Identification and Analysis of Glioblastoma Biomarkers Based on Single Cell Sequencing

Quan Cheng[1,2†], Jing Li[3†], Fan Fan[1], Hui Cao[4], Zi-Yu Dai[1], Ze-Yu Wang[1] and Song-Shan Feng[1*]

[1] Department of Neurosurgery, Xiangya Hospital, Central South University, Changsha, China, [2] Department of Clinical Pharmacology, Xiangya Hospital, Central South University, Changsha, China, [3] Department of Rehabilitation, The Second Xiangya Hospital, Central South University, Changsha, China, [4] Department of Psychiatry, The Second People's Hospital of Hunan University of Chinese Medicine, Changsha, China

Glioblastoma (GBM) is one of the most common and aggressive primary adult brain tumors. Tumor heterogeneity poses a great challenge to the treatment of GBM, which is determined by both heterogeneous GBM cells and a complex tumor microenvironment. Single-cell RNA sequencing (scRNA-seq) enables the transcriptomes of great deal of individual cells to be assayed in an unbiased manner and has been applied in head and neck cancer, breast cancer, blood disease, and so on. In this study, based on the scRNA-seq results of infiltrating neoplastic cells in GBM, computational methods were applied to screen core biomarkers that can distinguish the discrepancy between GBM tumor and pericarcinomatous environment. The gene expression profiles of GBM from 2343 tumor cells and 1246 periphery cells were analyzed by maximum relevance minimum redundancy (mRMR). Upon further analysis of the feature lists yielded by the mRMR method, 31 important genes were extracted that may be essential biomarkers for GBM tumor cells. Besides, an optimal classification model using a support vector machine (SVM) algorithm as the classifier was also built. Our results provided insights of GBM mechanisms and may be useful for GBM diagnosis and therapy.

Keywords: glioblastoma biomarkers, scRNA-seq, mRMR method, support vector machine, pericarcinomatous environment

INTRODUCTION

Glioblastoma (GBM), with an annual incidence of 3.19 per 100,000, maintains the most common and aggressive primary adult brain tumor (Stupp et al., 2007, 2017; Chinot et al., 2014; Gilbert et al., 2014; Ostrom et al., 2016). Currently, the standard therapeutic regimen has been established, including surgical resection, followed by radiotherapy with concurrent chemotherapy (temozolomide), then followed by maintenance therapy (temozolomide for 6–12 months) (Stupp et al., 2005). However, the diffuse nature of GBMs makes it invariably recur after treatment, rendering local therapies invalid, because the migrating GBM cells outside of the neoplasm core are usually unaffected by local therapies and hence cause recurrence of GBMs (Darmanis et al., 2017). The mean disease-free survival is just over 6 months and the mean overall survival also remains gloomy, with an approximately 25% 2-year survival rate after diagnosis and a 5–10% 5-year survival rate (Stupp et al., 2005, 2017; Das and Marsden, 2013).

Tumor heterogeneity poses a great challenge to the treatment of GBM, which is determined by both heterogeneous GBM cells and a complex tumor microenvironment. It is critical important

for researchers to understand how different types of GBM cells interact with neoplasm cells through profiling of different types of cell from cell population in paraneoplastic environment, as well as identifying the lineage and phenotypes (Darmanis et al., 2017). Verhaak et al. (2010) has proved bulk tumor sequencing methods were useful in generating classification schemas of GBM subtypes, but the heterogeneity of GBM was not unveiled in essence (Cancer Genome Atlas Research Network, 2008). Until recently, RNA profiling was limited to ensemble-based approaches, averaging over bulk cell populations. Therefore, the advent of single-cell RNA sequencing (scRNA-seq) enables the transcriptomes of great deal of individual cells to be assayed in an unbiased manner (Stegle et al., 2015) and has been applied in head and neck cancer (Puram et al., 2017), breast cancer (Bajikar et al., 2017), blood disease (Zhao et al., 2017), and so on. Patel et al. (2014) profiled 430 cells from five GBM patients using scRNA-seq and described inter-patient variation and molecular diversity of tumor cells within individual GBM patients. The diversities of GBM cells within tumors are responsible for cancer progression and finally result in treatment failure.

Currently, in order to improve future treatment options, an increasing number of researchers have focused on the targeted agents or genes (Liu et al., 2013; Xiao et al., 2014; Li et al., 2018). Furnari et al. (2007) have identified genetic molecular mechanisms in GBM patients: (1) dysregulation of growth factor signaling through amplification and mutational activation of receptor tyrosine kinase (RTK) genes; (2) activation of the phosphatidyl inositol 3-kinase (PI3K) pathway; and (3) deactivation of the p53 and retinoblastoma tumor suppressor pathways. Moreover, four distinct GBM subclasses, including neural, proneural (PGFRA/IDH1 events), classical (focal EGFR events), and mesenchymal (NF1 mutation and loss), were defined by gene expression studies from The Cancer Genome Atlas (TCGA) (Verhaak et al., 2010), which also found the majority of GBM neoplasms had abnormalities in the pathways (RB, TP53, and RTK) through projecting copy number and mutation data on these pathways, revealing that this is a crucial step for GBM pathogenesis. Apart from such researches focused on tumor or microenvironment, many studies analyzed the gene expression of immune cells in GBM via scRNA-seq. Muller et al. (2017) identified 66 new gene sets which can be applied as biomarkers (such as P2RY12, CD49D, and HLA-DRA) to distinguish the different lineages of the macrophage cell subsets.

In this study, based on the scRNA-seq results of infiltrating neoplastic cells in GBM, computational methods were applied to screen core biomarkers that can distinguish the discrepancy between GBM tumor and pericarcinomatous environment. The gene expression profiles of GBM from 2343 tumor cells and 1246 periphery cells were analyzed by maximum relevance minimum redundancy (mRMR) (Peng et al., 2005). Upon further analysis of the feature lists yielded by the mRMR method, 31 important genes were extracted that may be essential biomarkers for GBM tumor cells. Besides, an optimal classification model using a support vector machine (SVM) algorithm (Ding and Dubchak, 2001) as the classifier was also built.

MATERIALS AND METHODS

The Single Cell Gene Expression Profiles of Tumor and Surrounding Tissues

We download the single cell gene expression profiles of 2343 cells of tumor core and 1246 cells of surrounding tissue from Gene Expression Omnibus (GEO) with accession number of GSE84465 (Darmanis et al., 2017). 23,460 genes were measured using Illumina NextSeq 500. Within each sample, we counted the number of expressed genes, i.e., the number of genes with mapped reads. The average number of expressed genes in each sample was 2,581. Our goal is to discriminate the 2343 tumor cells (positive samples) and 1246 surrounding cells (negative samples).

The mRMR Ranking of Discriminative Genes

There have been many statistics methods for identifying the differentially expressed genes (DEGs). But these methods did not consider the relationships between genes. Usually, the number of DEGs was too large to apply as biomarker. Therefore, we adopted the information theory-based mRMR (minimal Redundancy Maximal Relevance) method (Peng et al., 2005) to overcome this problem. The mRMR method not only considers the associations between genes and samples, but also the redundancy between genes. If several genes are similar, only the most representative gene will be selected. This approach has been proven to be effective and has been widely used for many biomedical feature selection problems (Niu et al., 2013; Zhao et al., 2013; Zhou et al., 2015; Zhang et al., 2016; Liu et al., 2017), especially in single cell RNA-Seq analysis (Zhang et al., 2019). The sample size of single cell data was large and the gene expression was spare. It was easy to get too many redundant significant genes using traditional statistical based method, such as t-test. Therefore, the mRMR was suitable for analyzing single cell data to get small number of non-redundant biomarkers.

Let's describe the method mathematically. All genes, selected genes, to be selected genes can be represented as Ω, Ω_s, and Ω_t, respectively. The relevance of gene g from Ω_t with tissue type t can be measured with mutual information (I) (Sun et al., 2012; Huang and Cai, 2013):

$$D = I(g, t). \qquad (1)$$

And the redundancy R of the gene g with the selected genes in Ω_s are

$$R = \frac{1}{m}\left(\sum_{g_i \in \Omega_s} I(g, g_i)\right) \qquad (2)$$

The goal of this algorithm is to get the gene g_j from Ω_t that has maximum relevance with tissue type t and minimum redundancy with the selected genes in Ω_s, i.e., maximize the mRMR function

$$\max_{g_j \in \Omega_t}\left[I(g_j, t) - \frac{1}{m}\left(\sum_{g_i \in \Omega_s} I(g_j, g_i)\right)\right] \quad (j = 1, 2, \ldots, n) \quad (3)$$

The evaluation procedure will be continued for N rounds, and all the genes will be ranked as a list

$$S = \{g'_1, g'_2, \ldots, g'_h, \ldots, g'_N, \} \qquad (4)$$

The index h reflects the trade-off between relevance with tissue type and redundancy with selected genes. The smaller index h is, the better discriminating power the gene has.

The Single Cell GBM Biomarker Optimization

Based on the top 100 mRMR genes, we constructed 100 SVM classifiers and applied an incremental feature selection (IFS) method (Jiang et al., 2013; Li et al., 2014; Shu et al., 2014; Zhang et al., 2014, 2015) to identify the optimal number of genes as biomarker. The svm function from R package e10171 was used to implement the SVM method. Each candidate gene set $S_k = \{g'_1, g'_2, \ldots, g'_k\}(1 \leq k \leq 100)$ included the top k genes in the mRMR list.

We used leave-one-out cross validation (LOOCV) (Cui et al., 2013; Yang et al., 2014) to evaluate the prediction performance of each SVM classifier. During LOOCV, all of the N samples were tested one-by-one. In each round, one sample was used for testing of the prediction model trained with all the other $N-1$ samples. After N rounds, all samples were tested one time, and the predicted tissue types were compared with the actual tissue types.

Since the positive and negative sample sizes were imbalance and Mathew's correlation coefficient (MCC) can consider both sensitivity and specificity (Huang et al., 2015), MCC was used in IFS optimization. MCC can be calculated as follows:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \qquad (5)$$

where TP, TN, FP, and FN stand for true positive, true negative, false positive, and false negative, respectively.

Based on the LOOCV MCC of each candidate gene set, an IFS curve can be plotted. The x-axis denoted the number of top genes that were used in the SVM classifier, and the y-axis denoted the LOOCV MCCs of the SVM classifiers. Based on the IFS curve, we can choose the right number of genes which had a good prediction performance as final biomarkers.

RESULTS AND DISCUSSION

The Discriminative Importance of Genes

We applied mRMR algorithm to evaluate the discriminative importance of features iteratively. We want to find the features that were strongly associated with samples groups and were not redundant with other selected features. Using the mRMR method, we identified the top 100 most important genes. These genes were listed in **Supplementary Table S1**.

The Optimal GBM Biomarker Genes Selected With IFS Method

After we got the top 100 mRMR genes, we still did not know how many genes should be selected. To optimize the selected biomarker genes, we adopted IFS method. Each time, we added one feature into the previous feature set and got a new feature set. Then SVM classifiers were built to predict each sample's labels during LOOCV. The IFS curve with the number of genes as x-axis and the prediction performance (LOOCV MCC) as y-axis were plotted in **Figure 1**. The peak MCC was 0.812 when 31 genes were used. These 31 genes were selected as optimal GBM biomarker genes. The 31 genes were listed in **Table 1**. The confusion matrix of the 31 genes were given in **Table 2**. The sensitivity, specificity, and accuracy were 0.948, 0.855, and 0.915, respectively.

Since the tumor tissues are usually a mixture of tumor cells and normal cells, the tumor purity may cause the misclassifications. To check this, **Figures 2A,B** showed the t-distributed stochastic neighbor embedding (t-SNE) plots of predicted GBM cells and predicted non-GBM cells, respectively. In **Figure 2A**, it can be seen that the false positive samples (red dots) and the true positive samples (black dots) were mixed and they were difficult to classify. Similarly, in **Figure 2B**, it can be seen that the false negative samples (black dots) and the true negative samples (red dots) were mixed. These t-SNE plots suggested that the GBM tissues may contain non-GBM cells and the non-GBM tissues may contain GBM cells, but most cells from the corresponding tissue were similar and the machine learning algorithm we used can get the robust single cell biomarkers even when there were tissue purity issues.

The Biological Functions of the Selected Genes

Upon analysis by the mRMR method, 31 important genes were extracted that may be essential biomarkers of GBM. We did Gene

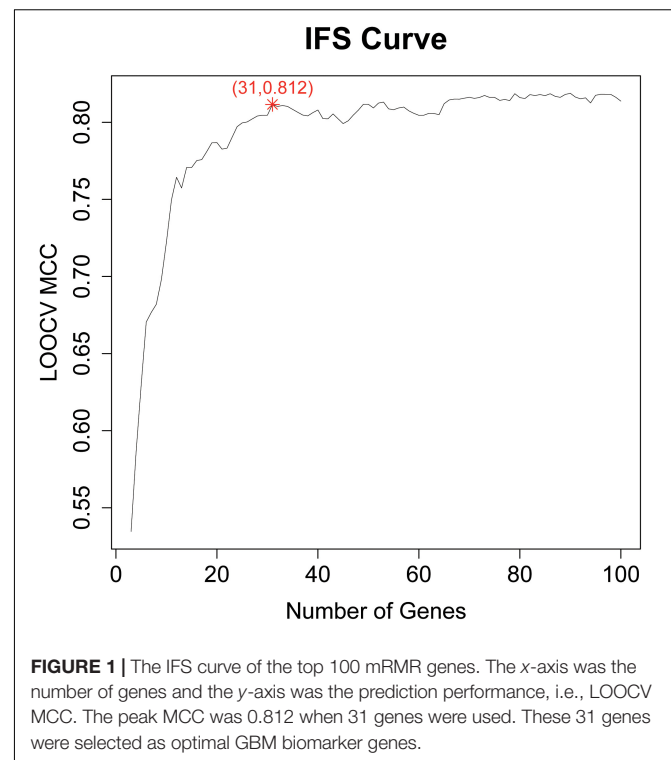**FIGURE 1 |** The IFS curve of the top 100 mRMR genes. The x-axis was the number of genes and the y-axis was the prediction performance, i.e., LOOCV MCC. The peak MCC was 0.812 when 31 genes were used. These 31 genes were selected as optimal GBM biomarker genes.

TABLE 1 | The 31 selected GBM biomarker genes.

Rank	Gene	Rank	Gene
1	TMSB4X	17	VIM
2	IPCEF1	18	ATP1A2
3	MTSS1	19	RPL41
4	S100A10	20	EGR3
5	HTRA1	21	OMG
6	DHRS9	22	LDHA
7	TPI1	23	P2RY12
8	SNX22	24	SPOCK1
9	FCGBP	25	NAMPT
10	TMSB10	26	C1QL2
11	CCL3	27	PTN
12	SLC6A1	28	CCL4
13	SMOC1	29	PDZD2
14	SEC61G	30	LGALS1
15	TGFBI	31	CLDN10
16	CDR1		

TABLE 2 | The confusion matrix of the 31 selected genes.

	Predicted GBM	Predicted non-GBM
Actual GBM	2220	123
Actual non-GBM	181	1065

TABLE 3 | The GO enrichment results of the 31 selected genes.

GO term	FDR	P-value	Genes
GO:0007155 cell adhesion	0.0068	$8.26E-07$	EGR3, LGALS1, OMG, PTN, S100A10, CCL4, SPOCK1, TGFBI, CLDN10, MTSS1, PDZD2, P2RY12
GO:0022610 biological adhesion	0.0068	$8.74E-07$	EGR3, LGALS1, OMG, PTN, S100A10, CCL4, SPOCK1, TGFBI, CLDN10, MTSS1, PDZD2, P2RY12
GO:0031012 extracellular matrix	0.0029	$1.57E-06$	LGALS1, OMG, HTRA1, PTN, SPOCK1, TGFBI, VIM, SMOC1
GO:0005615 extracellular space	0.0107	$1.56E-05$	LGALS1, OMG, HTRA1, PTN, CCL3, CCL4, SPOCK1, TGFBI, TMSB4X, TPI1, NAMPT
GO:0005576 extracellular region	0.0107	$1.87E-05$	ATP1A2, LDHA, LGALS1, OMG, HTRA1, PTN, S100A10, CCL3, CCL4, SPOCK1, TGFBI, TMSB4X, TPI1, VIM, FCGBP, NAMPT, PDZD2, SMOC1, C1QL2
GO:0005578 proteinaceous extracellular matrix	0.0107	$2.30E-05$	LGALS1, OMG, PTN, SPOCK1, TGFBI, SMOC1
GO:0044421 extracellular region part	0.0108	$2.89E-05$	ATP1A2, LDHA, LGALS1, OMG, HTRA1, PTN, S100A10, CCL3, CCL4, SPOCK1, TGFBI, TMSB4X, TPI1, VIM, FCGBP, NAMPT, SMOC1

Ontology (GO) enrichment analysis of these 31 genes. The GO enrichment results were given in **Table 3**. It can be seen that their main function was cell adhesion and their main subcellular location was extracellular.

We compared the 31 genes with reported GBM signatures in GeneSigDB (Culhane et al., 2012) and found that the 31 genes were significantly overlapped with a signature called "Human

FIGURE 2 | The *t*-SNE plots of predicted GBM cells and predicted non-GBM cells. **(A)** The *t*-SNE plots of predicted GBM cells. It can be seen that the false positive samples (red dots) and the true positive samples (black dots) were mixed and they were difficult to classify. **(B)** The *t*-SNE plots of predicted non-GBM cells. It can be seen that the false negative samples (black dots) and the true negative samples (red dots) were mixed. These *t*-SNE plots suggested that the GBM tissues may contain non-GBM cells and the non-GBM tissues may contain GBM cells, but most cells from the corresponding tissue were similar and the machine learning algorithm we used can get the robust single cell biomarkers even when there were tissue purity issues.

Glioblastoma_Morandi08_22genes" which were from Table 1 of Morandi et al. (2008): the 22 up-regulated genes following camptothecin (CPT) treatment in both U87-MG and DBTRG-05 cells. The hypergeometric test p-value was 0.0157.

Among the 31 genes, several of them plays roles in tumor metastasis. Thymosin β4 (TMSB4X/Tβ4) is associated with tumor metastasis and progression which plays a role in cell proliferation, migration, and differentiation through a TGFβ/MRTF Signaling Axis (Morita and Hayashi, 2018). TMSB4X expression was associated with cancers in a stage- and histology-specific manner and could be an effective prognostic parameter and prognostic index. Thus far, the relationship between TMSB4X and GBM remain unknown. IPCEF1 is the C-terminal half of CNK3 which is required for HGF-dependent Arf6 activation and migration during cancer metastasis (Attar et al., 2012). MTSS1 plays an important role in cancer metastasis. Previous researches indicated that MTSS1 as a potential tumor biomarker and its reduced expression associated with bad prognosis in many cancers. In GBM, MTSS1was reported as a potential tumor suppressor and prognostic biomarker which could suppress cell migration and invasion (Zhang and Qi, 2015).

Several genes can facilitate cancer progression. S100A10 is a calcium binding protein which is found to be significantly correlated with poor survival in patients with gliomas (Sethi et al., 2012). S100A10 has been involved in cancer progression, but the unique function is not well understood (O'Connell et al., 2010). HTRA1 encodes a ubiquitously expressed serine protease with prominent expression in the vasculature. Inhibition of HTRA1 could deregulate angiogenesis in the tumor stroma which plays an important role in tumor progression (Chien et al., 2006; He et al., 2010; Klose et al., 2018).

There are several other reported tumor genes. DHRS9 is a member of the short-chain dehydrogenases/reductases (SDR) family. Recent research found that SDR family members have been involved in tumors (Hu et al., 2016). TPI1 encodes an enzyme, consisting of two identical proteins, which catalyzes the isomerization of glyceraldehydes-3-phosphate (G3P) and dihydroxy-acetone phosphate (DHAP) in glycolysis and gluconeogenesis. TPI1 was down-regulated in response to LLL12 treatment and validated using immunoblot (Jain et al., 2015). It may serve as potential therapeutic targets in GBM (Jain et al., 2015).

CONCLUSION

Glioblastoma is the most aggressive and incurable primary brain cancer in adults. The most common survival time after diagnosis is 12–15 months, with 5-year survival rate <5%. Symptoms of GBM are non-specific at early stage and the cause of GBM remains elusive. We analysis the data from 2343 tumor cells and 1246 periphery cells using mRMR and IFS method to characterize infiltrating tumor cells, and to define the cellular diversity.

DATA AVAILABILITY STATEMENT

The datasets generated for this study can be found in the https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE84465.

AUTHOR CONTRIBUTIONS

S-SF and QC conceived and designed the study. QC, JL, Z-YD, and S-SF performed the data mining and statistical analyses. FF, HC, and Z-YW prepared the figures and tables. QC and JL drafted the initial manuscript. S-SF made critical comments and revision for the initial manuscript. S-SF, QC, and JL had primary responsibility for the final content. All authors reviewed and approved the final manuscript.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fbioe.2020.00167/full#supplementary-material

TABLE S1 | The top 100 mRMR genes.

REFERENCES

Attar, M. A., Salem, J. C., Pursel, H. S., and Santy, L. C. (2012). CNK3 and IPCEF1 produce a single protein that is required for HGF dependent Arf6 activation and migration. *Exp. Cell Res.* 318, 228–237. doi: 10.1016/j.yexcr.2011.10.018

Bajikar, S. S., Wang, C. C., Borten, M. A., Pereira, E. J., Atkins, K. A., and Janes, K. A. (2017). Tumor-suppressor inactivation of GDF11 occurs by precursor sequestration in triple-negative breast cancer. *Dev. Cell* 43, 418–435.e13. doi: 10.1016/j.devcel.2017.10.027

Cancer Genome Atlas Research Network (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061–1068. doi: 10.1038/nature07385

Chien, J., Aletti, G., Baldi, A., Catalano, V., Muretto, P., Keeney, G. L., et al. (2006). Serine protease HtrA1 modulates chemotherapy-induced cytotoxicity. *J. Clin. Invest.* 116, 1994–2004. doi: 10.1172/JCI27698

Chinot, O. L., Wick, W., Mason, W., Henriksson, R., Saran, F., Nishikawa, R., et al. (2014). Bevacizumab plus radiotherapy-temozolomide for newly diagnosed glioblastoma. *N. Engl. J. Med.* 370, 709–722. doi: 10.1056/NEJMoa1308345

Cui, W., Chen, L., Huang, T., Gao, Q., Jiang, M., Zhang, N., et al. (2013). Computationally identifying virulence factors based on KEGG pathways. *Mol. Biosyst.* 9, 1447–1452. doi: 10.1039/c3mb70024k

Culhane, A. C., Schröder, M. S., Sultana, R., Picard, S. C., Martinelli, E. N., Kelly, C., et al. (2012). GeneSigDB: a manually curated database and resource for

analysis of gene expression signatures. *Nucleic Acids Res.* 40, D1060–D1066. doi: 10.1093/nar/gkr901

Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., et al. (2017). Single-cell RNA-seq analysis of infiltrating neoplastic cells at the migrating front of human glioblastoma. *Cell Rep.* 21, 1399–1410. doi: 10.1016/j.celrep.2017.10.030

Das, S., and Marsden, P. A. (2013). Angiogenesis in glioblastoma. *N. Engl. J. Med.* 369, 1561–1563. doi: 10.1056/NEJMcibr1309402

Ding, C. H., and Dubchak, I. (2001). Multi-class protein fold recognition using support vector machines and neural networks. *Bioinformatics* 17, 349–358. doi: 10.1093/bioinformatics/17.4.349

Furnari, F. B., Fenton, T., Bachoo, R. M., Mukasa, A., Stommel, J. M., Stegh, A., et al. (2007). Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev.* 21, 2683–2710. doi: 10.1101/gad.1596707

Gilbert, M. R., Dignam, J. J., Armstrong, T. S., Wefel, J. S., Blumenthal, D. T., Vogelbaum, M. A., et al. (2014). A randomized trial of bevacizumab for newly diagnosed glioblastoma. *N. Engl. J. Med.* 370, 699–708. doi: 10.1056/NEJMoa1308573

He, X., Ota, T., Liu, P., Su, C., Chien, J., and Shridhar, V. (2010). Downregulation of HtrA1 promotes resistance to anoikis and peritoneal dissemination of ovarian cancer cells. *Cancer Res.* 70, 3109–3118. doi: 10.1158/0008-5472.CAN-09-3557

Hu, L., Chen, H. Y., Han, T., Yang, G. Z., Feng, D., Qi, C. Y., et al. (2016). Downregulation of DHRS9 expression in colorectal cancer tissues and its prognostic significance. *Tumour Biol.* 37, 837–845. doi: 10.1007/s13277-015-3880-6

Huang, T., and Cai, Y.-D. (2013). An information-theoretic machine learning approach to expression QTL analysis. *PLoS One* 8:e67899. doi: 10.1371/journal.pone.0067899

Huang, T., Wang, M., and Cai, Y.-D. (2015). Analysis of the preferences for splice codes across tissues. *Protein Cell* 6, 904–907. doi: 10.1007/s13238-015-0226-5

Jain, R., Kulkarni, P., Dhali, S., Rapole, S., and Srivastava, S. (2015). Quantitative proteomic analysis of global effect of LLL12 on U87 cell's proteome: an insight into the molecular mechanism of LLL12. *J. Proteomics* 113, 127–142. doi: 10.1016/j.jprot.2014.09.020

Jiang, Y., Huang, T., Chen, L., Gao, Y. F., Cai, Y., and Chou, K. C. (2013). Signal propagation in protein interaction network during colorectal cancer progression. *Biomed Res. Int.* 2013:287019. doi: 10.1155/2013/287019

Klose, R., Adam, M. G., Weis, E. M., Moll, I., Wustehube-Lausch, J., Tetzlaff, F., et al. (2018). Inactivation of the serine protease HTRA1 inhibits tumor growth by deregulating angiogenesis. *Oncogene* 37, 4260–4272. doi: 10.1038/s41388-018-0258-4

Li, B. Q., You, J., Huang, T., and Cai, Y. D. (2014). Classification of non-small cell lung cancer based on copy number alterations. *PLoS One* 9:e88300. doi: 10.1371/journal.pone.0088300

Li, Z., Guo, J., Ma, Y., Zhang, L., and Lin, Z. (2018). Oncogenic role of MicroRNA-30b-5p in glioblastoma through targeting proline-rich transmembrane protein 2. *Oncol. Res.* 26, 219–230. doi: 10.3727/096504017x14944585873659

Liu, J., Albrecht, A. M., Ni, X., Yang, J., and Li, M. (2013). Glioblastoma tumor initiating cells: therapeutic strategies targeting apoptosis and microRNA pathways. *Curr. Mol. Med.* 13, 352–357. doi: 10.2174/15665241380507 6830

Liu, L., Chen, L., Zhang, Y. H., Wei, L., Cheng, S., Kong, X., et al. (2017). Analysis and prediction of drug-drug interaction by minimum redundancy maximum relevance and incremental feature selection. *J. Biomol. Struct. Dyn.* 35, 312–329. doi: 10.1080/07391102.2016.1138142

Morandi, E., Severini, C., Quercioli, D., D'Ario, G., Perdichizzi, S., Capri, M., et al. (2008). Gene expression time-series analysis of camptothecin effects in U87-MG and DBTRG-05 glioblastoma cell lines. *Mol. Cancer* 7:66. doi: 10.1186/1476-4598-7-66

Morita, T., and Hayashi, K. (2018). Tumor progression is mediated by thymosin-beta4 through a TGFbeta/MRTF signaling axis. *Mol. Cancer Res.* 16, 880–893. doi: 10.1158/1541-7786.MCR-17-0715

Muller, S., Kohanbash, G., Liu, S. J., Alvarado, B., Carrera, D., Bhaduri, A., et al. (2017). Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* 18:234. doi: 10.1186/s13059-017-1362-4

Niu, B., Huang, G., Zheng, L., Wang, X., Chen, F., Zhang, Y., et al. (2013). Prediction of substrate-enzyme-product interaction based on molecular descriptors and physicochemical properties. *BioMed Res. Int.* 2013:674215. doi: 10.1155/2013/674215

O'Connell, P. A., Surette, A. P., Liwski, R. S., Svenningsson, P., and Waisman, D. M. (2010). S100A10 regulates plasminogen-dependent macrophage invasion. *Blood* 116, 1136–1146. doi: 10.1182/blood-2010-01-264754

Ostrom, Q. T., Gittleman, H., Xu, J., Kromer, C., Wolinsky, Y., Kruchko, C., et al. (2016). CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2009-2013. *Neuro Oncol.* 18(Suppl. 5), v1–v75. doi: 10.1093/neuonc/now207

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401. doi: 10.1126/science.1254257

Peng, H., Long, F., and Ding, C. (2005). Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* 27, 1226–1238. doi: 10.1109/tpami.2005.159

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., et al. (2017). Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* 171, 1611–1624.e24. doi: 10.1016/j.cell.2017.10.044

Sethi, M. K., Buettner, F. F., Ashikov, A., Krylov, V. B., Takeuchi, H., Nifantiev, N. E., et al. (2012). Molecular cloning of a xylosyltransferase that transfers the second xylose to O-glucosylated epidermal growth factor repeats of notch. *J. Biol. Chem.* 287, 2739–2748. doi: 10.1074/jbc.M111.302406

Shu, Y., Zhang, N., Kong, X., Huang, T., and Cai, Y. D. (2014). Predicting A-to-I RNA editing by feature selection and random forest. *PLoS One* 9:e110607. doi: 10.1371/journal.pone.0110607

Stegle, O., Teichmann, S. A., and Marioni, J. C. (2015). Computational and analytical challenges in single-cell transcriptomics. *Nat. Rev. Genet.* 16, 133–145. doi: 10.1038/nrg3833

Stupp, R., Hegi, M. E., Gilbert, M. R., and Chakravarti, A. (2007). Chemoradiotherapy in malignant glioma: standard of care and future directions. *J. Clin. Oncol.* 25, 4127–4136. doi: 10.1200/JCO.2007.11.8554

Stupp, R., Mason, W. P., van den Bent, M. J., Weller, M., Fisher, B., Taphoorn, M. J., et al. (2005). Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N. Engl. J. Med.* 352, 987–996. doi: 10.1056/NEJMoa04 3330

Stupp, R., Taillibert, S., Kanner, A., Read, W., Steinberg, D. M., Lhermitte, B., et al. (2017). Effect of tumor-treating fields plus maintenance temozolomide vs maintenance temozolomide alone on survival in patients with glioblastoma: a randomized clinical trial. *JAMA* 318, 2306–2316. doi: 10.1001/jama.2017.18718

Sun, L., Yu, Y., Huang, T., An, P., Yu, D., Yu, Z., et al. (2012). Associations between ionomic profile and metabolic abnormalities in human population. *PLoS One* 7:e38845. doi: 10.1371/journal.pone.0038845

Verhaak, R. G., Hoadley, K. A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M. D., et al. (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110. doi: 10.1016/j.ccr.2009.12.020

Xiao, S., Yang, Z., Lv, R., Zhao, J., Wu, M., Liao, Y., et al. (2014). miR-135b contributes to the radioresistance by targeting GSK3beta in human glioblastoma multiforme cells. *PLoS One* 9:e108810. doi: 10.1371/journal.pone.0108810

Yang, J., Chen, L., Kong, X., Huang, T., and Cai, Y. D. (2014). Analysis of tumor suppressor genes based on gene ontology and the KEGG pathway. *PLoS One* 9:e107202. doi: 10.1371/journal.pone.0107202

Zhang, G.-L., Pan, L.-L., Huang, T., and Wang, J.-H. (2019). The transcriptome difference between colorectal tumor and normal tissues revealed by single-cell sequencing. *J. Cancer* 10, 5883–5890. doi: 10.7150/jca.32267

Zhang, N., Huang, T., and Cai, Y. D. (2014). Discriminating between deleterious and neutral non-frameshifting indels based on protein interaction networks and hybrid properties. *Mol. Genet. Genomics* 290, 343–352. doi: 10.1007/s00438-014-0922-5

Zhang, N., Wang, M., Zhang, P., and Huang, T. (2016). Classification of cancers based on copy number variation landscapes. *Biochim. Biophys. Acta* 1860(11 Part B), 2750–2755. doi: 10.1016/j.bbagen.2016.06.003

Zhang, P. W., Chen, L., Huang, T., Zhang, N., Kong, X. Y., and Cai, Y. D. (2015). Classifying ten types of major cancers based on reverse phase protein array profiles. *PLoS One* 10:e0123147. doi: 10.1371/journal.pone.0123147

Zhang, S., and Qi, Q. (2015). MTSS1 suppresses cell migration and invasion by targeting CTTN in glioblastoma. *J. Neurooncol.* 121, 425–431. doi: 10.1007/s11060-014-1656-2

Zhao, T. H., Jiang, M., Huang, T., Li, B. Q., Zhang, N., Li, H. P., et al. (2013). A novel method of predicting protein disordered regions based on sequence features. *BioMed Res. Int.* 2013:414327. doi: 10.1155/2013/414327

Zhao, X., Gao, S., Wu, Z., Kajigaya, S., Feng, X., Liu, Q., et al. (2017). Single-cell RNA-seq reveals a distinct transcriptome signature of aneuploid hematopoietic cells. *Blood* 130, 2762–2773. doi: 10.1182/blood-2017-08-803353

Zhou, Y., Zhang, N., Li, B. Q., Huang, T., Cai, Y. D., and Kong, X. Y. (2015). A method to distinguish between lysine acetylation and lysine ubiquitination

with feature selection and analysis. *J. Biomol. Struct. Dyn.* 33, 2479–2490. doi: 10.1080/07391102.2014.1001793