Research Paper

Predicting chromosome 1p/19q codeletion by RNA expression profile: a comparison of current prediction models

Zhi-liang Wang[1], Zheng Zhao[1], Zheng Wang[2], Chuan-bao Zhang[2], Tao Jiang[1,2,3,4]

[1]Beijing Neurosurgical Institute, Capital Medical University, Beijing, China
[2]Department of Neurosurgery, Beijing Tiantan Hospital, Capital Medical University, Beijing, China
[3]China National Clinical Research Center for Neurological Diseases, Beijing, China
[4]Center of Brain Tumor, Beijing Institute for Brain Disorders, Beijing, China

Correspondence to: Tao Jiang, Chuan-bao Zhang, Zheng Wang; **email:** taojiang1964@163.com, chuanbao123@139.com, wangzheng1024@126.com

Keywords: oligodendroglioma, 1p/19q codeletion, prediction, smoother, method
Received: October 20, 2018 **Accepted:** January 24, 2019 **Published:** February 2, 2019

ABSTRACT

Background: Chromosome 1p/19q codeletion is increasingly being recognized as the crucial genetic marker for glioma patients and have been included in WHO classification of glioma in 2016. Fluorescent in situ hybridization, a widely used method in detecting 1p/19q status, has some methodological limitations which might influence the clinical management for doctors. Here, we attempted to explore an RNA sequencing computational method to detect 1p/19q status.

Methods: We included 692 samples with 1p/19q status information from TCGA cohort as training set and 222 samples with 1p/19q status information from REMBRANDT cohort as validation set. We reviewed and compared five tools: TSPairs, GSVA, PAM, Caret, smoother, with respect to their accuracy, sensitivity and specificity.

Results: In TCGA cohort, the GSVA method showed the highest accuracy (98.4%) in predicting 1p/19q status (sensitivity=95.5%, specificity=99.6%) and smoother method showed the second-highest accuracy (accuracy=97.8%, sensitivity=96.4%, specificity=98.3%). While in REMBRANDT cohort, smoother method exhibited the highest accuracy (98.6%) (sensitivity= 96.7%, specificity=98.9%) in 1p/19q status prediction.

Conclusions: Our independent assessment of five tools revealed that smoother method was selected as the most stable and accurate method in predicting 1p/19q status. This method could be regarded as a potential alternative method for clinical practice in future.

INTRODUCTION

Glioma is the most common and deadliest malignant primary brain tumor in adults [1]. Oligodendroglial tumors, including oligodendrogliomas and oligoastrocytomas, are the second common type of glioma [2–5]. Chromosome 1p/19q codeletion, complete deletion of both the short arm of chromosome 1 and the long arm of chromosome 19, is the specific hallmark of oligodendrogliomas. The frequency of this genetic aberration in oligoastrocytoma and oligodendroglioma are 50%~ 70% and 70% ~80%, respectively [6].

Nowadays, 1p/19q codeletion is increasingly recognized as a crucial genetic aberration in glioma patients and was first time included in the WHO classification of brain tumor in 2016 [7]. This pathognomonic biomarker is thought to commonly occur in the early phase of glioma development [8]. In addition, numerous studies explored the clinical significance of 1p/19q codeletion and found that it is a strong independent favorable prognosticator of overall survival (OS) and progression free survival (PFS) for glioma patients [8-11], and patients with this aberration would benefit from radiation therapy plus chemotherapy in comparison with radiation therapy alone after surgery [10]. Hence, prediction of 1p/19q status accurate become particularly critical for the precision medical in glioma patients.

Fluorescent in situ hybridization (FISH), targeting 1p36/1p21 and 19q13/19p13 regions via fluorophore-labelled DNA probes [12], was used as standard protocol to detect 1p/19q status in most hospitals [13]. However, FISH has some methodological limitations neuro-pathologists need to be aware of in clinical practice. Firstly, probes designed for chromosome 1p and 19q span a long region that may not identify small interstitial and terminal deletions. Secondly, FISH may not detect hemizygous deletions if there is loss of one allele and reduplication of the other allele [14]. Thirdly, FISH analysis is time-consuming and subjective which requires experienced pathologist to ensure result accuracy [15]. The incorrect 1p/19q status detecting by FISH may cause improper treatment strategy for patients [16]. Moreover, using FISH to detect 1p/19q status exerts a financial burden on glioma patients and fails to get more genetic alterations information.

Nowadays, next generation RNA sequencing (RNA-seq) technologies greatly promote the exploration of the complex and dynamic nature of cancer [17] and could provide insights to previously undetected changes occurring in disease [18]. RNA sequencing data has been successfully applied in identifying single nucleotide variants mutation [19], alternative splicing [20], fusion genes [21] and RNA editing [22]. Comprehensive understanding of the gene expression profile variation caused by copy number variation provide us possibility to detect 1p/19q status.

However, the comprehensive study that integrated RNA-seq data analysis methods to predict 1p/19q status has not been conducted yet. Therefore, in this study, we reviewed and assessed five methods which were designed to detect gene expression variations with RNA-seq data, and attempted to find out a precise, objective and cost-effective method to replace FISH for identifying 1p/19q status in clinical practice in the future.

RESULTS

Chromosome 1p/19q co-deleted patients exhibited a distinct expression profile

In order to assess the feasibility of predicting 1p/19q status with RNA expression data, we used whole genome expression profiling (20501 genes) from TCGA dataset to explore the relationship between 1p/19q status and gene expression profiles. As shown in Supplementary Figure 1A, hierarchical cluster method separated the dendrogram into two branches. The first branch was consisted by five subgroups while the second branch contained only one subgroup. The majority of 1p/19q co-deleted patients were in group 1, 3 and 5, while the 1p/19q intact patients were mainly in the rest of groups (group 2, 4, 6). The result indicated that there were some obvious differences in expression profiles between 1p/19q intact patients and 1p/19q co-deleted patients. However, the classification process was interfered by noisy genes and it was hard for us to predict 1p/19q status clearly with raw RNA sequencing data. Then we tried to change the threshold of gene expression to improve the classification accuracy in detecting 1p/19q status. Previously studies used MAD value to evaluate highly variable expression genes with RNA sequencing data [2, 23]. Here, we used highly variable genes (MAD >2, 886 genes) for hierarchical clustering, and the distance between 1p/19q co-deleted samples (Supplementary Figure 1B) was closer than clustering with whole gene expression profiles (Supplementary Figure 1A). And the hierarchical assignment in clustering chromosome 1p and 19q genes (n=1775) expression profile exhibited a similar result (Supplementary Figure 1C). Those results indicated that the significant differences of RNA sequencing data between 1p/19q co-deleted and intact patients could provide a feasibility to predict 1p/19q status by RNA expression data.

Overview of five methods in predicting 1p/19q status

Based on this finding, we selected five methods which have been used to process RNA sequencing data to identify 1p/19q status.

TSPairs method which compared the two genes expression ratio was used in breast cancer and lung adenocarcinoma for classification and prognosis [24, 25]. In the training dataset, with the *tspcalc* function, HDAC1(Histone deacetylase 1) and DRG2 (Developmentally regulated GTP binding protein 2) gene pair which had the highest TSP score (0.962) was selected as the most consistent switch (Figure 1A). In 1p/19q co-deleted group, 89.6% (172/192) 1p/19q co-deleted patients had a lower expression of HDAC1 than DRG2, while 100% (500/500) 1p/19q intact patients had a lower expression of DRG2 than HDAC1 (Figure 1B, p=0.0063, chi-square test). And the Receiver Operating Characteristic (ROC) curve showed that this model achieved an area under the curve (AUC) of 0.9878 (Figure 1D). With this prediction model, in the validation dataset, we got a similar result (Figure 1C, p<0.0001, chi-square test) and the AUC was 0.9165 (Figure 1E). Previously studies showed that HDAC1, locating at chromosome 1p, might serve as a good diagnostic and prognostic marker for lung cancer [26]. Overexpressing DRG2, locating at chromosome 17p, could delay cell-cycle arrest and apoptosis [27].

GSVA method was designed for integrating genes that shared common biological functions or chromosomal locations and was widely used in cancer research [24, 28, 29]. Gene set enrichment score of genes on 1p and genes on 19q were calculated with *gsva* function, respectively. Hierarchical clustering was performed based on the enrichment scores in the training dataset. Two branches were identified: 95.5% (169/177) 1p/19q co-deleted samples were clustered into the first branch while 99.6% (513/515) 1p/19q intact samples were clustered into the second branch (Figure 2A). The ROC curves of the enrichment score of 1p and 19q showed an AUC of 0.97 (Figure 2B) and 0.845 (Figure 2C), respectively. In the validation dataset, two branches were identified: 81.6% (31/38) 1p/19q co-deleted samples were clustered into the first branch, while 98.9% (182/184) 1p/19q intact samples were clustered into the second branch (Figure 2D). In the validation dataset, ROC curves of the enrichment score of 1p and 19q showed an AUC of 0.595 (Figure 2E) and 0.567 (Figure 2F), respectively.

PAM method exhibited powerful predictive capabilities in rectal cancer by selecting a group of genes [30]. The centroid shrinkage value of 10.984 which containing a minimum of 53 genes and 7 misclassification errors was selected in the training dataset (Figure 3A). Then the RNA expression profile of the 53 genes in the two

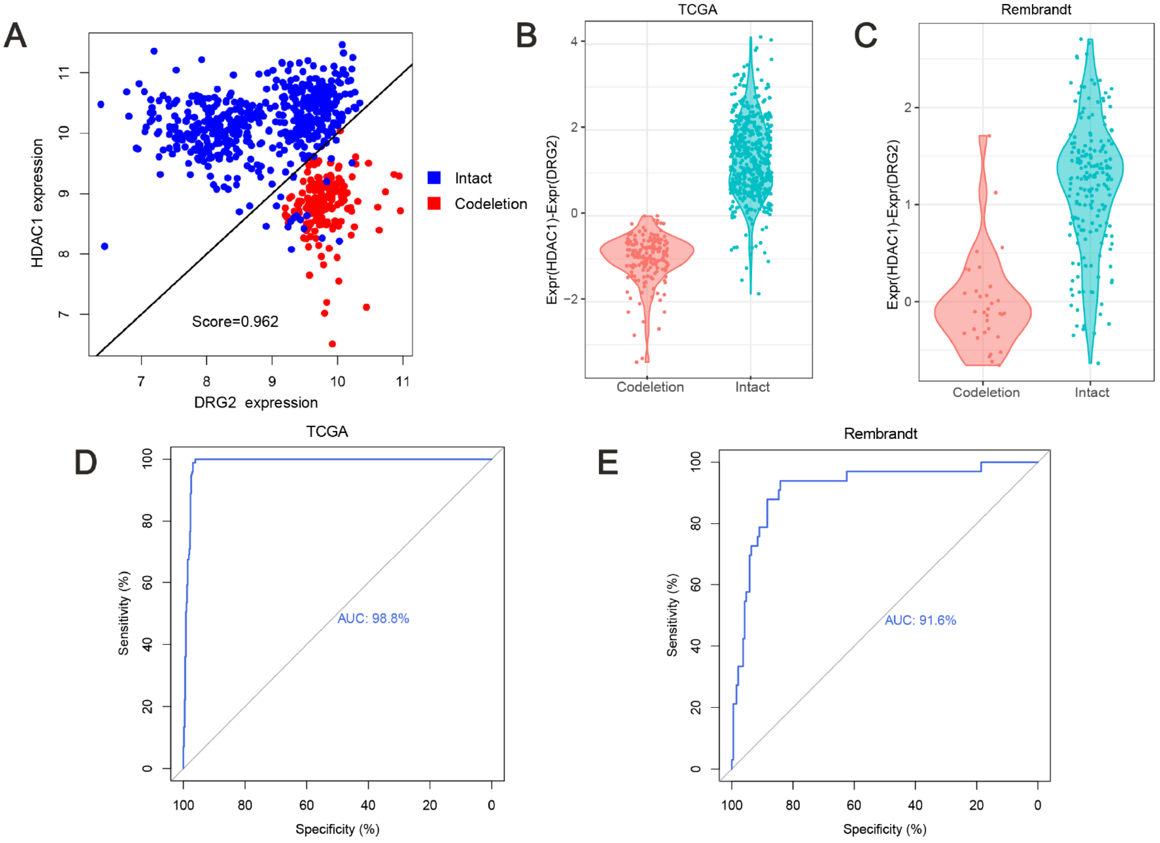


Figure 1. Predicting 1p/19q status by TSPair algorithm. (**A**) HDAC1 and DRG2 pair was the top scoring pair in predicting 1p/19q (score=0.962). The expression values of training set were normalized as Expr=log2(RSEM+1). (**B**) the values (HDAC1 - DRG2 expression values) were significantly different (p=0.0063) between 1p/19q co-deleted group and intact group in TCGA cohort. (**C**) the values (HDAC1 - DRG2 expression values) were significantly different (p<0.0001) between 1p/19q co-deleted group and intact group in Rembrandt cohort. (**D** and **E**) ROC curve for 1p/19q status prediction in TCGA cohort and Rembrandt cohort, AUC=0.988 and 0.916, respectively.

datasets were clustered. In the training dataset, two branches were identified: the first branch contained 91.0% (172/189) 1p/19q co-deleted samples, while the second branch contained 100% (503/503) 1p/19q intact patients (Figure 3A). In the validation cohort, two branches were identified as expected. The first branch contained 45.5% (25/55) 1p/19q co-deleted patients. Meanwhile, 98.8% (164/167) 1p/19q intact patients were grouped into the second branch (Figure 3B).

Caret method, a powerful machine learning algorithm, was designed for predictive modeling in practice with gene expression data and has been applied in predicting clinical outcome of patients with Alzheimer's disease [31]. In each cross validation, 20% samples in the training dataset were selected as TCGA-Train data to build predictive model and the rest of samples in the training dataset were grouped into TCGA-Test data to estimate the efficiency of the model. In the TCGA-Train dataset, the partial least squares discriminant analysis (PLSDA) method [32] was performed to build regression models and 10-fold cross-validation was used to examine the predictive efficiency [33]. As shown in ROC curve, the AUC was more than 0.999 in every repeated cross-validation cohort (Figure 4A) and the maximum value was 1.0 (ncomp=1). Then we predicted 1p/19q status with the best model. In the TCGA-Test dataset, 93.48% (43/46) 1p/19q co-deleted patients and 100% (126/126)1p/19q intact patients were successfully predicted. And in the validation dataset, 85.7% (30/35) 1p/19q co-deleted patients and 98.4% (184/187) 1p/19q intact patients were successfully predicted. The AUC values for TCGA-Test dataset and validation dataset were 0.985 (Figure 4B) and 0.966 (Figure 4C), respectively.

Smoother method, which could modify individual noise, have been performed in the analysis of esophageal squamous cell carcinoma [34] and breast cancer [35]. Firstly, genes expression data in 1p and 19q from start to end were smoothed by a 100 genes window. Then the combined 1p and 19q smoothed RNA expression profile were clustered. In the training dataset, two branches were clustered: 96.4% (163/169) 1p/19q co-deleted patients were clustered into the first branch, while 98.3% (514/523) 1p/19q intact patients were clustered into the second branch (Figure 5A). In the validation dataset, 96.9% (31/32) 1p/19q co-deleted patients were in the first branch, while 98.9% (188/190) 1p/19q intact patients were in the second branch (Figure 5B).

Comparing the prediction accuracy of five methods

Finally, we summarized the similarities and differences between five methods and evaluated the most appropriate method to predict 1p/19q status with ROC curves. The 1p/19q status results that identified by single nucleotide polymorphism (SNP) array were used as golden standard in our study. As shown in the Figure 6A a total of 183 patients were classified into 1p/19q co-deleted group

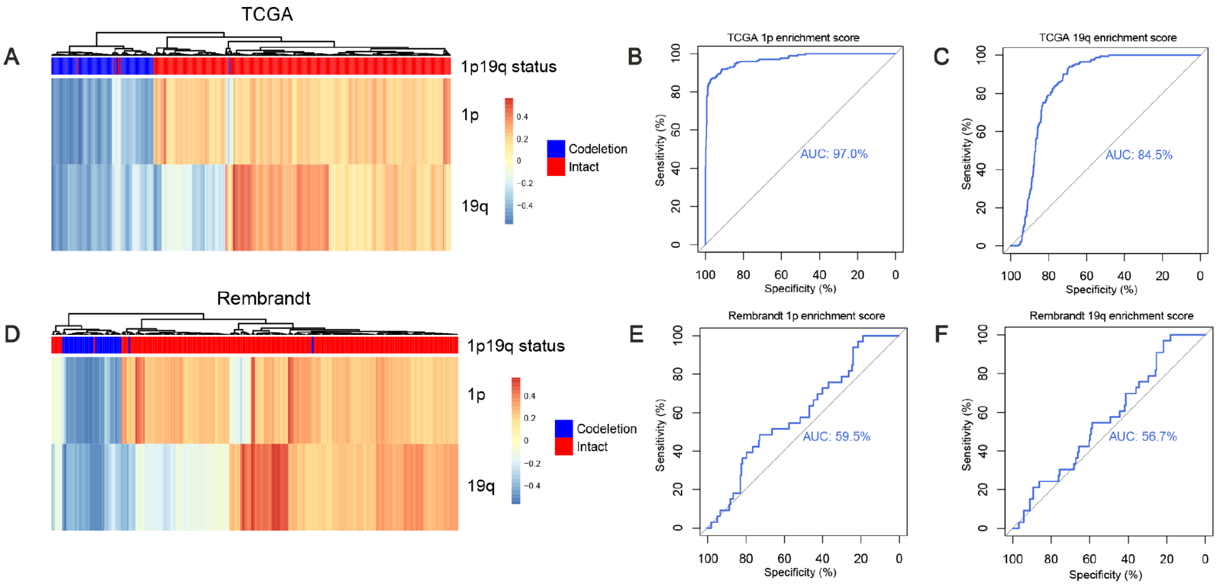


Figure 2. Predicting 1p/19q status by GSVA algorithm. (**A** and **D**) the hierarchical clustering of TCGA and Rembrandt cohorts based on the enrichment scores of 1p and 19q genes, respectively. (**B**, **E**) ROC for 1p/19q status prediction by 1p genes enrichment scores, AUC (TCGA cohort) = 0.970, AUC (Rembrandt cohort) = 0.595. (**C**, **F**) ROC for 1p/19q status prediction by 19q genes enrichment scores, AUC (TCGA cohort) = 0.845. AUC (Rembrandt cohort) = 0.567.

by five RNA processing methods and 181(88.3%) of them were identified by SNP array. While 99.7% (649/651) 1p/19q intact patients were identified by SNP array (Figure 6B).

Meanwhile, when integrated samples in TCGA and Rembrandt dataset, the smoother (96.5%) and PAM (82.3%) exhibited the highest sensitivity and lowest level of sensitivity among five methods, respectively (Figure 6C). And TSPair (97.9%) showed the lowest level of specificity. The sensitivity and specificity of five methods in predicting 1p/19q status in TCGA dataset exhibited similar trends (Figure 6D). Except TSPair (89.6%), the sensitivities of other methods were all more than 90% and smoother (96.5%) was still the highest one. The TSPair (100%) and PAM (100%) methods had the highest specificity while smoother method (98.3%) had the lowest specificity. However, in Rembrandt dataset, the sensitivities of GSVA (81.6%), PAM (45.5%), Caret (85.7%) and TSPair (64.5%) were all decreased obviously. Only smoother (sensitivity = 96.7%, specificity = 98.8%) method remained a higher accuracy in predicting the 1p/19q status.

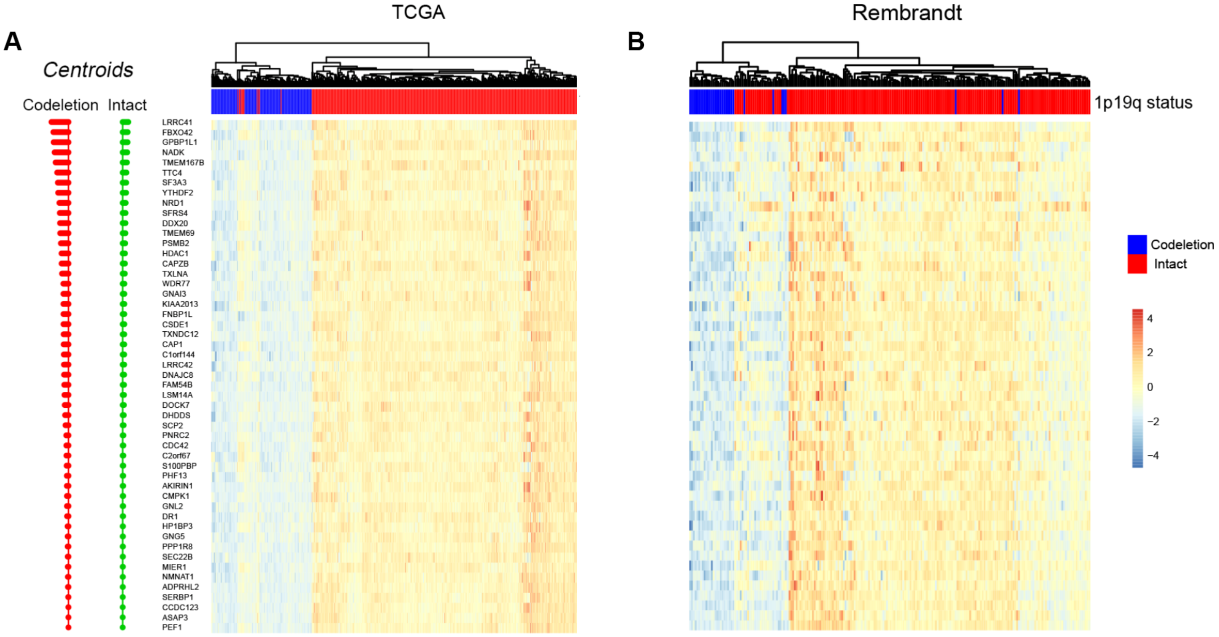


Figure 3. Predicting 1p/19q status by PAM algorithm. The hierarchical clustering samples in TCGA and Rembrandt cohorts using 53 signature genes, respectively.

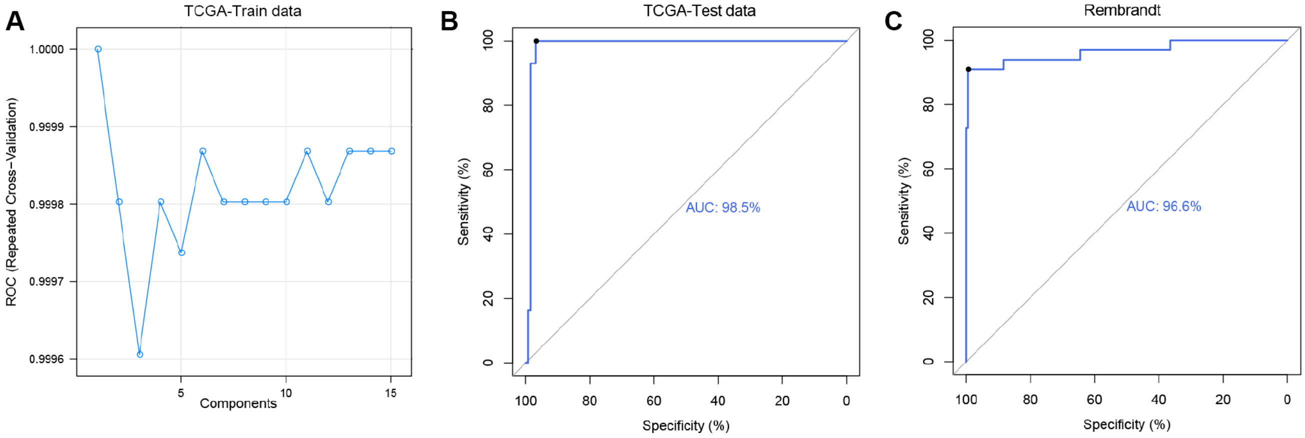


Figure 4. Predicting 1p/19q status by Caret algorithm. (**A**) The ROC values of 15 PLS models were compared to select the optimal prediction model (ncomp.1) using TCGA-Test data. (**B** and **C**) ROC curves for 1p/19q status prediction by applying the ncomp.1 model, AUC (TCGA-Test data) = 0.985, AUC (Rembrandt) = 0.966.

After synthetical comparing the differences of sensitivity, specificity and accuracy among five methods in two individual datasets, we found that smoother method clearly outperformed the other methods in predicting 1p/19q status with little room for improvement.

DISCUSSION

With the rapid expansion of multi-platform integrated analysis of glioma, molecular markers have greatly facilitated the understanding of the genetic progress

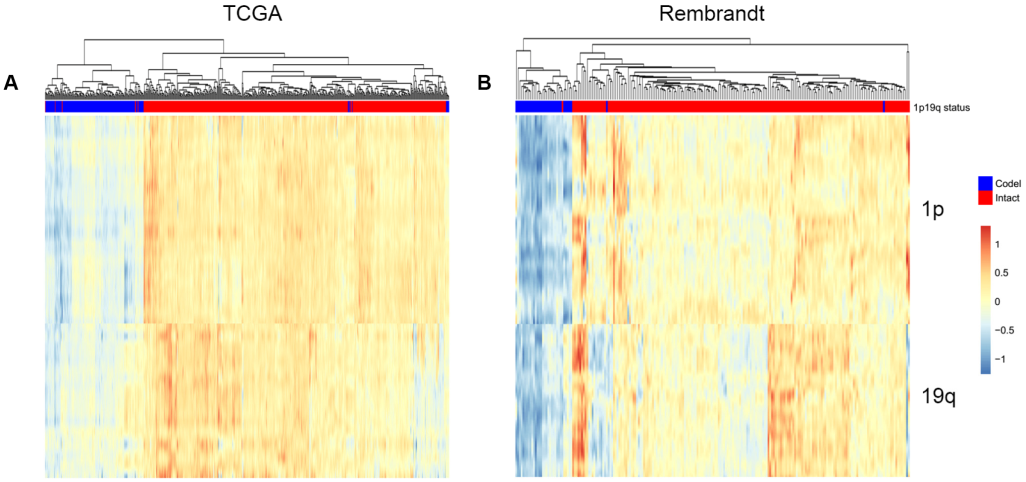


Figure 5. Predicting 1p/19q status by smoother algorithm. The hierarchical clustering of TCGA and Rembrandt cohorts by using smoothed gene expression on 1p and19q respectively.

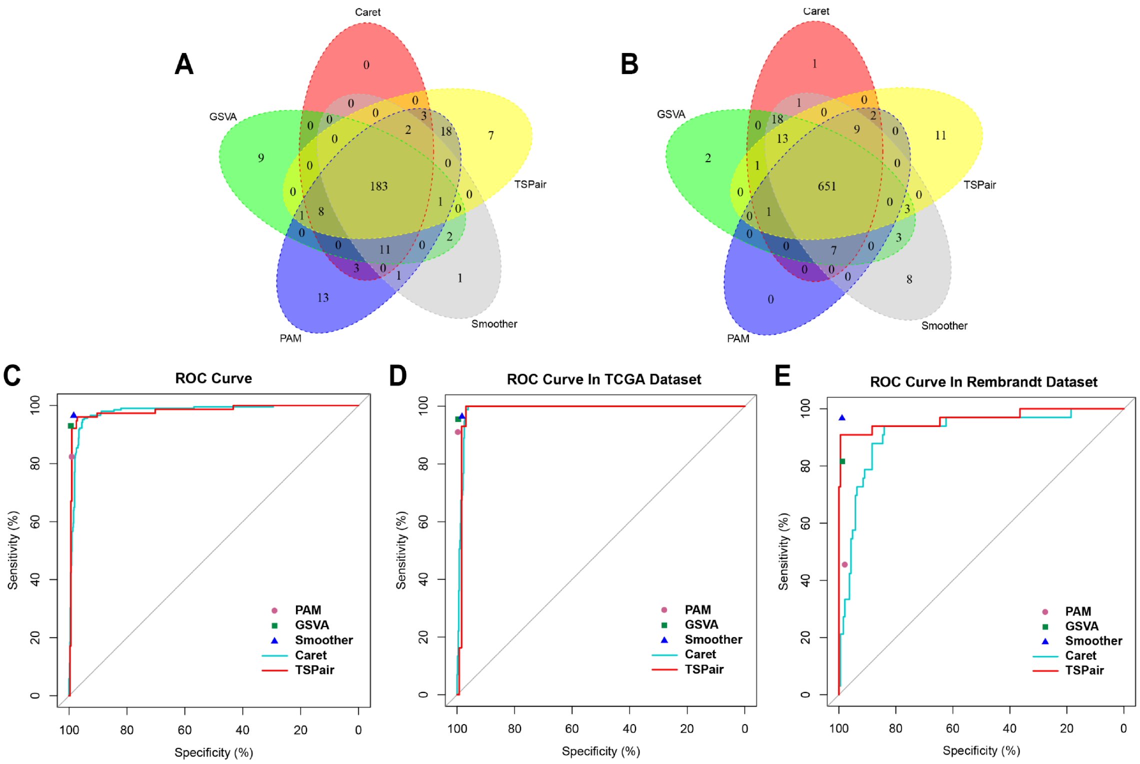


Figure 6. Comparing the sensitivity, specificity and accuracy of five algorithms. (**A**) Overlap among of the 1p/19q co-deleted samples found by five methods. (**B**) Overlap among of the 1p/19q intact samples found by five methods. (**C**) Comparing the sensitivity and specificity of five algorithms in predicting 1p/19q status in 914 samples. (**D**) Comparing the sensitivity and specificity of five algorithms in predicting 1p/19q codeletion in TCGA dataset. (**E**) Comparing the sensitivity and specificity of five algorithms in predicting 1p/19q codeletion in Rembrandt dataset.

underlying the progress of cancer and provided key insights on precision medicine. The status of chromosome 1p/19q is one of the most crucial molecular markers in glioma, which has shown well-established association with the diagnosis and prognosis of patients [36].

DNA sequencing, processing the order of nucleotides in DNA, is considered as the most accurate method in detecting large regions chromosome variation [37]. However, the cost to generate a high-quality DNA sequencing was almost $2,500 and the medical insurance does not cover the cost for genetic tests [38]. DNA sequencing as routine tool to detect 1p/19q status was limited due to the high cost. There is some uncertainly about detecting 1p/19q status with FISH method due to the limitations. Thus, quality assurance remains an issue, finding a more cost-effective and accurate way to obtain 1p/19q status information would greatly promote the widespread of genetic-guided precision medicine of glioma. Next-generation sequencing technologies have revolutionarily advanced genome-related research with the advantages of high-throughput, high-sensitivity, and low-cost [39]. RNA-seq is now being used widely for detecting molecular aberrations in cancer research [40]. It is well known that the copy number variations (CNVs) typically result in a corresponding gene expression changes, especially large chromosome amplification or deletion [41]. CNV-related gene expression changes supported us detecting 1p/19q status with RNA sequencing data.

The main purpose of this study was to explore an appropriate RNA sequencing computational method to detect 1p/19q status. In this study, several analysis pipelines have been applied to analyze gene expression data for predicting 1p/19q status. TSPairs method, which identifies an alteration by comparing ratio of two genes expression, has been used in breast cancer and lung adenocarcinoma for molecular classification and predicting prognosis [24, 25]. GSVA method is designed for interpreting gene expression data. By using GSVA method, scientists are able to get further insights into leukemia and lung cancer [42, 43]. PAM method exhibits powerful predictive capabilities in rectal cancer by selecting a group of genes [30]. Caret method relates the practice of predictive modeling and has been applied in predicting prognosis of Alzheimer's disease [31]. Smoother method is designed for removing the noise and scatter of RNA sequencing data [44]. Finally, smoother method was selected as the most stable and accurate method in TCGA and Rembrandt datasets.

There were some advantages in inferring 1p/19q status with gene expression data. Firstly, the cost in RNA sequencing is much less than FISH. Prediction 1p/19q status with RNA sequencing data could reduce financial burden for patients. Secondly, identifying 1p/19q status by gene expression data could eliminate the limitation of FISH probes of testing only two regions. Thirdly, the RNA sequencing data could also be applied to call or predict IDH mutation, ATRX mutation, TERT mutation, EGFRvIII deletion, fusion genes and so on. The integrated analysis of DNA sequencing technology, RNA sequencing technology and FISH were shown in Supplementary Table 1. The experiment in RNA sequencing used the shortest time and least amount of money.

However, several limitations should be considered in this research. Firstly, the prediction models based on five RNA sequencing data processing methods were established retrospectively. The prospective longitudinal study was need to estimate this research in future. Second, the consistency between FISH and prediction models in predicting 1p/19q status could not be evaluated due to the limitation of FISH information of glioma patients in TCGA database and REMBRANDT database.

In summary, along with the application of next-generation sequencing in clinical practice, we believed that RNA sequencing processing method will show great potential as the standard detection method to detect various genetic and molecular alterations. For detecting 1p/19q, we would recommend smoother method using RNA sequencing data, which was more cost-effective and convenient in clinical practice.

MATERIALS AND METHODS

Data collection

Training set: RNA-sequencing data of 692 glioma samples downloaded from The Cancer Genome Atlas (TCGA, http://cancergenome.nih.gov/), containing 172 1p/19q co-deleted and 520 1p/19q intact patients. TCGA RNA-seq expression data was log2 transformed before using. Validation set: RNA microarray expression data of 222 glioma samples from Repository for Molecular Brain Neoplasia Data (Georgetown Database of Cancer G-DOC https://gdoc.georgetown.edu/gdoc/), containing 33 1p/19q co-deleted and 189 1p/19q intact patients. The characteristics of glioma patients were described in Table 1.

Running predictors

We evaluated the efficacy of five different methods for identifying 1p/19q status in two datasets. These methods were described as bellow:

Top scoring pairs (TSPairs) contains functions for selecting top scoring pairs whose relative rankings can be used to accurately classify individuals into one of two classes (20501*20500 =420500270 gene pairs) [45].

Table 1. Clinicopathological characteristics of the patients.

Variable		TCGA dataset	Rembrandt dataset
Age	≥45	333	69
	<45	296	103
	NA	63	50
Gender	Male	364	124
	Female	265	56
	NA	63	42
Preoperative KPS score	≥80	320	-
	<80	70	-
	NA	302	222
Grade	II	223	27
	III	245	29
	IV	161	131
	NA	63	35
IDH1/2 status	Mutation	440	-
	Wild type	242	-
	NA	10	222
1p/19q status	Codeleted	172	33
	Intact	520	189
Molecular subtype	Astrocytoma (II, III)	170	50
	Oligoastrocytoma (II, III)	118	6
	Oligodendroglioma (II, III)	180	35
	Glioblastoma	161	131
	NA	63	0

Gene Set Variation Analysis (GSVA) is a non-parametric, unsupervised method for estimating variation of gene set enrichment through the samples of an expression data set and bypasses the conventional approach of explicitly modeling phenotypes within the enrichment scoring algorithm [28].

Prediction analysis for microarrays (PAM), which can be defined as a ‘nearest shrunken centroid classifier’ is a statistical method for class prediction by adding a ‘fudge-factor’ to each statistic's denominator [46]. TCGA and Rembrandt datasets were removed batch effects before using pam method.

Functions Relating to the Smoothing of Numerical Data (smoother) could smooth numerical data, blur images and remove detail and noise. The gaussian window smoothing function allows users to infer the pattern of DNA aberration from gene expression [47].

Classification and Regression Training(caret), a machine learning algorithm, could integrate with the features in training and the modeled interaction features. The model was evaluated independently through stratified (k=10)-folds cross-validation [48].

Hierarchical clustering was performed using complete agglomeration algorithm and a distance metric equal correlation coefficient with data processed by PAM, smoother and GSVA methods. We obtained TSPairs method from R package “tspair”, GSVA method from R package “GSVA”, PAM method from R package “pamr”, smoother method from R package “smoother”, caret method from R package “caret”, Combat function for removing batch effects [49] among TCGA and Rembrandt datasets from R package “sva” and hierarchically clustered function from R package “pheatmap”. All packages were based on the statistical software environment R, version 3.3.4 (http://www.r-project.org).

AUTHOR CONTRIBUTIONS

Zhiliang Wang design and wrote the manuscript. Zheng Wang and Zheng Zhao gave suggestion on study design, discussed and interpreted the data. Tao Jiang and Chuanbao Zhang designed and supervised study. All authors read and approved the final manuscript.

ACKNOWLEDGMENTS

We appreciate the generosity of TCGA and Rembrandt network for sharing the huge amount of data.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

FUNDING

1. The National Key Research and Development Plan (No. 2016YFC0902500); 2. Beijing Science and Technology Plan (Z141100000214009); 3. Capital Medical Development Research Fund (2016-1-1072); 4. Beijing Municipal Administration of Hospitals Clinical Medicine Development of Special.

REFERENCES

1. Hu H, Mu Q, Bao Z, Chen Y, Liu Y, Chen J, Wang K, Wang Z, Nam Y, Jiang B, Sa JK, Cho HJ, Her NG, et al. Mutational Landscape of Secondary Glioblastoma Guides MET-Targeted Trial in Brain Tumor. Cell. 2018; 175:1665–1678.e18. https://doi.org/10.1016/j.cell.2018.09.038

2. Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, et al, and Cancer Genome Atlas Research Network. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010; 17:98–110. https://doi.org/10.1016/j.ccr.2009.12.020

3. Wen PY, Kesari S. Malignant gliomas in adults. N Engl J Med. 2008; 359:492–507. https://doi.org/10.1056/NEJMra0708126

4. Ostrom QT, Gittleman H, Farah P, Ondracek A, Chen Y, Wolinsky Y, Stroup NE, Kruchko C, Barnholtz-Sloan JS. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2006-2010. Neuro-oncol. 2013 (Suppl 2); 15:ii1–56. https://doi.org/10.1093/neuonc/not151

5. Ostrom QT, Gittleman H, Liao P, Vecchione-Koval T, Wolinsky Y, Kruchko C, Barnholtz-Sloan JS. CBTRUS Statistical Report: primary brain and other central nervous system tumors diagnosed in the United States in 2010-2014. Neuro-oncol. 2017 (suppl_5); 19:v1–88. https://doi.org/10.1093/neuonc/nox158

6. Felsberg J, Erkwoh A, Sabel MC, Kirsch L, Fimmers R, Blaschke B, Schlegel U, Schramm J, Wiestler OD, Reifenberger G. Oligodendroglial tumors: refinement of candidate regions on chromosome arm 1p and correlation of 1p/19q status with survival. Brain Pathol. 2004; 14:121–30. https://doi.org/10.1111/j.1750-3639.2004.tb00044.x

7. Pekmezci M, Rice T, Molinaro AM, Walsh KM, Decker PA, Hansen H, Sicotte H, Kollmeyer TM, McCoy LS, Sarkar G, Perry A, Giannini C, Tihan T, et al. Adult infiltrating gliomas with WHO 2016 integrated diagnosis: additional prognostic roles of ATRX and TERT. Acta Neuropathol. 2017; 133:1001–16. https://doi.org/10.1007/s00401-017-1690-1

8. Jenkins RB, Blair H, Ballman KV, Giannini C, Arusell RM, Law M, Flynn H, Passe S, Felten S, Brown PD, Shaw EG, Buckner JC. A t(1;19)(q10;p10) mediates the combined deletions of 1p and 19q and predicts a better prognosis of patients with oligodendroglioma. Cancer Res. 2006; 66:9852–61. https://doi.org/10.1158/0008-5472.CAN-06-1796

9. Brat DJ, Verhaak RG, Aldape KD, Yung WK, Salama SR, Cooper LA, Rheinbay E, Miller CR, Vitucci M, Morozova O, Robertson AG, Noushmehr H, Laird PW, et al, and Cancer Genome Atlas Research Network. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. N Engl J Med. 2015; 372:2481–98. https://doi.org/10.1056/NEJMoa1402121

10. Cairncross G, Wang M, Shaw E, Jenkins R, Brachman D, Buckner J, Fink K, Souhami L, Laperriere N, Curran W, Mehta M. Phase III trial of chemoradiotherapy for anaplastic oligodendroglioma: long-term results of RTOG 9402. J Clin Oncol. 2013; 31:337–43. https://doi.org/10.1200/JCO.2012.43.2674

11. van den Bent MJ, Brandes AA, Taphoorn MJ, Kros JM, Kouwenhoven MC, Delattre JY, Bernsen HJ, Frenay M, Tijssen CC, Grisold W, Sipos L, Enting RH, French PJ, et al. Adjuvant procarbazine, lomustine, and vincristine chemotherapy in newly diagnosed anaplastic oligodendroglioma: long-term follow-up of EORTC brain tumor group study 26951. J Clin Oncol. 2013; 31:344–50. https://doi.org/10.1200/JCO.2012.43.2229

12. Belaud-Rotureau MA, Meunier N, Eimer S, Vital A, Loiseau H, Merlio JP. Automatized assessment of 1p36-19q13 status in gliomas by interphase FISH assay on touch imprints of frozen tumours. Acta

Neuropathol. 2006; 111:255–63. https://doi.org/10.1007/s00401-005-0001-4

13. Smith JS, Alderete B, Minn Y, Borell TJ, Perry A, Mohapatra G, Hosek SM, Kimmel D, O'Fallon J, Yates A, Feuerstein BG, Burger PC, Scheithauer BW, Jenkins RB. Localization of common deletion regions on 1p and 19q in human gliomas and their association with histological subtype. Oncogene. 1999; 18:4144–52. https://doi.org/10.1038/sj.onc.1202759

14. Woehrer A, Sander P, Haberler C, Kern S, Maier H, Preusser M, Hartmann C, Kros JM, Hainfellner JA, and Research Committee of the European Confederation of Neuropathological Societies. FISH-based detection of 1p 19q codeletion in oligodendroglial tumors: procedures and protocols for neuropathological practice - a publication under the auspices of the Research Committee of the European Confederation of Neuropathological Societies (Euro-CNS). Clin Neuropathol. 2011; 30:47–55.

15. Chaturbedi A, Yu L, Linskey ME, Zhou YH. Detection of 1p19q deletion by real-time comparative quantitative PCR. Biomark Insights. 2012; 7:9–17. https://doi.org/10.4137/BMI.S9003

16. Windle B, Draper BW, Yin YX, O'Gorman S, Wahl GM. A central role for chromosome breakage in gene amplification, deletion formation, and amplicon integration. Genes Dev. 1991; 5:160–74. https://doi.org/10.1101/gad.5.2.160

17. Kukurba KR, Montgomery SB. RNA Sequencing and Analysis. Cold Spring Harb Protoc. 2015; 2015:951–69. https://doi.org/10.1101/pdb.top084970

18. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. Am J Hum Genet. 2013; 93:641–51. https://doi.org/10.1016/j.ajhg.2013.08.008

19. Goya R, Sun MG, Morin RD, Leung G, Ha G, Wiegand KC, Senz J, Crisan A, Marra MA, Hirst M, Huntsman D, Murphy KP, Aparicio S, Shah SP. SNVMix: predicting single nucleotide variants from next-generation sequencing of tumors. Bioinformatics. 2010; 26:730–36. https://doi.org/10.1093/bioinformatics/btq040

20. Climente-González H, Porta-Pardo E, Godzik A, Eyras E. The Functional Impact of Alternative Splicing in Cancer. Cell Reports. 2017; 20:2215–26. https://doi.org/10.1016/j.celrep.2017.08.012

21. Di Stefano AL, Fucci A, Frattini V, Labussiere M, Mokhtari K, Zoppoli P, Marie Y, Bruno A, Boisselier B, Giry M, Savatovsky J, Touat M, Belaid H, et al. Detection, Characterization, and Inhibition of FGFR-TACC Fusions in IDH Wild-type Glioma. Clin Cancer Res. 2015; 21:3307–17. https://doi.org/10.1158/1078-0432.CCR-14-2199

22. Song J, Singh M. How and when should interactome-derived clusters be used to predict functional modules and protein function? Bioinformatics. 2009; 25:3143–50. https://doi.org/10.1093/bioinformatics/btp551

23. Yan W, Zhang W, You G, Zhang J, Han L, Bao Z, Wang Y, Liu Y, Jiang C, Kang C, You Y, Jiang T. Molecular classification of gliomas based on whole genome gene expression: a systematic report of 225 samples from the Chinese Glioma Cooperative Group. Neuro-oncol. 2012; 14:1432–40. https://doi.org/10.1093/neuonc/nos263

24. Patnaik SK, Yendamuri S, Kannisto E, Kucharczuk JC, Singhal S, Vachani A. MicroRNA expression profiles of whole blood in lung adenocarcinoma. PLoS One. 2012; 7:e46045. https://doi.org/10.1371/journal.pone.0046045

25. Marchionni L, Afsari B, Geman D, Leek JT. A simple and reproducible breast cancer prognostic test. BMC Genomics. 2013; 14:336. https://doi.org/10.1186/1471-2164-14-336

26. Cao LL, Song X, Pei L, Liu L, Wang H, Jia M. Histone deacetylase HDAC1 expression correlates with the progression and prognosis of lung cancer: A meta-analysis. Medicine (Baltimore). 2017; 96:e7663. https://doi.org/10.1097/MD.0000000000007663

27. Jang SH, Kim AR, Park NH, Park JW, Han IS. DRG2 Regulates G2/M Progression via the Cyclin B1-Cdk1 Complex. Mol Cells. 2016; 39:699–704. https://doi.org/10.14348/molcells.2016.0149

28. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics. 2013; 14:7. https://doi.org/10.1186/1471-2105-14-7

29. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, Miller DK, Christ AN, Bruxner TJ, Quinn MC, Nourse C, Murtaugh LC, Harliwong I, et al, and Australian Pancreatic Cancer Genome Initiative. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016; 531:47–52. https://doi.org/10.1038/nature16965

30. Kalady MF, Dejulius K, Church JM, Lavery IC, Fazio VW, Ishwaran H. Gene signature is associated with early stage rectal cancer recurrence. J Am Coll Surg. 2010; 211:187–95. https://doi.org/10.1016/j.jamcollsurg.2010.03.035

31. Toledo JB, Korff A, Shaw LM, Trojanowski JQ, Zhang J. CSF α-synuclein improves diagnostic and prognostic performance of CSF tau and Aβ in Alzheimer's disease. Acta Neuropathol. 2013; 126:683–97. https://doi.org/10.1007/s00401-013-1148-z

32. Partial Least Squares and Principal Component

Regression. 2016.

33. Kuhn M, Contributions from Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, the R Core Team, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C, Hunt T. caret: Classification and Regression Training. 2017.

34. Cheng C, Zhou Y, Li H, Xiong T, Li S, Bi Y, Kong P, Wang F, Cui H, Li Y, Fang X, Yan T, Li Y, et al. Whole-Genome Sequencing Reveals Diverse Models of Structural Variations in Esophageal Squamous Cell Carcinoma. Am J Hum Genet. 2016; 98:256–74. https://doi.org/10.1016/j.ajhg.2015.12.013

35. Flister MJ, Tsaih SW, Stoddard A, Plasterer C, Jagtap J, Parchur AK, Sharma G, Prisco AR, Lemke A, Murphy D, Al-Gizawiy M, Straza M, Ran S, et al. Host genetic modifiers of nonproductive angiogenesis inhibit breast cancer. Breast Cancer Res Treat. 2017; 165:53–64. https://doi.org/10.1007/s10549-017-4311-8

36. Eckel-Passow JE, Lachance DH, Molinaro AM, Walsh KM, Decker PA, Sicotte H, Pekmezci M, Rice T, Kosel ML, Smirnov IV, Sarkar G, Caron AA, Kollmeyer TM, et al. Glioma Groups Based on 1p/19q, IDH, and TERT Promoter Mutations in Tumors. N Engl J Med. 2015; 372:2499–508. https://doi.org/10.1056/NEJMoa1407279

37. Mahdieh N, Rabbani B. An overview of mutation detection methods in genetic disorders. Iran J Pediatr. 2013; 23:375–88.

38. McPherson E. Genetic diagnosis and testing in clinical practice. Clin Med Res. 2006; 4:123–29. https://doi.org/10.3121/cmr.4.2.123

39. Han Y, Gao S, Muegge K, Zhang W, Zhou B. Advanced Applications of RNA Sequencing and Challenges. Bioinform Biol Insights. 2015 (Suppl 1); 9:29–46. 10.4137/BBI.S28991

40. Hellmann MD, Nathanson T, Rizvi H, Creelan BC, Sanchez-Vega F, Ahuja A, Ni A, Novik JB, Mangarin LM, Abu-Akeel M, Liu C, Sauter JL, Rekhtman N, et al. Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer. Cancer Cell. 2018; 33:843–852.e4. https://doi.org/10.1016/j.ccell.2018.03.018

41. Tang YC, Amon A. Gene copy-number alterations: a cost-benefit analysis. Cell. 2013; 152:394–405. https://doi.org/10.1016/j.cell.2012.11.043

42. Zhao N, Stoffel A, Wang PW, Eisenbart JD, Espinosa R 3rd, Larson RA, Le Beau MM. Molecular delineation of the smallest commonly deleted region of chromosome 5 in malignant myeloid diseases to 1-1.5 Mb and preparation of a PAC-based physical map. Proc Natl Acad Sci USA. 1997; 94:6948–53. https://doi.org/10.1073/pnas.94.13.6948

43. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, Ladd C, Beheshti J, Bueno R, Gillette M, Loda M, Weber G, Mark EJ, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci USA. 2001; 98:13790–95. https://doi.org/10.1073/pnas.191502998

44. Hu X, Martinez-Ledesma E, Zheng S, Kim H, Barthel F, Jiang T, Hess KR, Verhaak RG. Multigene signature for predicting prognosis of patients with 1p19q co-deletion diffuse glioma. Neuro-oncol. 2017; 19:786–95. https://doi.org/10.1093/neuonc/now285

45. Leek JT. The tspair package for finding top scoring pair classifiers in R. Bioinformatics. 2009; 25:1203–04. https://doi.org/10.1093/bioinformatics/btp126

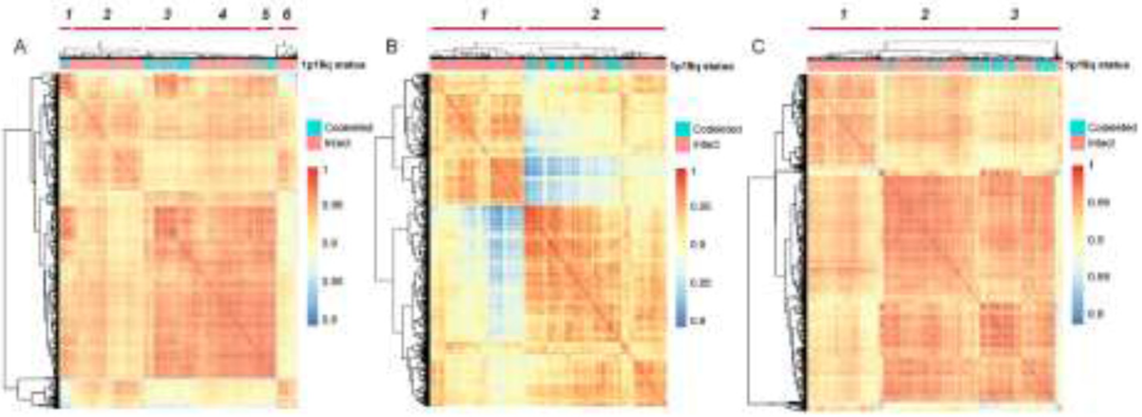
46. Hastie T, Tibshirani R, Narasimhan B, Chu G. Pam: prediction analysis for microarrays. CRAN 2014-08-27.

47. Hamilton N. Functions Relating to the Smoothing of Numerical Data. 2014.

48. Kuhn M. Contributions from Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Cooper T, Mayer Z, Kenkel B, the R Core Team, Benesty M, Lescarbeau R, Ziem A, Scrucca L, Tang Y, Candan C, Hunt T. Classification and Regression Training. 2017.

49. Leek JT, Johnson WE, Parker HS, Fertig EJ, Jaffe AE, Storey JD, Zhang Y, Torres LC. Surrogate Variable Analysis. R package version 3.24.4. 2017.

SUPPLEMENTARY MATERIAL



Supplementary Figure 1. Expression profile revealed a distinct gene expression between Chromosome 1p19q codeleted and intact patients. (**A**) Hierarchical clustering based on whole genome expression profiling (20501 genes), there were 92.4% (159/172) 1p/19q co-deleted patients in group1, 3 and 5 and 85.7% (446/520) 1p/19q intact patients in the rest of groups. (**B**) Hierarchical clustering based on highly variable expression genes (MAD >2, 886 genes), there were 96.5% (166/172) 1p/19q co-deleted patients in group2. (**C**) Hierarchical clustering based on genes on chromosome 1p and 19q (1775 genes), there were 75% (129/172) 1p/19q co-deleted patients in group3

Supplementary Table 1. Integrated analysis of methods for detecting 1p/19q status.

Methods	Cost	Time	Accuracy
Whole genome sequencing	~$2500	~1month	100%
RNA sequencing	~$200	~2weeks	97.8%
FISH	~$340	~2weeks	1p36/1q21 and 19q13/19p13 Deletion Probe Kit