

PRIMARY RESEARCH

Open Access



Multi-dimensional omics characterization in glioblastoma identifies the purity-associated pattern and prognostic gene signatures

Yi Xiong^{1,2} , Zujian Xiong^{1,2}, Hang Cao^{1,2} , Chang Li^{1,2}, Siyi Wanggou^{1,2} and Xuejun Li^{1,2*} 

Abstract

Background: The presence of tumor-associated stroma and tumor-infiltrated immune cells have been largely reported across glioblastomas. Tumor purity, defined as the proportion of tumor cells in the tumor, was associated with the genomic and clinicopathologic features of the tumor and may alter the interpretation of glioblastoma biology.

Methods: We use an integrative approach to infer tumor purity based on multi-omic data and comprehensively evaluate the impact of tumor purity on glioblastoma (GBM) prognosis, genomic profiling, and the immune microenvironment in the Cancer Genome Atlas Consortium (TCGA) cohort.

Results: We found that low tumor purity was significantly associated with reduced survival time. Additionally, we established a purity-relevant 5-gene signature that was an independent prognostic biomarker and validated it in the TCGA, CGGA and GSE4412 cohort. Moreover, we correlated tumor purity with genomic characteristics and tumor microenvironment. We identified that gamma delta T cells in glioblastoma microenvironment were positively correlated with purity and served as a marker for favorable prognosis, which was validated in both TCGA and CGGA dataset.

Conclusions: We observe the potential confounding effects of tumor purity on GBM clinical and molecular information interpretation. GBM microenvironment could be purity-dependent, which provides new insights into the clinical implications of glioblastoma.

Keywords: Tumor purity, Glioblastoma, Tumor immunity, Tumor heterogeneity, Tumor microenvironment

Background

Glioblastoma (GBM), Grade IV glioma, is an incurable CNS malignancy of adults with high heterogeneity. Despite advances in surgery, radiotherapy, and chemotherapy, the prognosis of GBM patients has not improved significantly, and the median survival remains around 15 months [1]. Recently, increasing evidence has shown that tumor microenvironment plays a pivotal role in tumor biology, including tumor progression and drug

resistance [2, 3]. The presence of specific immune infiltrates or the absence of immunosuppressive signaling were found to indicate positive prognostic features [4]. Much emphasis was placed on tumor-associated macrophages (TAMs), which could participate in tumor progression and metastasis, and influence response to chemotherapy or radiotherapy [5, 6].

Tumor purity is defined as the proportion of tumor cells in the tumor tissue. Over the past few years, tumor purity is routinely determined by pathologists through visual inspections such as immunohistochemistry (IHC) staining, which could be affected by the sensitivity of histopathology, interobserver bias, and variability in accuracy [7]. Several alternative purity estimation methods by

*Correspondence: lxjneuro@csu.edu.cn

¹ Department of Neurosurgery, Xiangya Hospital, Central South University, 87 Xiangya RD, Kaifu District, Changsha 410008, China
Full list of author information is available at the end of the article



© The Author(s) 2020. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

computational approaches were developed recently, and they were based on transcriptome data, DNA methylation data or genome data [8–10]. However, purity estimates inferred by one certain omics data still confines the interpretation of purity in tumor biology systemically in previous studies [11, 12]. To overcome this, a recent study proposed a computational method for calculating the value of purity, namely, the consensus purity estimation (CPE), which was based on ABSOLUTE, ESTIMATE, LUMP and IHC methods [13].

Despite these discoveries, however, little is known regarding the association between the purity and the genomic or clinicopathological features in glioblastoma. Besides, the relationship between purity and glioblastoma microenvironment remains unclear. In this study, we employed the CPE method to estimate the tumor purity and sought to identify the potential confounding effects between tumor purity and clinical or molecular characteristics (Fig. 1a). Thus, we investigated the correlation between tumor purity and genomic alterations, biological pathways as well as immune cell compositions in the microenvironment, which could deepen our understanding of glioblastoma biology and provide new insights into the clinical management of glioblastoma.

Materials and methods

Datasets and data processing

A total of 583 patients with glioblastoma from the TCGA cohort were enrolled in this study. Clinical information and data of molecular biomarkers were acquired from TCGA publications as previously described [14]. In CGGA (Chinese Glioma Genome Atlas) cohort, IlluminaHiSeq RNA-seq data and clinical and molecular information from 144 patients with glioblastoma were obtained via the CGGA database (<http://www.cgga.org.cn>). The raw data for the GSE4412 dataset and corresponding clinical data were acquired from the Gene Expression Omnibus (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE4412>).

Tumor purity analysis

Tumor purity scores were inferred by the consensus purity estimation (CPE) method as previously described [13]. The tumor purity score was derived from the median value estimated from ABSOLUTE, ESTIMATE, LUMP and IHC methods after normalization [9, 10]. To prevent confusion, we defined the CPE score as purity score and used it throughout the article, unless specifically noted.

Transcriptomic data analysis

TCGA AffyU133a gene expression array data and IlluminaHiSeq RNA-seq data of GBMs were downloaded from <https://tcga-data.nci.nih.gov/> via Xena Browser

developed by UCSC. Statistical ranking for purity scores by the top and bottom quartiles were defined as high-purity and low-purity, respectively. Differentially expressed genes (DEGs) between high-purity and low-purity samples were identified using the R/DESeq 2 package or R/Limma package. DEGs with false Discovery Rate (FDR) < 0.05 and fold change > 2 (or < 0.5) were included in further analysis. Both the enrichment analysis and GSEA were performed using R package *clusterProfiler* and *ReactomePA* [15–17]. GSVA analysis of hallmark pathways and metabolic pathways was implemented as previously described [18] by R/GSVA package [19]. GO Enrichment network was drawn using EnrichmentMap software [20] for interpretation.

Methylation analysis

Illumina Infinium DNA methylation platform arrays HumanMethylation450 data in the TCGA-GBM cohort were downloaded from <https://tcga-data.nci.nih.gov/> via Xena Browser developed by UCSC. Data were separated into different datasets according to purity. Data were further normalized and processed by using *ChAMP* package with default parameters [21].

Survival analysis

R package *survival* and *survminer* were used for overall survival analysis. Cox proportional hazard (PH) model was executed by *Coxph* and *Survfit* functions from R packages. The Kaplan–Meier curves were employed to estimate overall survival distribution.

Genomics analysis, intra-tumor heterogeneity (ITH) analysis

Somatic copy number alterations (SCNA) data (minus germline CNV) were downloaded from GDAC Firehose and separated into different datasets according to the purity. SCNA events were detected by GISTIC2.0 using the segmented Affymetrix SNP 6.0 microarray data [22]. Somatic variants files of GBMs in MAF format were downloaded from <https://tcga-data.nci.nih.gov/> via Xena Browser developed by UCSC and downstream analysis was performed by R/maftools package [23]. For somatic nucleotide variations (SNVs), we calculated the total mutation count. We used Mutant-allele tumor heterogeneity (MATH) as a quantitative estimate of intra-tumor heterogeneity (ITH) [24]. Subclone numbers within each sample were inferred by *pyclone* software as previously described and were normalized with tumor purity [25].

Immune cellular fraction estimates

The relative fractions of 24 immune cell types within the leukocyte compartment were estimated using gene sets introduced by Gabriela et al. [26]. IlluminaHiSeq

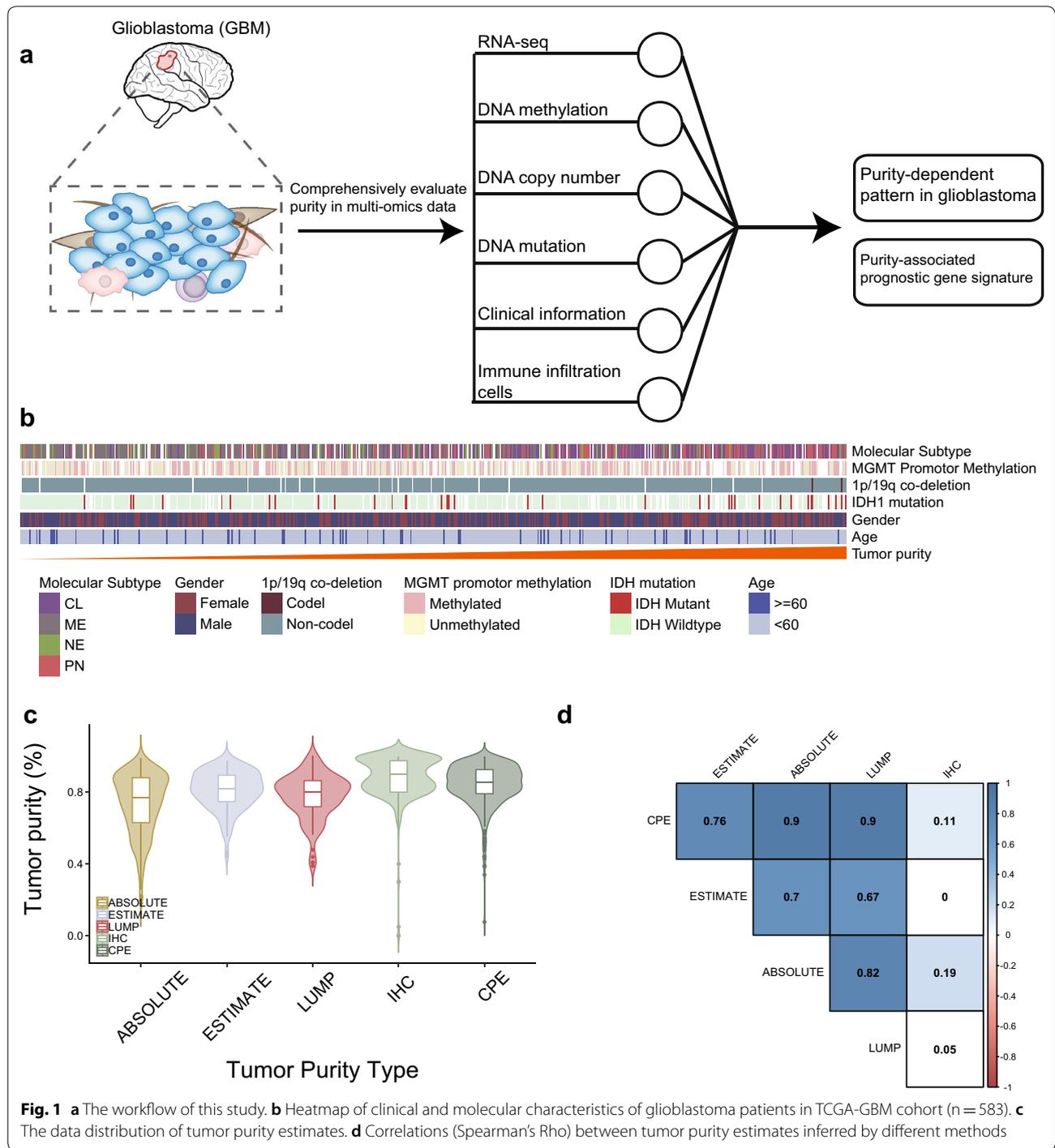


Fig. 1 **a** The workflow of this study. **b** Heatmap of clinical and molecular characteristics of glioblastoma patients in TCGA-GBM cohort (n = 583). **c** The data distribution of tumor purity estimates. **d** Correlations (Spearman's Rho) between tumor purity estimates inferred by different methods

RNA-seq raw counts across all genes for each sample were divided by the gene's maximum transcript length to represent a coverage depth estimate, which was then rescaled to sum to a total depth of 1e6 and thus can be interpreted as transcripts per million (TPM) [27]. We used the RNA-seq TPM data as input and the enrichment

of an immune cell type meta-gene in a given sample were then scored using single-sample gene set enrichment (ssGSEA) analysis [28], as implemented in the GSVA R package [19], with subsequent z-scoring across samples. Note that these enrichments should not be interpreted as deconvolutions of actual cell-type proportions.

Statistics analysis

All statistical analyses were performed using R software, version 3.5.1 (The R Foundation for Statistical Computing, <http://www.rproject.org/>). Continuous variables between groups were compared by the Student's *t* test, one-way analysis of variance (ANOVA) test or the Wilcoxon rank-sum test. Correlations between continuous variables were evaluated by Spearman or Pearson correlation analyses. For all statistical analyses, the *P* value of 0.05 was taken as the significant threshold in all tests.

Results

Tumor purity and clinicopathological and molecular features

An overview of the analytical strategy here is shown in Fig. 1a. Tumor purity scores were calculated by consensus purity estimation (CPE) method based on ABSOLUTE, ESTIMATE, LUMP and IHC algorithms (Additional file 1: Table S1). The purity inferred by CPE is the normalized purity scores inferred from the above four methods (Fig. 1b, c). We observed that tumor purity inferred by the CPE method was significantly positively correlated with purity calculated based on ABSOLUTE, LUMP, ESTIMATE (Spearman's correlation, $\rho=0.90$, 0.90 , 0.76 , respectively) (Fig. 1d), suggesting the rationality of this method.

We next identified the relationship between tumor purity and clinicopathological/molecular features in the TCGA-GBM cohort (Fig. 1b). We observed that tumor purity was significantly enriched in IDH-mutant samples or MGMT-promoter-methylated samples (student's *t*-test, $P<0.001$, $P=0.024$, respectively) (Additional file 2: Fig. S1), which were associated with favorable prognosis. Meanwhile, we analyzed the purity distribution among four GBM molecular subtypes, namely proneural, classical, mesenchymal and neural, based on transcriptome profile [29]. We found that decreased tumor purity levels were enriched in neural or mesenchymal molecular subtypes, which were generally connected with the malignant progression of glioma (Fig. 2a). These findings emphasized that purity was closely related to specific clinicopathological/molecular features.

The prognostic role of tumor purity in glioblastoma

To illustrate the correlation between purity and overall survival, patients were divided into three groups according to the quantiles of the purity score. Kaplan–Meier curve showed that low purity GBM samples (tumors in the bottom 25th percentile) display significantly worse clinical outcomes (High vs Low, $HR=0.75$, $p=0.028$, Log-Rank Test) (Fig. 2b). Subgroup analysis revealed that the low tumor purity has dismal prognosis in the female,

G-GIMP subtype, proneural subtype, and MGMT promoter methylated patients (Additional file 2: Fig. S2).

Construction of a purity-associated gene signature using transcriptomic data

The workflow of data processing and signature construction is shown in Fig. 2c. We first take TCGA RNA-seq dataset as a training set. We divided patients into high-purity cohort (tumors in the top 25th percentile) and low-purity cohort (tumors in the bottom 25th percentile) and differentially expressed genes (DEGs) analysis were performed. By comparing low-purity samples with high-purity samples, we found that 3307 differentially expressed genes, including 2465 upregulated (highly expressed genes in low-purity samples) and 842 down-regulated genes (highly expressed genes in high-purity samples). We next evaluated the prognostic impact of those genes in training set using univariate Cox regression analysis. Finally, we identified a purity-relevant 5-gene signature using the least absolute shrinkage and selection operator (LASSO) Cox regression algorithm. We calculated a risk score by integrating the *z*-score gene expression data and the corresponding coefficients derived from the multivariate Cox regression analysis. The risk score is as follow: $\text{risk score} = 0.152 \cdot \text{SNCB} + 0.003 \cdot \text{KCNN4} + 0.012 \cdot \text{FCGR2C} + 0.348 \cdot \text{PLAUR} + 0.067 \cdot \text{LSP1}$. As a result, a significant difference in overall survival (OS) between the high-risk group and the low-risk group in the training set ($HR=2.24$, 95% CI 1.48–3.38, $p<0.001$) (Fig. 2d). Moreover, we validated the prognostic value of this gene signature in the TCGA-GBM microarray set, CGGA-GBM RNA-seq set and GSE4412 set (Fig. 2d, Additional file 2: Fig. S3). Finally, we performed a multivariate Cox analysis including tumor purity, gender, age, IDH mutation status, risk score as covariates. We identified that the purity-relevant gene signature was an independent prognostic indicator ($HR = 1.17$, 95% CI 1.01–1.36, $p = 0.031$) (Fig. 2e, Additional file 1: Table S2).

Functional annotation of transcriptomic and methylation analysis in tumor purity

We first performed unsupervised clustering of transcriptomic data based on t-SNE or PCA (principle components analysis), which could also divide patients into different groups according to purity (Additional file 2: Fig. S4). To further elucidate the mechanism underlying purity subgroups in transcriptomic architecture, we annotated DEGs using either functional enrichment pathway analysis or gene set enrichment analysis (GSEA) in the TCGA-GBM RNA-seq dataset. GO enrichment analysis of biological processes for the upregulated genes in low-purity samples revealed significant enrichment in “immune response” GO terms (Fig. 3a). Further, GSEA of

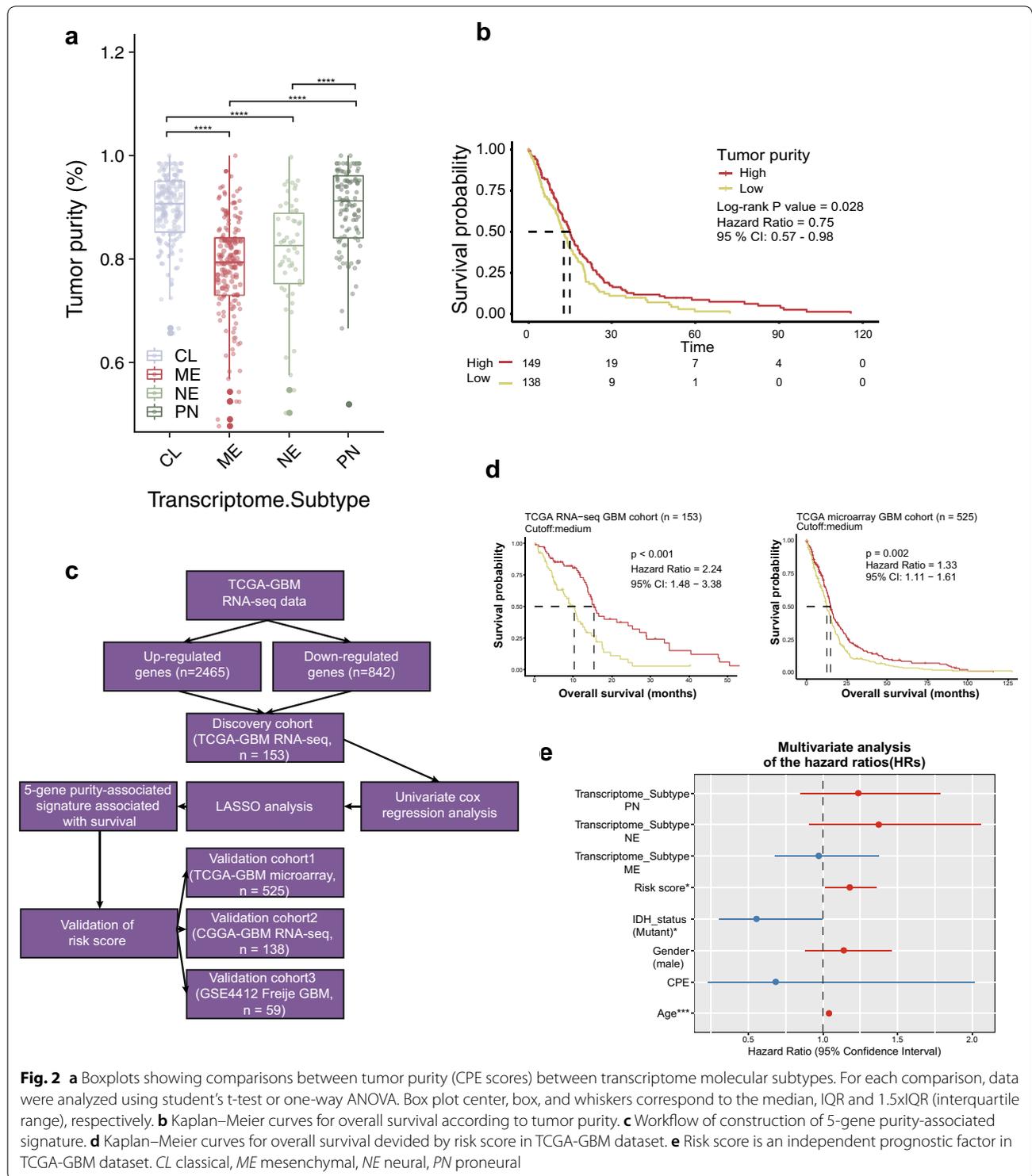


Fig. 2 **a** Boxplots showing comparisons between tumor purity (CPE scores) between transcriptome molecular subtypes. For each comparison, data were analyzed using student's t-test or one-way ANOVA. Box plot center, box, and whiskers correspond to the median, IQR and 1.5xIQR (interquartile range), respectively. **b** Kaplan–Meier curves for overall survival according to tumor purity. **c** Workflow of construction of 5-gene purity-associated signature. **d** Kaplan–Meier curves for overall survival divided by risk score in TCGA-GBM dataset. **e** Risk score is an independent prognostic factor in TCGA-GBM dataset. *CL* classical, *ME* mesenchymal, *NE* neural, *PN* proneural

the pre-ranked gene list revealed that low-purity samples were significantly enriched in immune-related pathways, including B cell receptor signaling pathway, Fc gamma R-mediated phagocytosis, and IL-17 signaling pathway

(Fig. 3b). In addition, to illustrate activated reactions, pathways and biological processes in all samples, Reactome enrichment analysis was performed. As expected, samples with low-purity was significantly enriched in

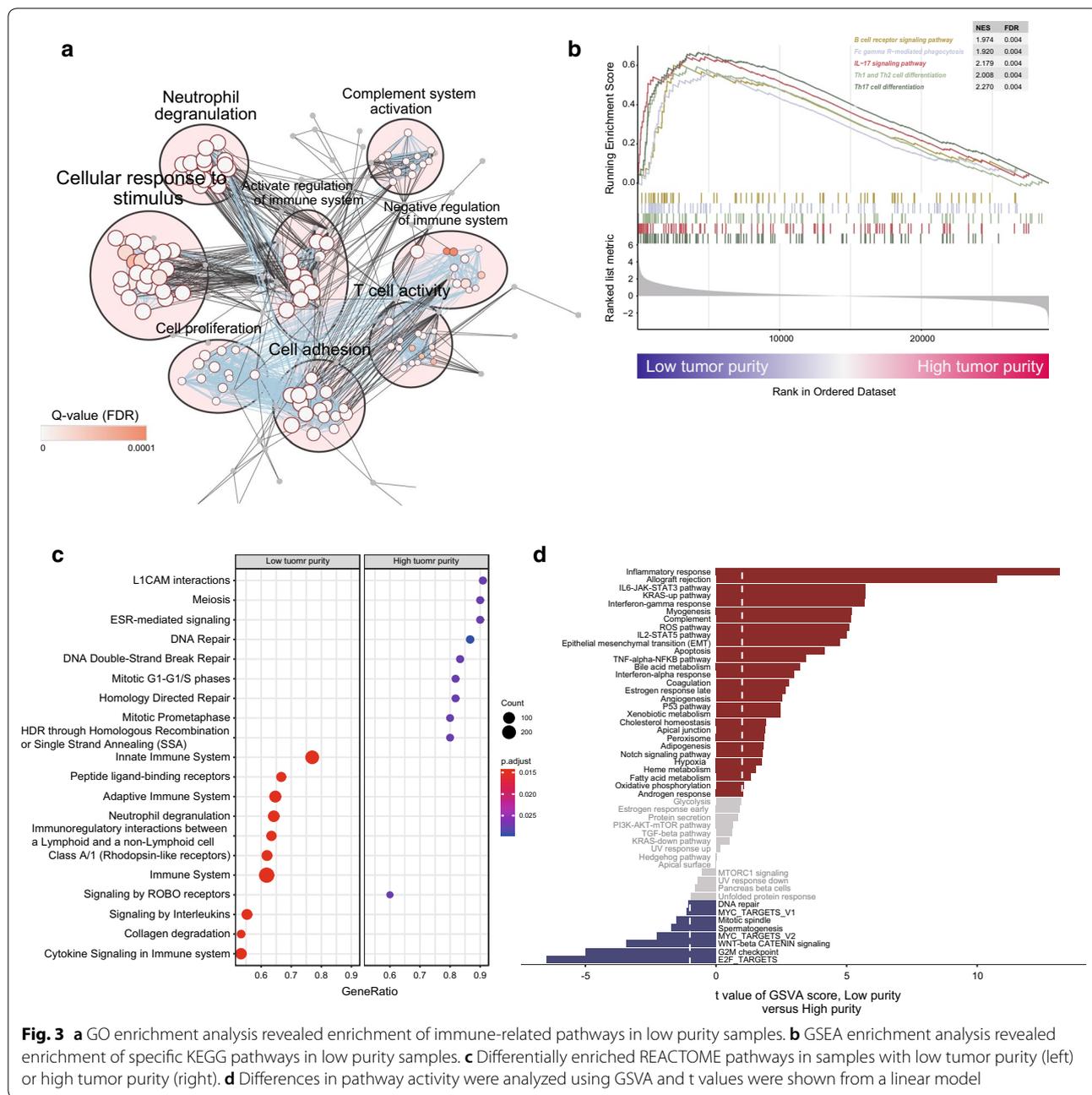


Fig. 3 **a** GO enrichment analysis revealed enrichment of immune-related pathways in low purity samples. **b** GSEA enrichment analysis revealed enrichment of specific KEGG pathways in low purity samples. **c** Differentially enriched REACTOME pathways in samples with low tumor purity (left) or high tumor purity (right). **d** Differences in pathway activity were analyzed using GSVA and t values were shown from a linear model

immune-related signaling pathways and immunoregulatory interactions whereas samples with high-purity showed significant enrichment in cell cycle regulation and DNA repair pathways (Fig. 3c).

Furthermore, we set out to identify the differences in pathway activity among two purity subgroups. We performed gene set variation analysis (GSVA) to assign pathway activity estimates to individual samples. We observed that immune-regulation pathways such as IL6-JAK-STAT3 signaling pathway, and IL2-STAT5 signaling

pathway showed high pathway activity in the low-purity, whereas cell cycle regulation pathways such as G2M checkpoint signaling pathway, E2F signaling pathway displayed high pathway activity in the high-purity (Fig. 3d). All these results suggested the crucial role of regulation of the immune system in low-purity samples.

Comparing malignancy tissues with normal tissues is a common strategy to identify genes associated with tumor progression or tumor-specific markers. Thus, we used the R/Limma package to perform DEGs analysis to

compare the GBM samples (from TCGA cohort) with the normal brain samples (from GTEx cohort) [30]. As purity could be a confounding factor, we controlled purity in DEGs analysis. We observed marked differences in gene expression level before and after purity adjustment and large number of genes were identified as differentially expressed genes after purity adjustment (Additional file 1: Table S3). Genes upregulated in tumor samples could be a marker for tumorigenesis. Here we detected 7460 genes that were upregulated in tumor after purity adjustment, which could be novel genes altered in tumorigenesis (Additional file 2: Fig. S5). Meanwhile, expression values of immune checkpoints genes were important markers in immunotherapy. However, we found that programmed death 1 (PD-1, encoding by *PDCD1* gene) could be upregulated in traditional DEGs analysis. However, when purity was controlled, we did not detect statistically significant results in DEGs analysis (Additional file 1: Table S4).

We also compared DNA methylation profile of GBM samples between high-purity and low-purity samples. We identified differentially methylated probes (two-sided t-test FDR < 0.05) and probes resided in gene promoter were selected as we considered DNA methylation regulation in these genes could be purity-associated. As expected, KEGG enrichment analysis showed that several immune-related pathways are involved in these differentially methylated genes, which could partially explain the differentially expressed genes in the transcriptome (Additional file 2: Fig. S6). In summary, these observations suggested the importance of considering purity as a confounding factor in transcriptome and methylome analysis.

Genomics alterations and tumor purity

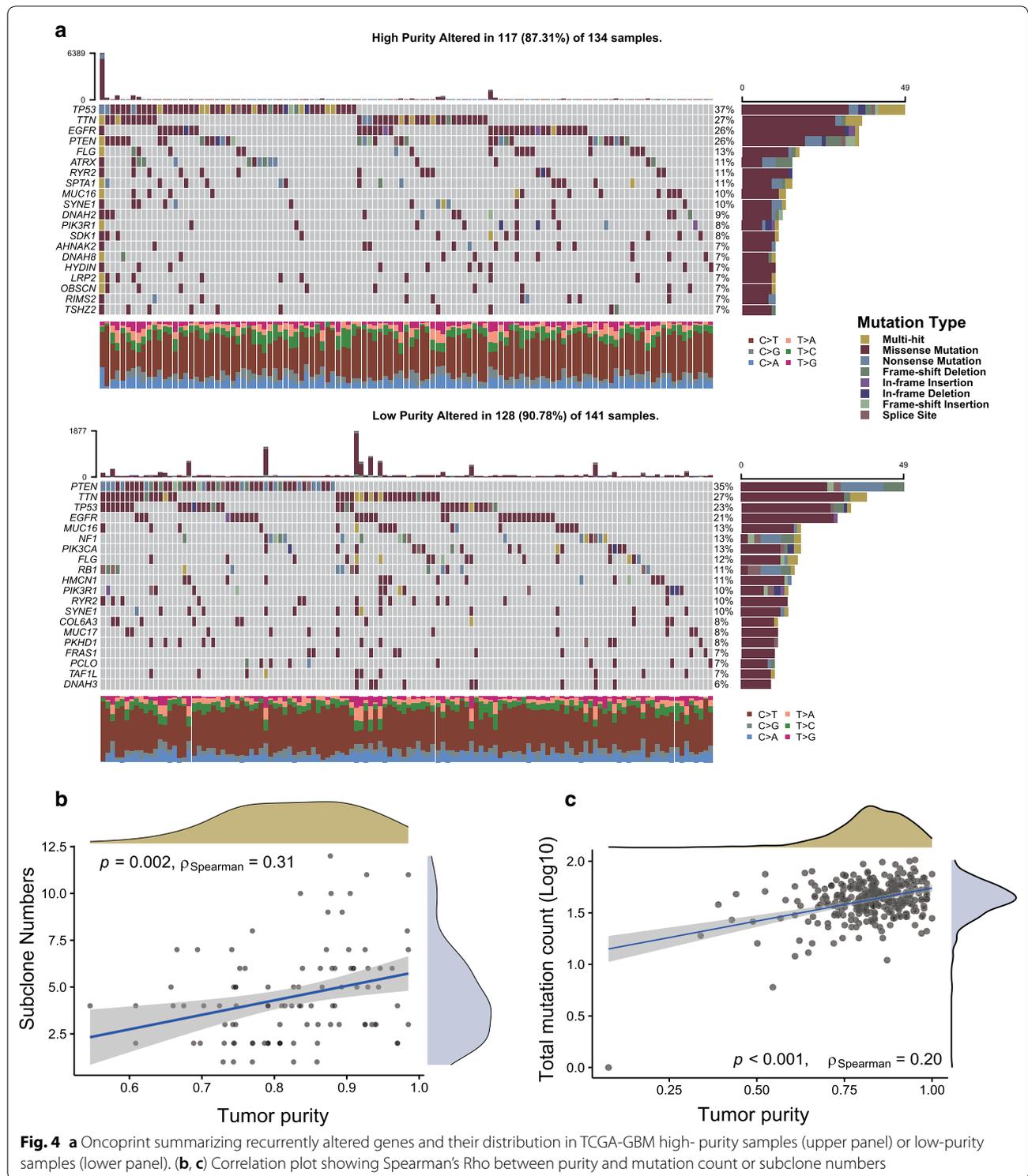
Genomic data from TCGA-GBM dataset were further analyzed to unveil the possible mechanisms affecting tumor purity in terms of the inter-patient genomic heterogeneity. The oncoprint showed the distinct landscape of somatic single nucleotide variants (SNVs) and indels in two tumor purity subgroups (Fig. 4a). *TP53*, *TTN*, *EGFR*, *PTEN* are the most frequently mutated genes in high-purity subgroups while *PTEN*, *TTN*, *TP53*, *EGFR* genes are the most frequently altered in the low-purity subgroup. We further explored the genomic mutations in pathways. By analyzing ten canonical oncogenic signaling pathways [31], we found the significantly high mutation frequency in cell cycle pathways for the low-purity group ($P = 0.0253$, Fisher's exact test) (Additional file 1: Table S5). In addition, we observed that mutation abundance was significantly positively correlated with purity (Fig. 4b).

Next, we explored the association between SCNA and purity (Additional file 2: Fig. S7a). A large number of cytobands are either significantly amplified or deleted regardless of the influence by purity (Additional file 2: Fig. S7b). We performed GO enrichment analysis of biological processes of genes that exclusively altered in either high-purity or low-purity subgroup, which showed mainly differences in the immune regulation process (Additional file 2: Fig. S8). Furthermore, after overlaid with DEGs identified by RNA-seq, we found that 606 genes locate within aberrantly amplified regions, including 35 (5.8%) genes up-regulated in low purity subgroup, suggesting that differential expression of these genes were partially due to copy number variations.

The impact of purity on clonal architecture remains poorly investigated. We next performed a clonality analysis and calculated the MATH value to infer intra-tumoral heterogeneity (ITH). However, there is no significant statistical difference in MATH value between the two groups (Wilcoxon rank-sum test, $P = 0.414$). Interestingly, we observed that decreased subclone numbers were associated with low-purity samples (Fig. 4c). We also found that high-purity was associated with high percent aneuploidy, suggesting genome instability may be enriched in the high-purity (Additional file 2: Fig. S9). Taken together, these findings confirm that purity can be a confounding factor in genomic architecture and purity was an important feature at the genomic level.

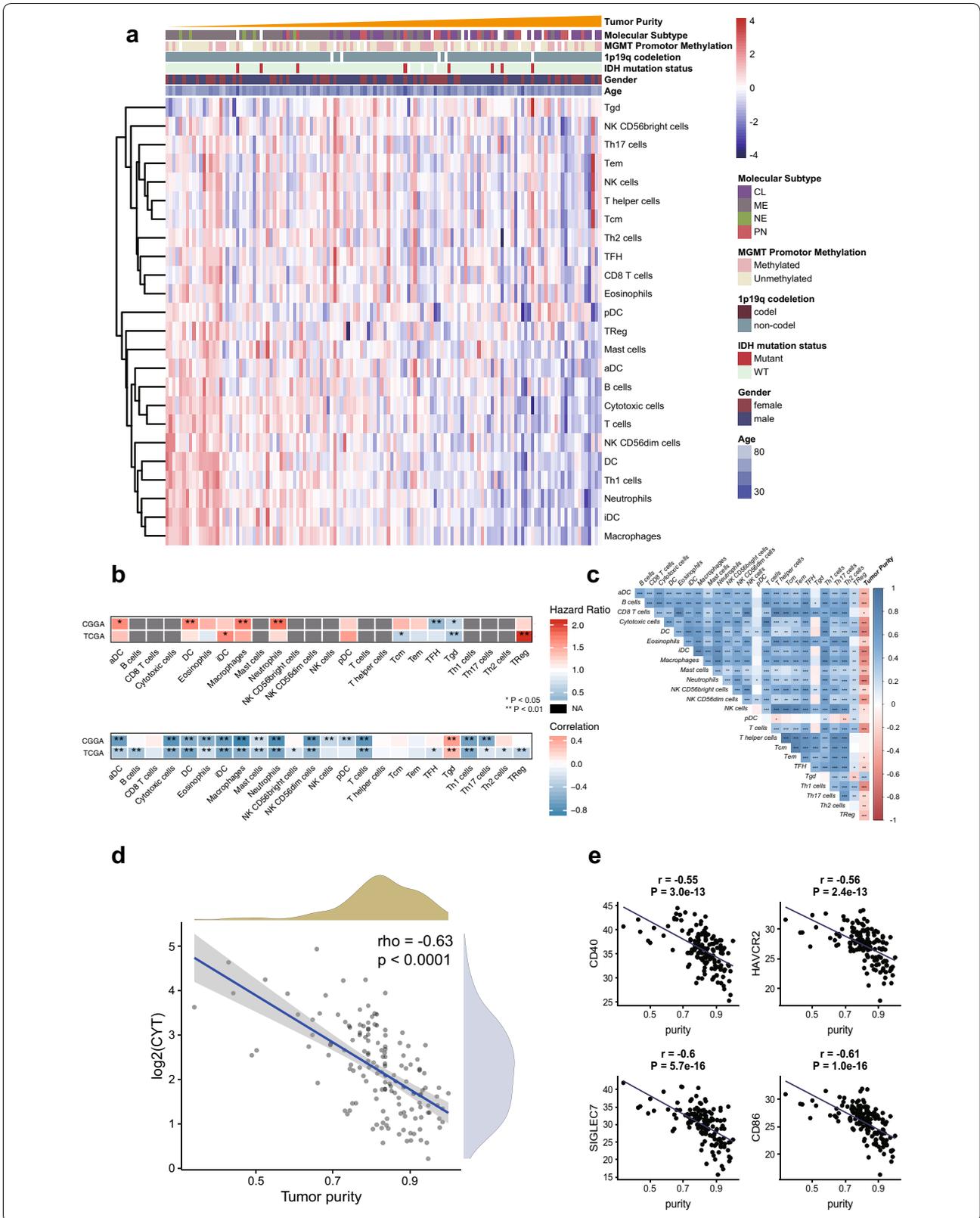
Tumor infiltration and tumor purity

To explore the tumor microenvironment of GBM, the cell abundance of tissue-infiltrating immune cells was estimated in the RNA-seq dataset from the TCGA and CGGA cohort. We estimated 24 subpopulations of immune cells by using ssGSEA strategy. As immune cells composed the majority of non-tumor components of the microenvironment, the proportion of main immune cells is inversely correlated with tumor purity (Fig. 5a, b). We also investigated the association between immune cell types and prognosis of patients (Fig. 5b). The association was varying in different cohorts. The cell types associated with worse prognosis are iDCs and Tregs in TCGA-GBM cohort while aDCs, DCs, macrophages, neutrophils correlated with worse prognosis in CGGA-GBM cohort. Interestingly, we found that gamma delta T cells (Tgd) were enriched in high-purity samples and were associated with favorable prognosis (Fig. 5b) in both TCGA and CGGA cohort (Log-Rank Test, $P < 0.05$). Furthermore, we observed a significant correlation between multiple cell types (Spearman's correlation, $P < 0.05$) (Fig. 5c). Also, we used a simple formula to estimate immune cytolytic activity (CYT), which is assessed by a geometric mean of *GZMA* and *PRF1* expression (TPM value).



We found that CYT was significantly correlated with tumor purity ($\rho = -0.63, P < 0.001$) (Fig. 5d). Consistently, the cytolytic activity could also be a biomarker for unfavorable prognosis [32]. However, we did not identify

a significant correlation between mutation abundance and CYT (Additional file 2: Fig. S10). Finally, we examine the impact of tumor purity on the expression of immune checkpoints genes. As expected, the expression level of



(See figure on previous page.)

Fig. 5 **a** The landscape of immune cell infiltrates sorted by increasing purity in TCGA-GBM RNA-seq dataset. Immune cell infiltrates were estimated by ssGSEA algorithm. **b** The correlation between the proportion of immune cell infiltrates and survival (upper panel) or purity estimates (lower panel) in TCGA-GBM or CGGA RNA-seq cohort. Purity values in CGGA cohort were inferred by ESTIMATE method. **c** Correlation plot showing Spearman's Rho between cell types in TCGA-GBM. **d** Scatter plot of correlation of tumor purity and CYT (a geometric mean of GZMA and PRF1; y-axis in log₂ scale). **e** Correlation between immune checkpoints gene expression (TCGA RNA-seq dataset) and tumor purity. Pearson's correlation coefficients (r) are stated

HAVCR2, *CD40*, *SIGLEC7*, *CD86* genes are inversely correlated with tumor purity (Fig. 5d). Taken together, these findings demonstrated that purity was an important characteristic of tumor microenvironment.

Discussion

To our knowledge, this is the first comprehensive study to investigate the confounding effects of tumor purity based on the CPE method in a series of clinicopathological, genomic, immune parameters for GBM patients. The key findings of this study were: 1. We identified purity-dependent distinct patterns associated with genomic and clinicopathologic features, which supports the hypothesis that tumor purity is an intrinsic characteristic of samples reflecting differences in the tumor microenvironment [13]. 2. We constructed a purity-associated gene signature which would be prognostically relevant. 3. By analyzing the GBM tumor microenvironment, we revealed that gamma delta T cells were positively associated with purity and were a favorable prognostic indicator.

Gene expression subtypes of GBM identified by unsupervised clustering have emerged as an important tool to understand GBM biology [29, 33]. We demonstrated that the low-purity was tightly associated with mesenchymal or neural subtypes. In accordance with previous studies, mesenchymal subtype was reported to be linked with tumor-associated glial and microglial cells and neural subtype was related to the tumor margin where normal neural tissue could be easily collected [33]. To analyze the confounding effect of purity based on transcriptomic data, we performed two parallel strategies. First, we assessed the DEGs between tumor and normal tissue samples, identified before and after purity adjustment, which is a well-established approach to screen tumor-associated biomarkers. We revealed that up-regulated genes identified after purity adjustment may play important roles in the biomarker setting, which requires further validation. Second, we also analyzed the upregulated and downregulated genes regarding their correlations with purity. We observed that, genes associated with high-purity, were significantly enriched in pathways related to the tumor-intrinsic characteristics such as abnormalities in cell cycle regulation and impaired DNA repair machinery; In low-purity samples, upregulated genes were commonly enriched in

immunoregulation and cellular interaction pathways, suggesting a phenotype of tumor microenvironment with increased immune-infiltration in these patients. Moreover, several pathways were activated in the low-purity phenotype. For instance, in tumor immunity, IL-10 signaling pathways plays a dual role that IL-10 promotes tumor immune escape by inhibiting inflammatory cytokines, and conversely induces tumor-specific CD8⁺ T cells infiltration and promotes their cytotoxic activity [34]. In addition, IL-17-producing cells, on the other side, were found promoting tumor infiltration and acquired survival benefits in the metabolites-deficiency tumor microenvironment [35]. We also found that the IL6-JAK-STAT3 signaling pathway, which drives the proliferation, survival, invasiveness, and metastasis of tumor cells and suppresses the antitumor immune response in the tumor microenvironment [36], displayed high pathway activity by GSVA analysis in this study. Hyperactivation of this pathway is generally associated with poor prognosis and thus this pathway could be therapeutically targeted by inhibitors [36].

In previous studies, purity levels were associated with histological subtypes and histological grades, as well as survival time [11–13, 37]. Consistent with these findings, tumor purity could be a potential prognostic indicator for GBM since low purity cases were associated with poor prognosis in our study. As tumor purity could not be an independent prognostic factor while controlling other factors such as IDH mutation status, age, etc., we were motivated to establish a purity-associated gene signature. In our 5-gene signature, to take the *PLAUR* gene as an example, the *PLAUR* gene encodes the urokinase receptor (uPAR) and had the largest positive coefficient in the Cox regression model. The prognostic value of *PLAUR* gene has been reported [38]. *PLAUR* could be functionally related to tumor growth and angiogenesis [39].

In this study, we also reported that the tumor mutation burden (TMB) of the samples was positively correlated with purity. Of note, we also identified that high-purity phenotype correlates with high aneuploidy, that is, increased genome instability. The possible explanations are that high genome instability is often associated with high TMB and increased pro-inflammatory activity that causes a higher percent of tumor necrosis component with decreased immune cell

infiltration [5]. Strikingly, we observed a decreased total number of subclonal numbers in low-purity samples, which is consistent with a recent study in renal cell carcinoma [40]. The possible mechanism behind this phenomenon is that during immune selection, increased immunoeediting was found to eliminate tumor clones in samples with high immune infiltration [41].

We set out to analyze the immune microenvironment architecture for glioblastoma. We demonstrated that most immune cells inversely correlate with the purity whereas Tgd (T gamma delta) cells are highly infiltrated in the high purity subgroup. Consistent with the previous study [42], increasing Tgd infiltration conferred favorable prognosis in both the TCGA and CGGA cohort, which could partially help to explain favorable prognosis in the high-purity subgroup. In mechanism, gamma delta T cells have been reported with well-established protective roles in cancer, largely based on their potent cytotoxicity and interferon- γ production [43], implying a potential predictive role in GBM immunotherapy.

One of the main advantages of our research was the use of CPE to evaluate the tumor purity and systematic analysis of the purity in multidimensional profiles for GBM. Nevertheless, the present study has several limitations. First, due to the retrospective setting of the TCGA and CGGA samples, results need to be carefully assessed and validated in future prospective studies. Second, immune infiltration analysis was performed based on transcriptomic data. Thus, our analyses were limited without cell phenotype confirmation by other methods. Due to the natural complexity of glioblastoma tumor microenvironment, further studies to promote its understanding might be helpful.

Conclusion

In summary, we systematically evaluated the role of tumor purity in the GBM prognosis, genomics, and transcriptome alterations as well as tumor immunity microenvironment. The tumor purity influences the intra-tumor heterogeneity and genomic architecture. It is important for further studies in glioblastoma biology to consider tumor purity as a confounding effect in the design.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12935-020-1116-3>.

Additional file 1: Table S1. The purity values in TCGA-GBM cohort. **Table S2.** Cox proportional hazards model in TCGA-GBM cohort. **Table S3.** Differentially expressed genes between tumor and normal samples.

Table S4. Expression of Immune checkpoint molecules before and after purity adjustment. **Table S5.** Differentially mutated gene frequency in oncogenic signaling pathways.

Additional file 2: Figure S1. Relation between purity and IDH mutation status or MGMT promotor methylation status. **Figure S2.** The prognostic value of purity in stratified GBMs. **Figure S3.** The prognostic role of purity-associated risk score in CGGA or GSE4412 cohort. **Figure S4.** Unsupervised analyses of global transcriptional similarities and differences between two purity subgroups. **Figure S5.** Adjustment of purity in differentially expressed genes analysis. **Figure S6.** Enrichment of KEGG pathways in differentially methylated genes. **Figure S7.** Relation between purity and genomic alterations. **Figure S8.** GO enrichment analysis of differentially amplified genes or differentially deleted genes between purity subgroups. **Figure S9.** Correlation between tumor purity and genomic instability. **Figure S10.** Correlation between CYT and mutation abundance.

Abbreviations

CGGA: Chinese Glioma Genome Atlas; CPE: consensus purity estimation; CNS: central neural system; CL: classical; CNV: copy number variation; DEG: differentially expressed gene; GBM: glioblastoma; G-CIMP: glioma CpG island methylator phenotype; GO: gene ontology; HR: hazard Ratio; ITH: intra-tumor heterogeneity; IHC: immunohistochemistry; KEGG: Kyoto Encyclopedia of Genes and Genomes; MATH: mutant-allele tumor heterogeneity; MES: mesenchymal; NE: neural; PN: proneural; PCA: principle components analysis; SCNA: somatic copy number alteration; SNV: somatic nucleotide variations; t-SNE: t-distributed stochastic neighbor embedding; TMB: tumor mutation burden; TCGA: The Cancer Genome Atlas Consortium.

Acknowledgements

We thank the TCGA and CGGA program for kindly providing publicly available data for analysis.

Authors' contributions

YX, ZJ-X and XJ-L conceived and designed the study, YX and ZJ-X crafted the literature search, figures and tables and was responsible for the writing and critical reading of the manuscript. YX, ZX, CL and HC contributed to the data analysis and the critical reading of manuscript. SY-WG, XJ-L supervised and contributed to the critical reading of manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the National Natural Science Foundation of China (No. 81472594, 81770781).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the ethics committee of Xiangya Hospital, and the written informed consent was obtained from all patients.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ Department of Neurosurgery, Xiangya Hospital, Central South University, 87 Xiangya RD, Kaifu District, Changsha 410008, China. ² Hunan International Scientific and Technological Cooperation Base of Brain Tumor Research, Xiangya Hospital, Central South University, Changsha 410008, Hunan, China.

Received: 15 November 2019 Accepted: 18 January 2020
Published online: 31 January 2020

References

- Jiang T, Mao Y, Ma W, Mao Q, You Y, Yang X, et al. CGCG clinical practice guidelines for the management of adult diffuse gliomas. *Cancer Lett*. 2016;375(2):263–73.
- Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell*. 2011;144(5):646–74.
- Quail DF, Joyce JA. Microenvironmental regulation of tumor progression and metastasis. *Nat Med*. 2013;19(11):1423–37.
- Fridman WH, Zitvogel L, Sautes-Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. *Nat Rev Clin Oncol*. 2017;14(12):717–34.
- Wang X, Li M. Correlate tumor mutation burden with immune signatures in human cancers. *BMC Immunol*. 2019;20(1):4.
- Junttila MR, de Sauvage FJ. Influence of tumour micro-environment heterogeneity on therapeutic response. *Nature*. 2013;501(7467):346–54.
- Cohen DA, Dabbs DJ, Cooper KL, Amin M, Jones TE, Jones MW, et al. Inter-observer agreement among pathologists for semiquantitative hormone receptor scoring in breast carcinoma. *Am J Clin Pathol*. 2012;138(6):796–802.
- Zheng X, Zhao Q, Wu HJ, Li W, Wang H, Meyer CA, et al. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. *Genome Biol*. 2014;15(8):419.
- Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun*. 2013;4:2612.
- Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30(5):413–21.
- Zhang C, Cheng W, Ren X, Wang Z, Liu X, Li G, et al. Tumor purity as an underlying key factor in glioma. *Clin Cancer Res*. 2017;23(20):6279–91.
- Mao Y, Feng Q, Zheng P, Yang L, Liu T, Xu Y, et al. Low tumor purity is associated with poor prognosis, heavy mutation burden, and intense immune phenotype in colon cancer. *Cancer Manag Res*. 2018;10:3569–77.
- Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun*. 2015;6:8971.
- Chen X, Wanggou S, Bodalia A, Zhu M, Dong W, Fan JJ, et al. A feedforward mechanism mediated by mechanosensitive ion channel PIEZO1 and tissue mechanics promotes glioma aggression. *Neuron*. 2018;100(4):799–815.e7.
- Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545–50.
- Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7.
- Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol Biosyst*. 2016;12(2):477–9.
- Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med*. 2018;24(8):1277–89.
- Hanzelmann S, Castelo R, Guinney J. GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinform*. 2013;14:7.
- Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS ONE*. 2010;5(11):e13984.
- Tian Y, Morris TJ, Webster AP, Yang Z, Beck S, Feber A, et al. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics*. 2017;33(24):3982–4.
- Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhir R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol*. 2011;12(4):41.
- Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747–56.
- Mroz EA, Rocco JW. MATH, a novel measure of intratumor genetic heterogeneity, is high in poor-outcome classes of head and neck squamous cell carcinoma. *Oral Oncol*. 2013;49(3):211–5.
- Andor N, Graham TA, Jansen M, Xia LC, Aktipis CA, Petritsch C, et al. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nat Med*. 2016;22(1):105–13.
- Bindea G, Mlecnik B, Tosolini M, Kirilovsky A, Waldner M, Obenaus AC, et al. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity*. 2013;39(4):782–95.
- Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*. 2015;160(1–2):48–61.
- Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*. 2009;462(7269):108–12.
- Verhaak RG, 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(1):98–110.
- Consortium GT. The genotype-tissue expression (GTEx) project. *Nat Genet*. 2013;45(6):580–5.
- Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*. 2018;173(2):321–337.e10.
- Wang ZL, Wang Z, Li GZ, Wang QW, Bao ZS, Zhang CB, et al. Immune cytolytic activity is associated with genetic and clinical properties of glioma. *Front Immunol*. 2019;10:1756.
- 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(1):e642–56.
- Mannino MH, Zhu Z, Xiao H, Bai Q, Wakefield MR, Fang Y. The paradoxical role of IL-10 in immunity and cancer. *Cancer Lett*. 2015;367(2):103–7.
- Veldhoen M. Interleukin 17 is a chief orchestrator of immunity. *Nat Immunol*. 2017;18(6):612–21.
- Johnson DE, O’Keefe RA, Grandis JR. Targeting the IL-6/JAK/STAT3 signalling axis in cancer. *Nat Rev Clin Oncol*. 2018;15(4):234–48.
- Schulze Heuling E, Knab F, Radke J, Eskilsson E, Martinez-Ledesma E, Koch A, et al. Prognostic relevance of tumor purity and interaction with MGMT methylation in glioblastoma. *Mol Cancer Res*. 2017;15(5):532–40.
- Hsu JB, Chang TH, Lee GA, Lee TY, Chen CY. Identification of potential biomarkers related to glioma survival by gene expression profile analysis. *BMC Med Genomics*. 2019;11(Suppl 7):34.
- Lakka SS, Gondi CS, Yanamandra N, Dinh DH, Olivero WC, Gujrati M, et al. Synergistic down-regulation of urokinase plasminogen activator receptor and matrix metalloproteinase-9 in SNB19 glioblastoma cells efficiently inhibits glioma cell invasion, angiogenesis, and tumor growth. *Cancer Res*. 2003;63(10):2454–61.
- Senbabaoglu Y, Gejman RS, Winer AG, Liu M, Van Allen EM, de Velasco G, et al. Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biol*. 2016;17(1):231.
- Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of metastases in space and time under immune selection. *Cell*. 2018;175(3):751–765.e16.
- Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, et al. The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nat Med*. 2015;21(8):938–45.
- Silva-Santos B, Serre K, Norell H. Gammadelta T cells in cancer. *Nat Rev Immunol*. 2015;15(11):683–91.

Publisher’s Note

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