the Humoral group. The Cellular-like group is enriched by samples with a high macrophage composition. Subgroups were associated with immune cell proportions using Fisher’s exact test. \* p < 0.05 \*\* p < 0.01, \*\*\* p < 0.001

Figure 4.

DNA methylation suppresses *HLA* gene expression in *IDH1* mutated samples. **A** In TCGA GBM cohort, *IDH1* mutated samples and *IDH1* wild-type samples with *CDK4-MARCH9* locus amplification (*CDK4-MARCH9* amp) have lower *HLA* gene expression than other *IDH1* wild-type samples (Wilcoxon rank-sum test). *HLA-A*, *HLA-B*, and *HLA-C* can each act as a MHC-I subunit. **B** Methylation level of *HLA* genes is higher in *IDH1* mutated samples compared to *IDH1* wild type samples (Wilcoxon rank-sum test). Representative probes (cg17608381,

cg13902357, cg15397231) are shown in the figure. **C** Methyltransferase inhibition and resulting demethylation results in increased RNA expression of *HLA* genes in an *IDH*-mutant glioma cell line BT142mut. Methyltransferases were inhibited with 5-aza-2'-deoxycytidine (DAC). Fold change of expression values with standard deviation are visualized in the figure. \*  $p < 0.05$  \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

Figure 1

A

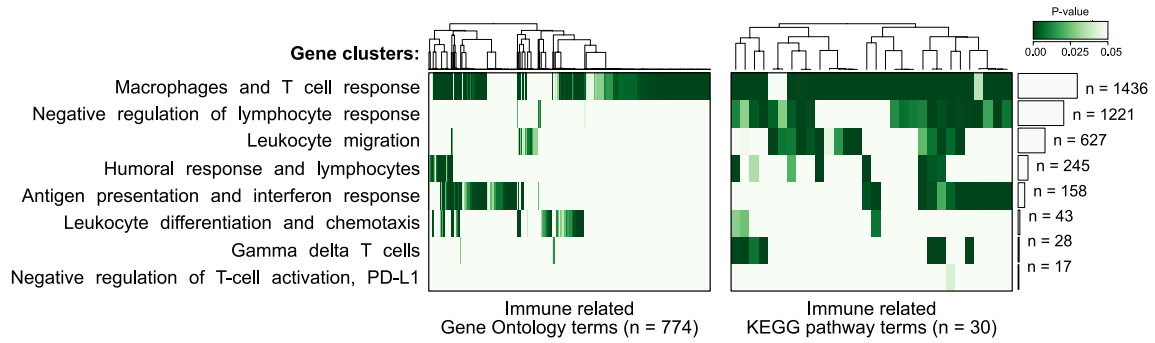


Figure 1

B

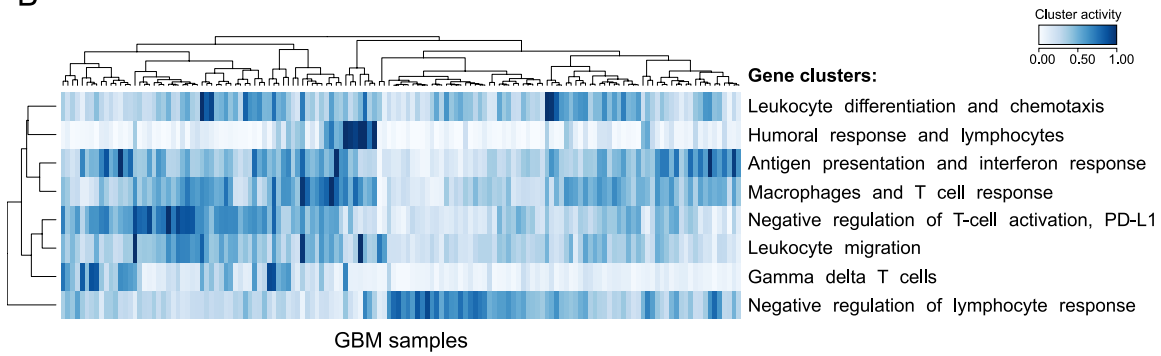


Figure 2

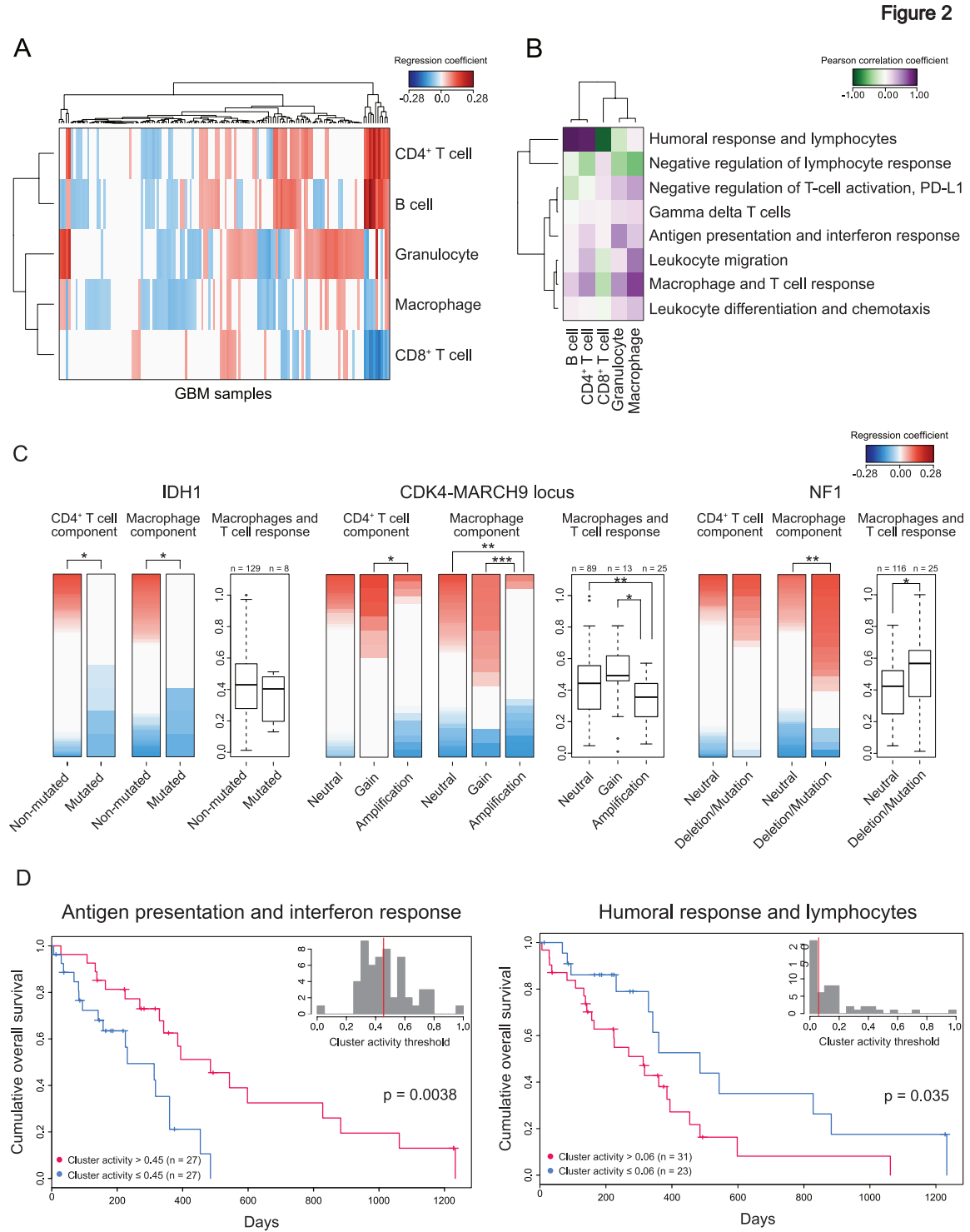


Figure 3

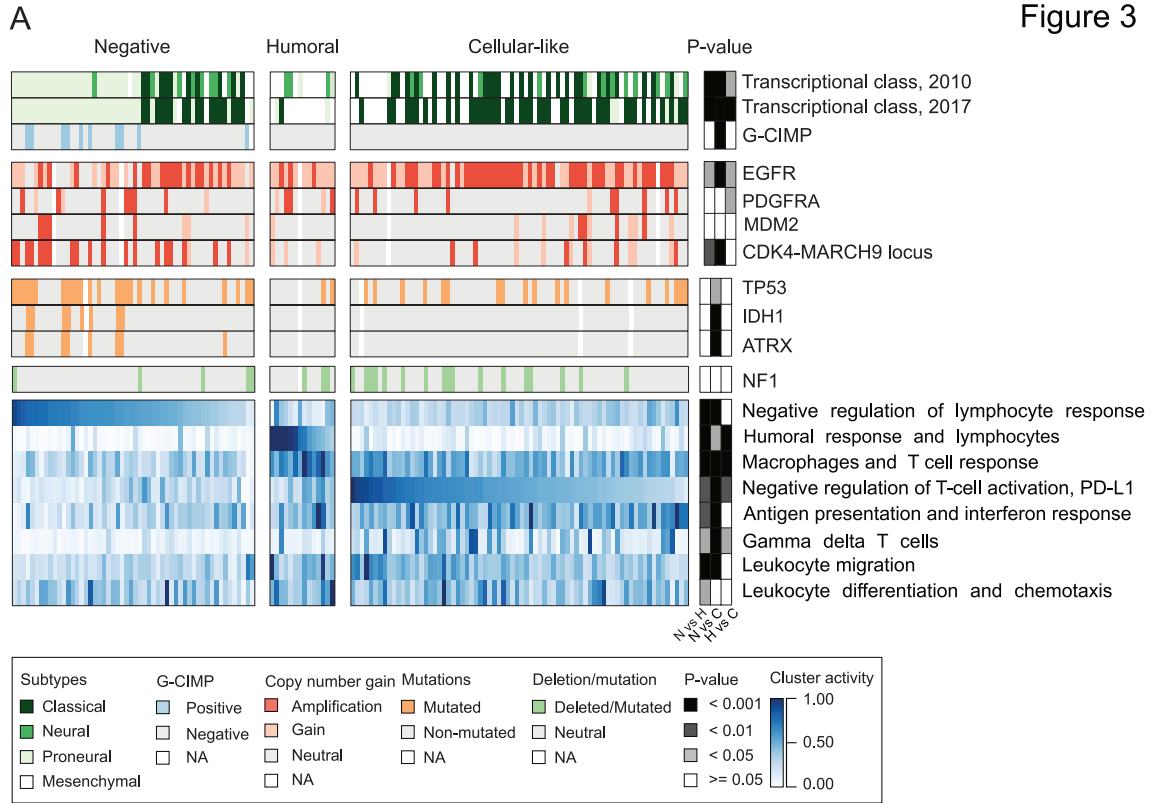


Figure 3

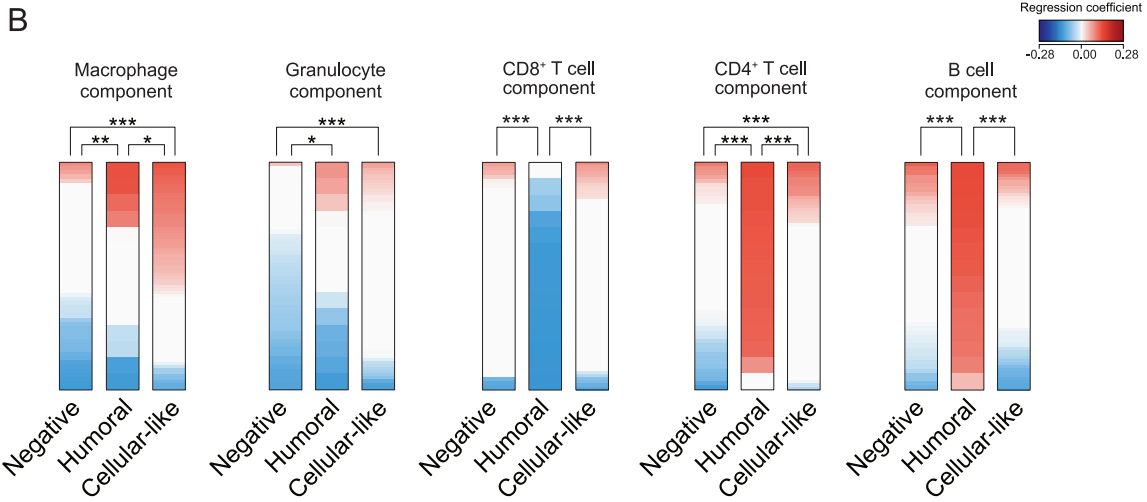
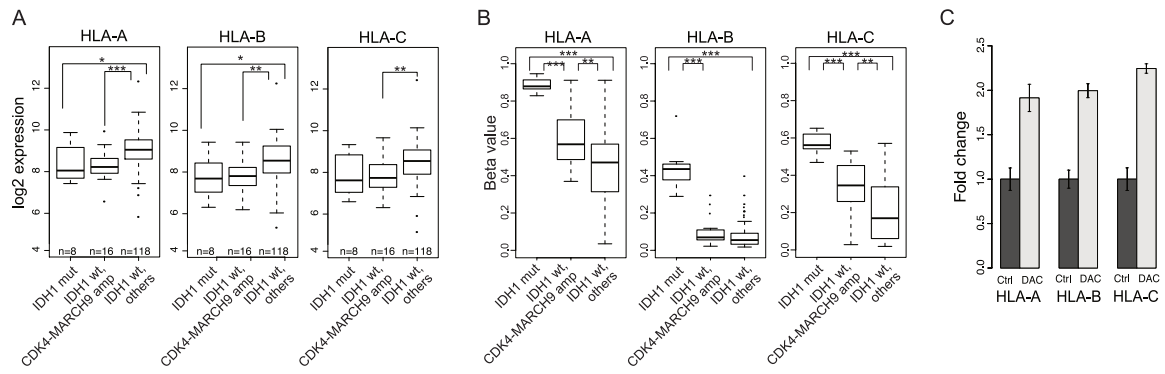
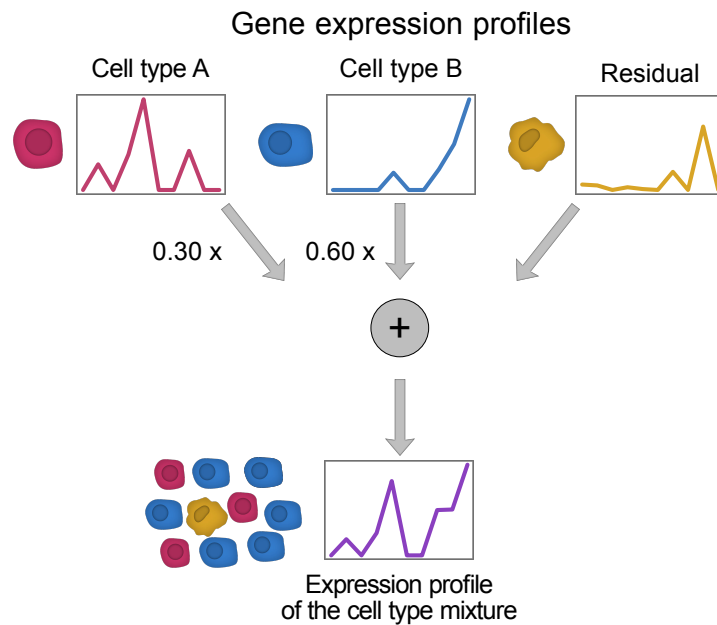


Figure 4



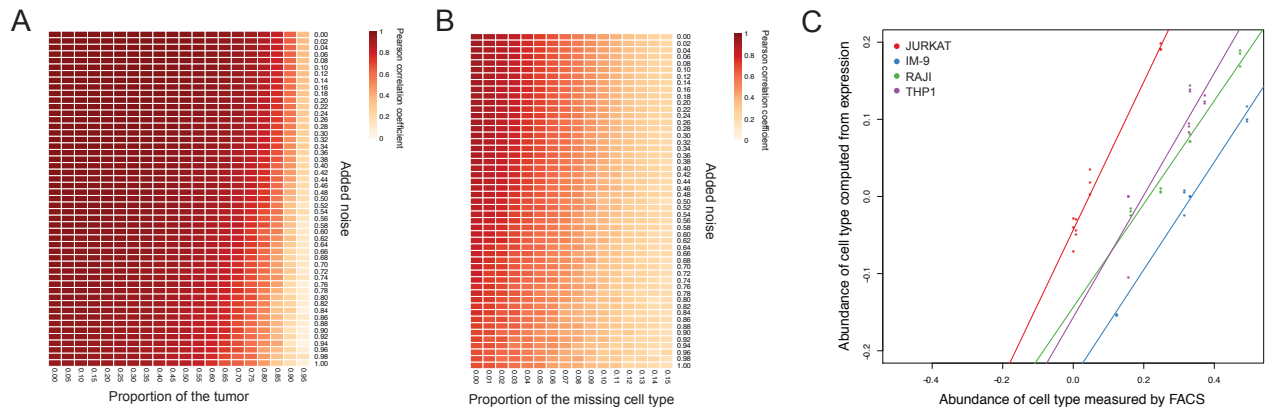


**Supplementary Figure S1: Estimation of cell type proportions from gene expression data utilizing reference cell types and linear regression.**



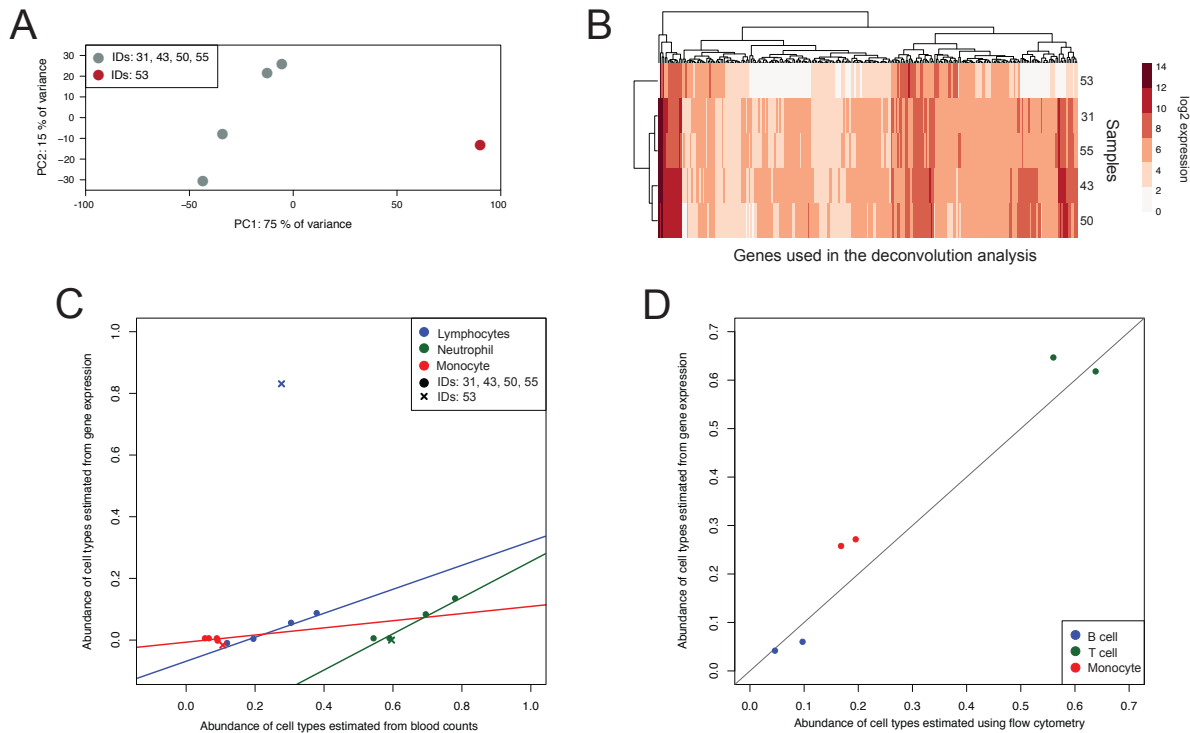
**Figure S1: Estimation of cell type proportions from gene expression data utilizing reference cell types and linear regression.** It can be thought that cell type mixture expression profile ( $y$  in the equations) is constructed from so-called reference expression profiles and residual. In this example, these reference profiles are from cell types A and B ( $X$  in the equations), and those profiles are mixed in proportions of 0.30 and 0.60 ( $\beta$  in the equations), respectively. The residual is the part of the expression profile that we can't model using the reference profiles. When these reference profiles are multiplied with the proportions and added together with the residual, we get the cell type mixture expression profile. In the deconvolution analysis, we know the reference cell type profiles and the cell type mixture profile. What we don't know, is the proportions in which the reference profiles must be added to end up with the cell type mixture profile. Those proportions (i.e. regression coefficients) can be inferred using linear regression.

**Supplementary Figure S2: Validation of the regression model using simulated and real measurement data.**



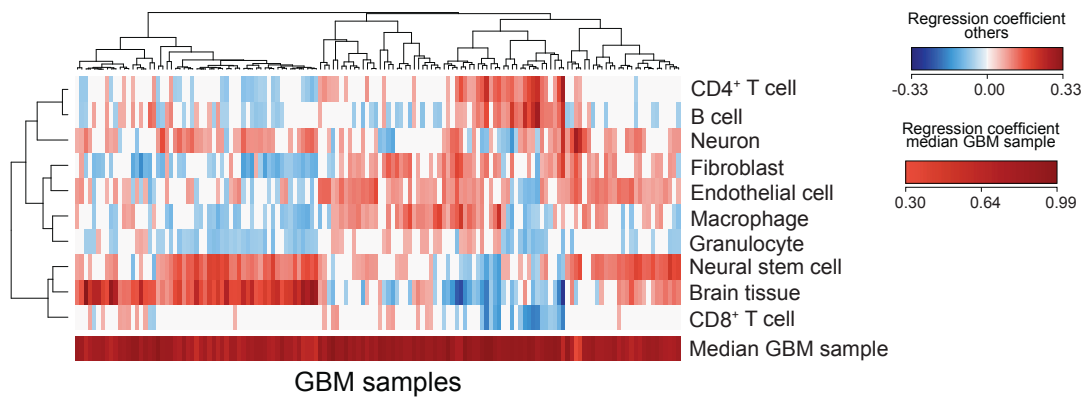
**Figure S2: Validation of the regression model using simulated and real measurement data.** **a** Validation of the regression model using simulated data with varying proportion of added noise and varying tumor proportion. One hundred samples were simulated using the RNA-seq data from four normal cell types (B cell, granulocyte, macrophage and neuron) and randomly chosen GBM sample. First, 20 different proportion of GBM sample was used from range [0, 1]. The proportions of four cell types were randomly chosen from range [0, 1] so that the proportions of cell types and the GBM sample added up to one for each sample. The expression profiles were weighted with the corresponding proportions and added together. Noise was added to the simulated samples (0-100% of the variance of the data). **b** Validation of the regression model with varying proportion of missing reference cell type using simulated data. One hundred samples were simulated with varying macrophage proportion. The proportion of the GBM sample was set to 0.8 and the proportion of macrophage was set to 16 different proportions from the range [0, 0.16]. Again, the proportions of other cell types were randomly chosen from the range [0, 1] so that the total proportion of all cell types and GBM sample was one for each sample and weighted expression profiles were added together. Noise was also added to these simulated samples (0-100% of the variance of the data). **c** Validation of the regression model using microarray data from mixtures of four transformed cell lines (Raji, IM-9, Jurkat, THP-1). The four cell lines together with a median sample were used in the regression analysis. (Pearson’s correlations for Jurkat, IM-9, Raji and THP-1, respectively: 0.989, 0.993, 0.987, 0.902)

**Supplementary Figure S3: Validation of the regression model using whole blood samples.**



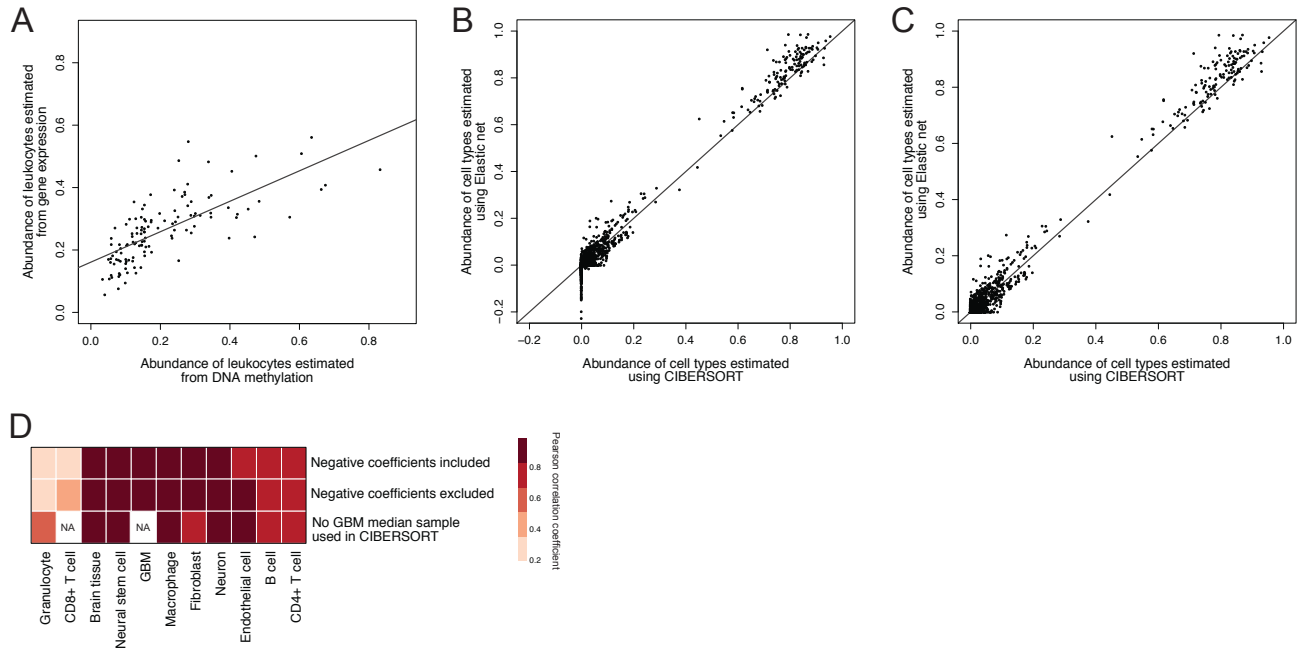
**Figure S3: Validation of the regression model using whole blood samples.** **a** Principal component analysis of whole blood samples from dataset 1 (GSE60424) using all genes in the dataset. The sample with ID 53 is clearly separated from the other samples. **b** Genes used in the deconvolution analysis visualized for dataset 1. The expression pattern of sample with ID 53 looks odd and differs extensively from the other samples. **c** Validation of the regression model using RNA-seq data from whole blood samples (dataset 1, GSE60424). The sample with ID 53 was left out from the median sample and calculation of correlations based on quality control. The four other samples were used as a median sample in the regression analysis. Obtained Pearson's correlations for Lymphocytes, Neutrophil and Monocyte are 0.989, 0.993, 0.984, 0.508, respectively. Macrophage was used as reference for monocyte which most likely causes the lower correlation for monocyte component. **d** Validation of the regression model using RNA-seq data from two whole blood samples (dataset 2, GSE64655). No median sample was used in the analysis due to low number of samples (Pearson's correlation: 0.975). Observed cell type proportions for these two datasets were obtained from article published by Racle et al.

**Supplementary Figure S4: The resulting coefficients for all cell types and GBM samples from regression analysis.**



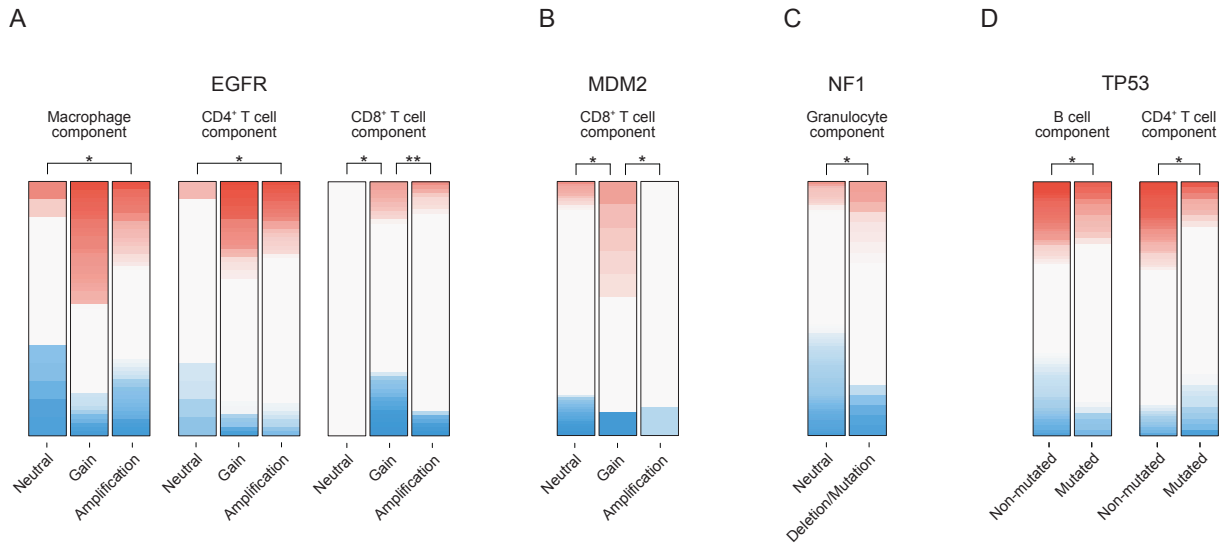
**Figure S4: The resulting coefficients for all cell types and GBM samples from regression analysis.** Proportions of 9 cell types and brain tissue together with median GBM samples were computationally estimated by applying linear regression with elastic net regularization.

**Supplementary Figure S5: Comparison of the cell type estimates from GBM gene expression data using our method to estimates using CIBERSORT or DNA methylation data.**



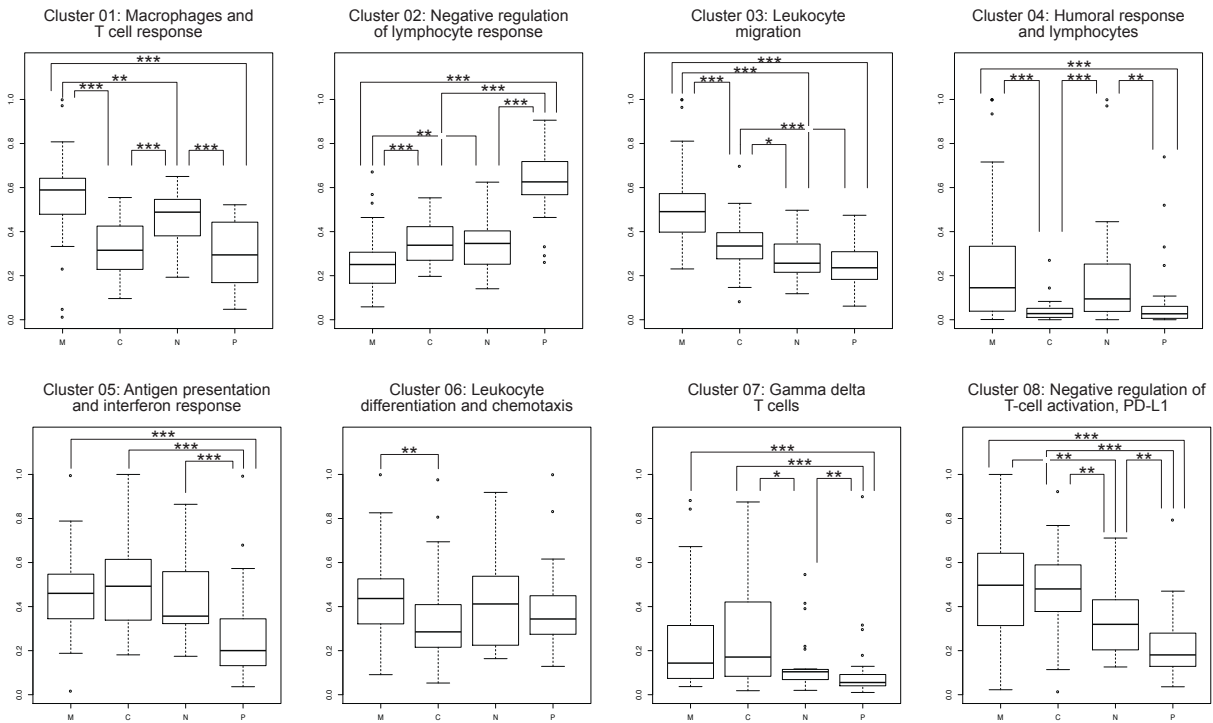
**Figure S5: Comparison of the cell type estimates from GBM gene expression data using our method to estimates using CIBERSORT or DNA methylation data.** **a** Leukocyte methylation score data for GBM samples was downloaded from TCGA PanCancer project from Synapse (syn1809223). Estimates for leukocytes from GBM gene expression were added together and compared to the leukocyte proportions estimated from the GBM DNA methylation data. Estimates from gene expression data correlated well with the DNA methylation estimates (Pearson’s correlation: 0.664) considering that from gene expression data not all leukocytes were estimated and that the DNA methylation estimates aren’t perfect either. Also, estimates from gene expression data are relative cell type proportions so the adding the cell types together is not accurately estimating the amount of the leukocytes. **b** Estimates from GBM gene expression using our method were compared with estimates using CIBERSORT. First, the same reference was used as in the estimates using elastic net. Estimates correlated well although the CIBERSORT is not designed to use median samples as reference and it sets negative values to zero (Pearson’s correlation: 0.99). **c** Comparison of estimates from GBM gene expression data using elastic net and CIBERSORT. Only positive values shown (Pearson’s correlation: 0.994). **d** Pearson’s correlations for each cell type separately. CIBERSORT was also run without the median GBM sample as a reference. This improved the cell type specific correlations. In the analysis without GBM median sample as reference the CIBERSORT estimated the CD8+ T cell proportions to be zero in all of the samples so the correlation coefficient is not available.

**Supplementary Figure S6: Statistically significant associations between estimated immune cell types and typical genetic alterations in GBM.**



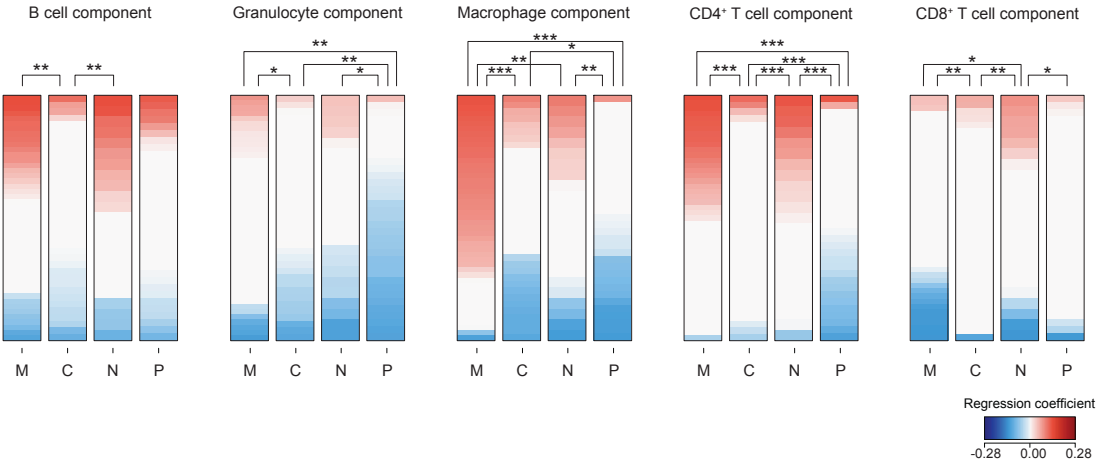
**Figure S6: Statistically significant associations between estimated immune cell types and typical genetic alterations in GBM.** **a** EGFR neutral ( $n = 14$ ) samples have lower estimated macrophage and CD4<sup>+</sup> T cell component compared to samples with EGFR amplification ( $n = 63$ ). Samples with EGFR gain ( $n = 60$ ) have lower estimated CD8<sup>+</sup> T cell component. **b** Samples with MDM2 gain ( $n = 11$ ) have higher CD8<sup>+</sup> T cell component compared to MDM2 neutral samples ( $n = 106$ ) and samples with MDM2 amplification ( $n = 9$ ). **c** NF1 neutral samples ( $n = 116$ ) have lower estimated granulocyte component compared to samples with inactivating NF1 mutation or deletion in NF1 ( $n = 25$ ). **d** Samples with non-mutated TP53 ( $n = 92$ ) have lower B cell and CD4<sup>+</sup> T cell components compared to mutated samples ( $n = 45$ ). \*  $p < 0.05$ , \*\*  $p < 0.01$ , Fisher's exact test.

**Supplementary Figure S7: Association between cluster activities and GBM transcriptional subclassification.**



**Figure S7: Association between cluster activities and GBM transcriptional classification.** Association between Mesenchymal (n = 47), Classical (n = 37), Neural (n = 23) and Proneural (n = 35) classes, and cluster activities were tested using Wilcoxon's rank-sum test. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

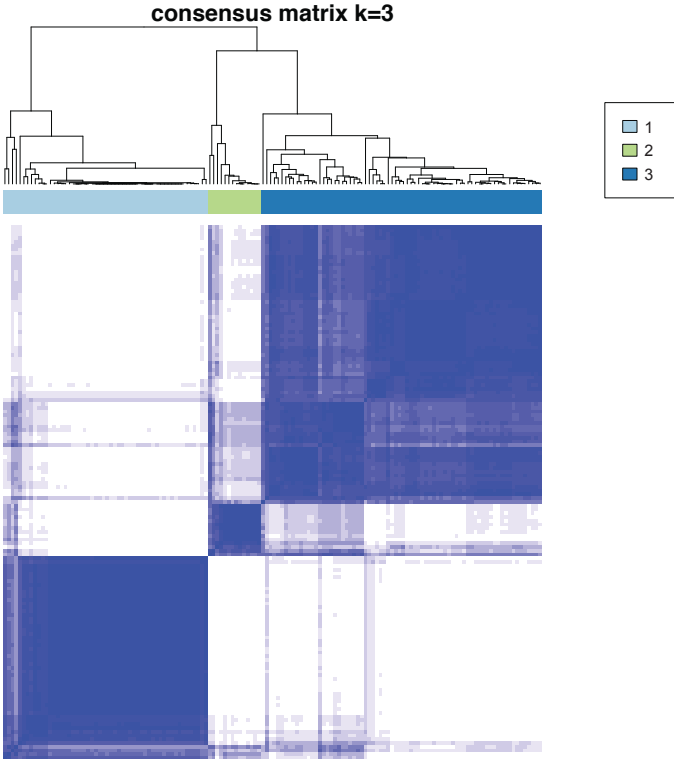
**Supplementary Figure S8: Association between the estimated immune cell type compositions and GBM transcriptional subclassification.**



**Figure S8: Association between the estimated immune cell type compositions and GBM transcriptional subclassification.** Association between Mesenchymal (n = 47), Classical (n = 37), Neural (n = 23) and Proneural (n = 35) classes, and estimated immune cell components were tested using Fisher's exact test. \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

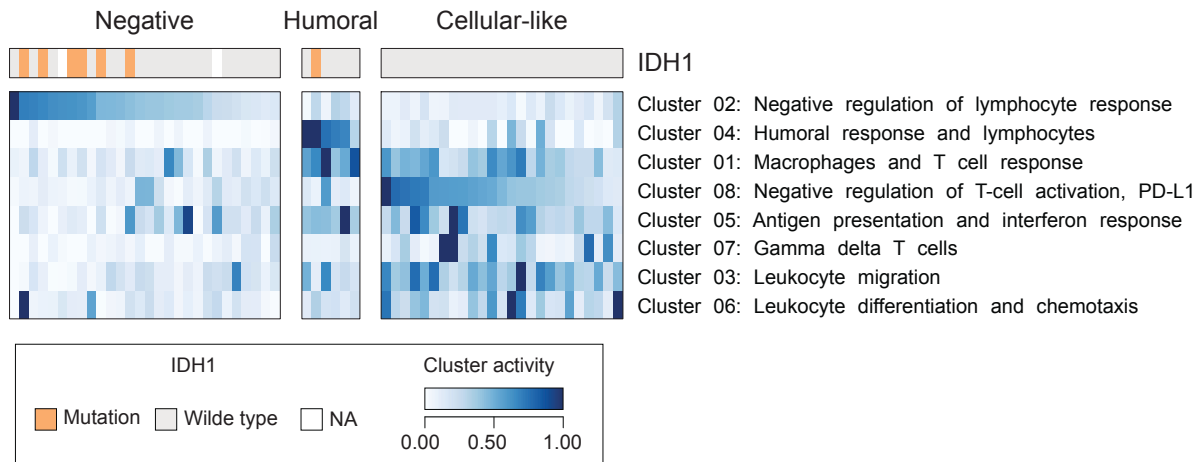


**Supplementary Figure S9: Consensus matrix for three groups.**



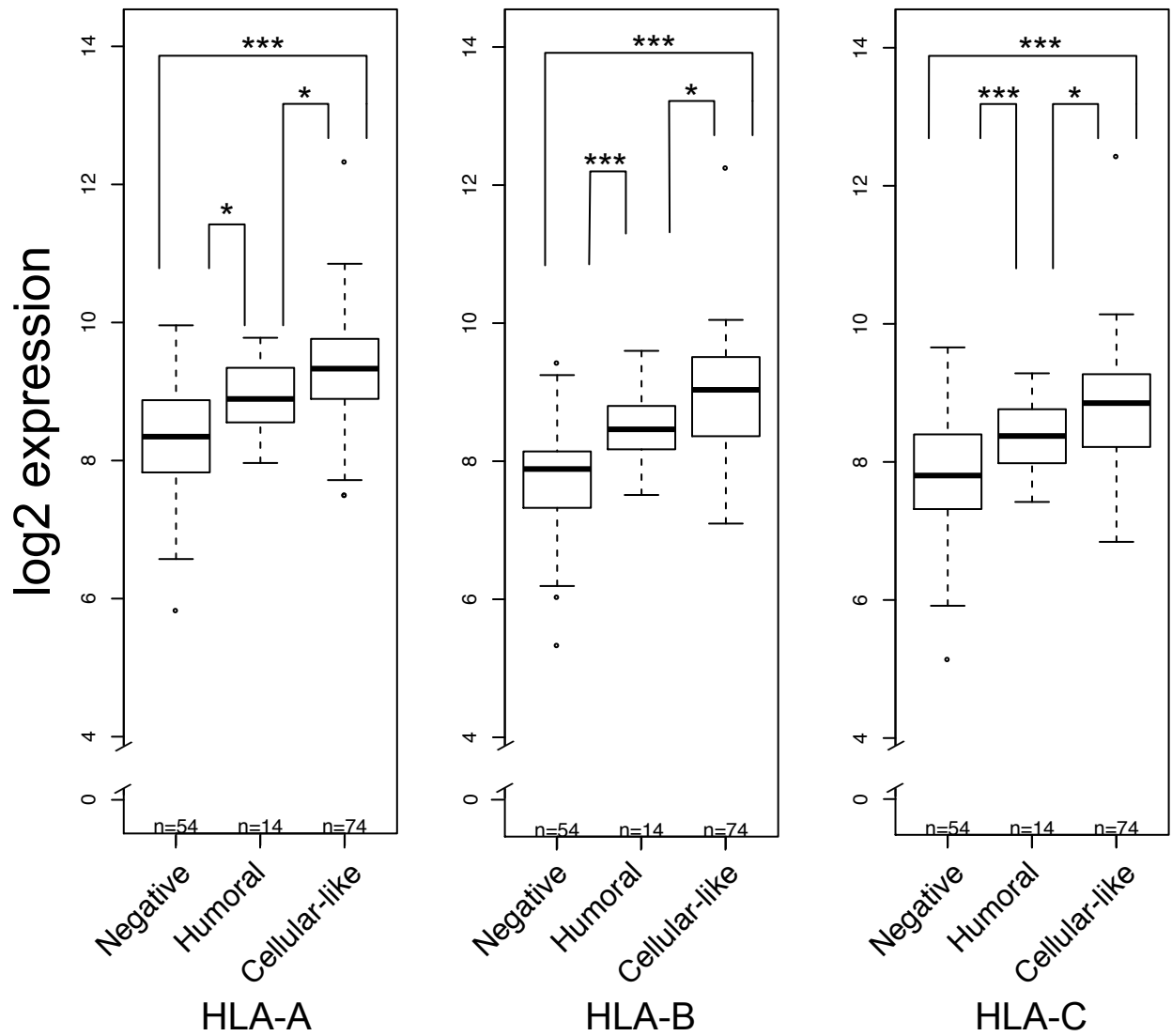
**Figure S9: Consensus matrix for three groups.** Consensus clustering was performed for GBM samples using cluster activities. Pearson’s correlation was used as a similarity metric and clustering was done using k-means.

**Supplementary Figure S10: Results from consensus clustering of the validation GBM dataset.**



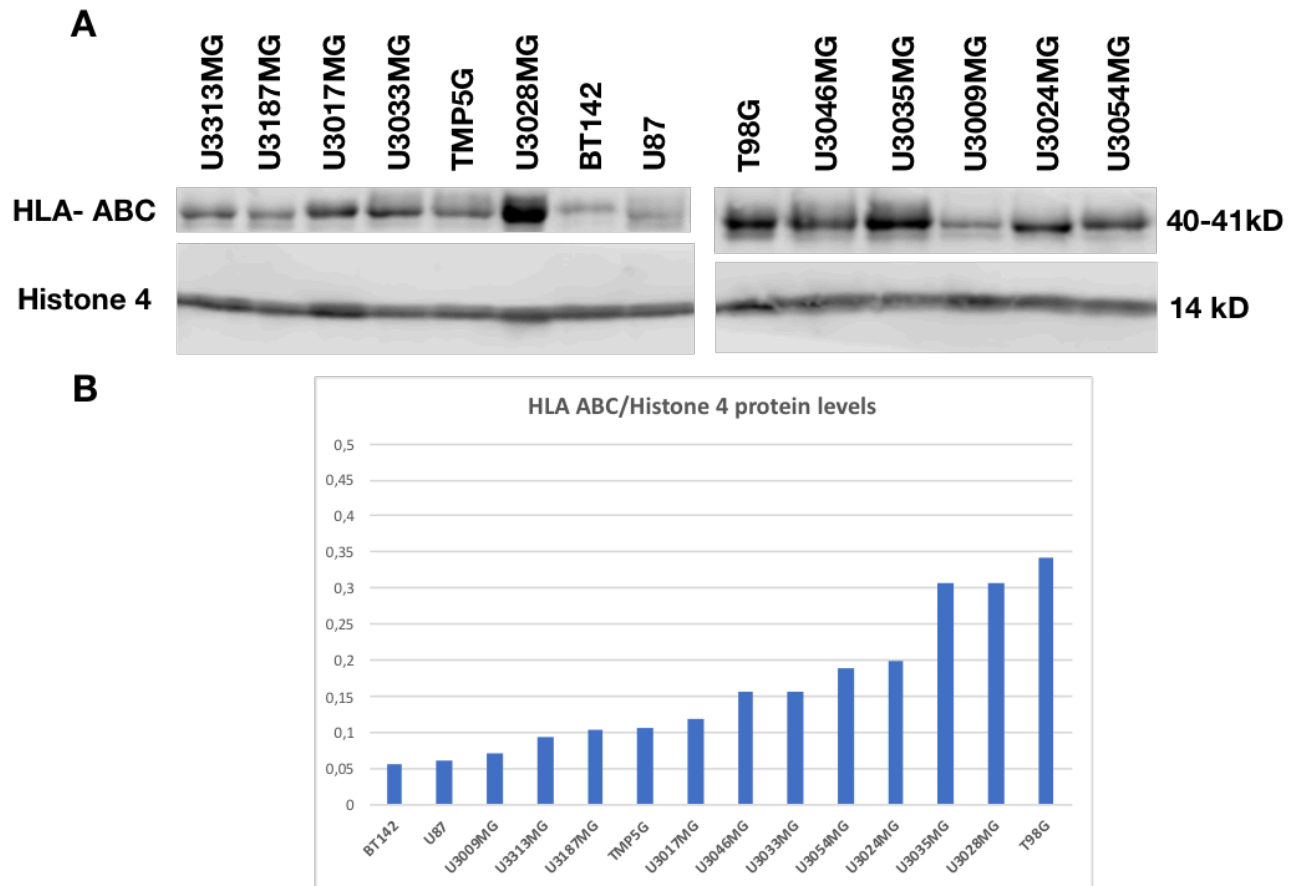
**Figure S10: Results from consensus clustering of the validation GBM dataset.** Consensus clustering was repeated for independent RNA-seq dataset containing 59 primary GBM samples (GSE48865). Negative (n = 28), Humoral (n = 6) and Cellular-like (n = 25) subgroups were identified also from this dataset.

**Supplementary Figure S11: Gene expression of *HLA* genes in different immune subgroups.**



**Supplementary Figure S11: Gene expression of *HLA* genes in different immune subgroups.** Expression of all *HLA* genes is statistically significantly lower in Negative subgroup compared to other two subgroups. Wilcoxon's rank-sum test, \*  $p < 0.05$ , \*\*\*  $p < 0.001$

**Supplementary Figure S12: Protein levels of *HLA-ABC* in glioma cell lines and patient-derived GBM cultures.**



**Supplementary Figure S12: Protein levels of *HLA-ABC* in glioma cell lines and patient-derived GBM cultures.** **A** Protein levels of *HLA-ABC* and Histone 4 (loading control) were analysed by immunoblotting from *IDH1* mutant cell line BT142 and from the following *IDH1* wild-type cell lines: U87, U3009MG, U3313MG, U3187MG, U3017MG, U3046MG, U3033MG, U3054MG, U3024MG, U3035MG, U3028MG and T98G; and the patient-derived GBM culture: TMP5G. For the experiment, cells were lysed in RIPA buffer plus Halt Phosphatase and Protease Inhibitor Cocktail (ThermoFisher Scientific). Equal protein amounts (40µg) were separated on 12% Bis-Tris acrylamide gels and transferred to nitrocellulose membrane (0.45µm, GE Healthcare, Chicago, IL, USA). After blocking (2% BSA in TBS 0.1% Tween-20 (TBS-T) for 1h), membranes were probed with primary antibodies against *HLA-ABC* [1:1,000] (detects *HLA-A*, *HLA-B*, and *HLA-C*) and Histone 4 (loading control) [1:10,000] (Abcam, Cambridge, UK) O/N. Membranes were then incubated with secondary antibodies (Goat anti-Mouse IgG (H+L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus DyLight 800 conjugated [1:10,000] and Goat anti-Rabbit IgG (H+L) DyLight 680 conjugated [1:10,000] (ThermoFisher Scientific)) for 1h. Odyssey CLx (LI-COR Biosciences, Lincoln, NE, USA) was used to visualize the signal. **B** *HLA-ABC* protein levels quantification for the cell lines shown in the westerns in **A**