

## Research Article

# Joint Covariate Detection on Expression Profiles for Selecting Prognostic miRNAs in Glioblastoma

**Chengqi Sun and Xudong Zhao**

*College of Information and Computer Engineering, Northeast Forestry University, Harbin 150001, China*

Correspondence should be addressed to Xudong Zhao; [zhaoxudong@nefu.edu.cn](mailto:zhaoxudong@nefu.edu.cn)

Received 4 October 2016; Revised 18 January 2017; Accepted 27 February 2017; Published 20 March 2017

Academic Editor: Xia Li

Copyright © 2017 Chengqi Sun and Xudong Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

An important application of expression profiles is to stratify patients into high-risk and low-risk groups using limited but key covariates associated with survival outcomes. Prior to that, variables considered to be associated with survival outcomes are selected. A combination of single variables, each of which is significantly related to survival outcomes, is always regarded to be candidates for posterior patient stratification. Instead of individually significant variables, a combination that contains not only significant but also insignificant variables is supposed to be concentrated on. By means of bottom-up enumeration on each pair of variables, we propose a joint covariate detection strategy to select candidates that not only correspond to close association with survival outcomes but also help to make a clear stratification of patients. Experimental results on a publicly available dataset of glioblastoma multiforme indicate that the selected pair composed of an individually significant and an insignificant miRNA keeps a better performance than the combination of significant single variables. The selected miRNA pair is ultimately regarded to be associated with the prognosis of glioblastoma multiforme by further pathway analysis.

## 1. Introduction

Survival analysis, which is a branch of statistics for analyzing time-to-event data, is commonly used in cancer research. In particular, it helps to assess the prognosis of patients having specific types of cancer in informing not only the categories of patients with differing survival outcomes but also the possible molecular cause of the risk of death. Narrow down to gliomas, expression profiles are utilized to discover the subtypes of patients with different survival risks [1]. This kind of data provides a supplementary predictor of survival due to the limited effectiveness of current clinical diagnoses. Numerous studies which attempted to use selected signatures from expression profiles for discrete stratification (e.g., recurrence, metastasis, and chemotherapy efficacy) have shown the effectiveness. Correspondingly, several methods that classified patients into subgroups with differing survival time have also been performed.

Considering the continuity of the observations' survival time with right censoring, Cox proportional hazards regression analysis [2] was extensively utilized to seek covariates

associated with the overall survival of patients in invasive breast cancer [3], non-small-cell lung cancer [4], follicular lymphoma [5], glioblastoma [6–8], and so forth. Due to the requirement of more observations than covariates, Cox proportional hazards regression model was combined with some methods for dimension reduction or shrinkage such as partial least squares [9] and principle component analysis [10]. However, these strategies can only provide a combination of variables other than reporting meaningful covariates. Since projections derived from these variables are made, one can only tell these variables together but not which variables are effective. Besides, top-down methods of tree-structured survival analysis [11] and random survival forests [12] associated with hazards regression were proposed for selection of covariates. Unlike bottom-up enumeration strategies, these heuristic approaches may get local optimal solutions although they infinitely approximate to global optimal solutions despite their efficiencies.

Hence, univariable regression analyses have been placed firmly in the mainstream. Due to the high-dimensional space of variables compared to the small observation size,

a penalized Cox hazards model using least-angle regression was proposed in order to solve the overfitting problem of parameter learning [13]. In addition, a sparse kernel method was proposed on condition that the correlation between the logarithm of the hazard ratio and covariates was linear, and a survival supporting vector machine that maximized the classification margin other than Cox regression was presented [14]. In practice, univariable Cox regression analysis was applied to each variable, which was regarded to be significant considering its correlation with survival time or its distinct stratification of patients. Significant variables were selected using either Wald  $t$ -test on regression coefficients [15] or log-rank test with permutations after dividing patients into high- and low-risk groups by univariable risk-score analysis [16, 17]. A risk score of each observation was obtained using a linear combination of the expression levels of selected variables weighted by multivariable regression coefficients. A cut-off threshold was derived from the median risk score or was determined by receiver operating characteristics (ROC) analysis [18], and patients within the training set were divided into high- and low-risk groups. The major problem of univariable Cox regression strategy roots in the assumption that covariates are derived from individual variables, each of which is significantly associated with survival outcomes. In essence, a meaningful set of covariates are probably composed of different variables, each of which is either correlated with or apparently unrelated to survival outcomes.

In order to solve this problem, we propose a joint covariate detection strategy for selection of variable pairs instead of a combination of variables individually correlated with survival time from expression profiles. Meanwhile, stratification of patients is also considered. That is, predictors not only associated with survival outcomes but also helpful to classify patients into high-risk and low-risk groups are chosen. Cox proportional hazards regression is used in order to detect variable pairs that are most associated with survival time. In order to overcome the overfitting problem, variable pairs which may most possibly help to stratify patients with differing survival risks are further selected. In particular, patients are stratified according to the corresponding risk-score analysis derived from Cox proportional hazards regression. Besides, log-rank test is performed for further confirmation whether the selected variable pairs contribute mainly to the stratification or not. In order to show the effectiveness of our method, miRNA expression profiles containing 548 patients with glioblastoma multiforme (GBM) downloaded from the Cancer Genome Atlas (TCGA) database are introduced in. The final selected miRNA pair of significance as representing the covariates not only most associated with survival outcomes but also effective to stratification of patients is ultimately testified using KEGG pathway analysis.

## 2. Materials and Methods

**2.1. Microarray Data.** We use the miRNA expression data (Level 3) of 548 patients with GBM downloaded from TCGA (<http://cancergenome.nih.gov>) in order to illustrate the effectiveness of identifying prognostic miRNAs in glioblastoma using the joint covariate selection method. In total, these

548 GBM cases with overall survival information are selected from 581 miRNA expression profiles, which were downloaded during May, 2015. We choose all the patients, for we discover that splitting samples using a random dichotomy or by balancing survival outcomes between training and testing group cannot achieve the same set of variables as it is derived from the whole samples. That is to say, how to reasonably split samples into training and testing ones is still under discussion. The reason derives from two aspects. One is that survival outcomes are continuous compared to discrete stratification (e.g., recurrence, metastasis, and chemotherapy efficacy). Thus, the distribution of survival outcomes is to be estimated before splitting samples. The other is that it is hard to estimate the distribution of survival outcomes because of including censored following time. Moreover, the survival time of each patient is recorded, which ranges between 0 and 3881 days. Among them, 450 are dead (uncensored) during the study and 98 are still alive (censored) at the end of the study. *MatlabR2013b* is selected as the experimental platform. Coefficients of Cox regression are obtained by calling the library function *coxphfit*.

**2.2. Joint Covariate Detection.** Here, it represents a twofold consideration on detection of variables, which are both associated with survival outcomes and helpful to classify patients into different risk groups. In order to seek variables associated with survival outcomes, Cox hazards regression is firstly introduced. The partial likelihood function is given by the expression

$$l(\boldsymbol{\beta}) = \prod_{i=1}^m \frac{e^{\mathbf{x}_{(i)}^T \boldsymbol{\beta}}}{\sum_{j \in R(t_{(i)})} e^{\mathbf{x}_{(j)}^T \boldsymbol{\beta}}}, \quad (1)$$

where the product is over the  $m$  distinct ordered survival time without any follow-up of right censoring assuming that there is no tied time.  $\mathbf{x}_{(i)}$  and  $\boldsymbol{\beta}$  denote the  $i$ th expression levels and the regression coefficients of the detected variables, respectively. The summation in the denominator is over all subjects in the risk set at ordered survival time  $t_{(i)}$ , denoted by  $R(t_{(i)})$ . The maximum partial likelihood estimator is obtained by differentiating the right hand side of the logarithm transformation of (1) with respect to  $\boldsymbol{\beta}$ , setting the derivative equal to zero, and solving for  $\boldsymbol{\beta}$ . As to each component of  $\boldsymbol{\beta}$ , a Wald statistic that represents the ratio of the estimated coefficient to its estimated standard error is presented. That is,

$$z_k = \frac{\hat{\beta}_k}{\text{SE}(\hat{\beta}_k)}. \quad (2)$$

The  $p$  value of the  $k$ th component of  $\boldsymbol{\beta}$  is obtained by looking up a table assuming that the Wald statistic in (2) follows the standard normal distribution. In order to enlarge the sample size, we make a permutation test by reordering the survival outcomes for  $B$  times. And the corresponding  $p$  value is expressed as follows:

$$p_k = \sum_{b=1}^B \frac{\#(|z_k^0| \geq |z_k|)}{B}, \quad (3)$$

where  $z_k^0$  denotes a null statistics by a random rearrangement of survival outcomes. Enumeration on each single variable or on each pair is made. Therefore, covariates significantly associated with survival outcomes are selected according to the individuals or the pairs with smallest  $p$  values.

Meanwhile, we consider a best stratification of patients with differing survival outcomes as an indicator for selection of covariates. In practice, patients are commonly classified into low-risk and high-risk groups, which conforms to the daily doctors' decision making process. Following the case, the risk score is the linear portion of Cox regression model, of which the estimator for the  $i$ th sample containing  $p$  covariates is

$$\hat{r}_i = \hat{r}(\mathbf{x}_i, \hat{\boldsymbol{\beta}}) = \sum_{k=1}^p \hat{\beta}_k x_{ik}. \quad (4)$$

Median risk score is utilized as a cut-off value for stratification, in order to keep the equivalent number between high-risk and low-risk patients. Assuming that the survival function is the same in each of the two groups, the estimator of the expected number of deaths in high-risk group is expressed as

$$\hat{e}_{1i} = \frac{n_{1i}d_i}{n_i}, \quad (5)$$

where  $n_i$  and  $d_i$  represent the number at risk and of deaths at the observation of ordered survival time  $t_{(i)}$ , respectively.  $n_{1i}$  denotes the number at risk in high-risk group. Correspondingly, the estimator of the variance of  $d_{1i}$  on the hypergeometric distribution is defined as follows:

$$\hat{v}_{1i} = \frac{n_{1i}n_{0i}d_i(n_i - d_i)}{n_i^2(n_i - 1)}, \quad (6)$$

where  $n_{0i}$  denotes the number at risk in low-risk group. Under the null hypothesis that survival functions of the two groups are the same, the statistic of log-rank test is expressed as follows:

$$Q = \frac{[\sum_{i=1}^m (d_{1i} - \hat{e}_{1i})]^2}{\sum_{i=1}^m \hat{v}_{1i}}. \quad (7)$$

The corresponding  $p$  value is obtained using the  $\chi^2$  distribution with one degree of freedom. In the same way, we make a permutation test similarly expressed in (3). That is,

$$p_r = \sum_{b=1}^B \frac{\#(|Q_r^0| \geq |Q_r|)}{B}, \quad (8)$$

where  $Q_r^0$  also represents a null statistics by a random rearrangement of survival outcomes. After enumerating on each individual variable or on each pair, covariates that significantly categorize patients with differing survival outcomes are detected according to smallest  $p$  values.

By enumeration on each variable and each pair, significant covariates most associated with survival time are chosen on condition that each component keeps a small  $p$  value as

expressed in (3). Moreover, the variables for stratification of patients using the risk score defined by (4) correspond to small  $p$  values as seen in (8). In fact, this conception derives from Integrative Hypothesis Testing (IHT) proposed by Xu [19]. The obtained covariates may indicate not only a close correlation with survival time but also distinct stratification of patients.

**2.3. KEGG Pathway Analysis.** In order to show the effectiveness of our method, we submit the final selected miRNA pair, which is not only most associated with survival outcomes but also effective to stratification of patients to low-risk and high-risk groups, to DIANA miRPath [20]. We only use TarBase [21] to select the targets of the miRNA pair, considering that it is a database of published experiments validated miRNA-gene interactions. Focusing on the pathways related to the selected miRNA pair instead of those corresponding to each component of the selected pair, we can find significant pathways, which may support our finding and show the effectiveness of our method.

### 3. Results

**3.1. Joint Covariate Detection for GBM Survival Analysis.** In this part, we apply joint covariate detection to seeking miRNAs which are associated with the risk of death and the stratification of high-risk and low-risk patients in GBM. The representation of "joint" is twofold. First, it is a strategy that combines Cox regression for seeking survival-associated variables with log-rank test on risk scores for evaluation of the classification results. Second, it also exhibits the steps from enumerations on each individual variable to those on enumerable covariate tuples. Considering the computational cost, joint covariate detection terminates after finishing enumeration on miRNA pairs.

For each miRNA,  $p$  values expressed in (3) and (8) were obtained after 10000 rounds of permutations. The miRNAs with  $p$  values  $\leq 0.01$  were regarded to be individually significant. We obtained six significant miRNAs, as listed in Table 1. Using the expression levels of the selected significant miRNAs, we made Kaplan-Meier survival analyses on high-risk and low-risk groups derived from cut-off values by calculating the median risk scores expressed in (4), as illustrated in Figure 1. Besides,  $p$  values of each significant miRNA were also shown in Figure 1. On assumption that hazard ratio (HR) is constant over survival time, we listed HRs in Figure 1, too.

As to each miRNA pair, permutations with 10000 rounds were made.  $p$  values corresponding to each component of every pair were calculated by (3). After a linear combination of the expression levels which regarded the learned Cox regression coefficients as its weights, risk scores were obtained using (4). The median risk score was utilized as a cut-off value; therefore, patients were classified into the high-risk and low-risk groups. Log-rank test expressed in (7) was performed, and the corresponding  $p$  value representing the significant differences of risks between the two groups was calculated by (8). The miRNA pairs with  $p$  values  $\leq 0.001$  were regarded as the significant pairs associated with the risk of death to patients in GBM. We obtained six significant pairs of

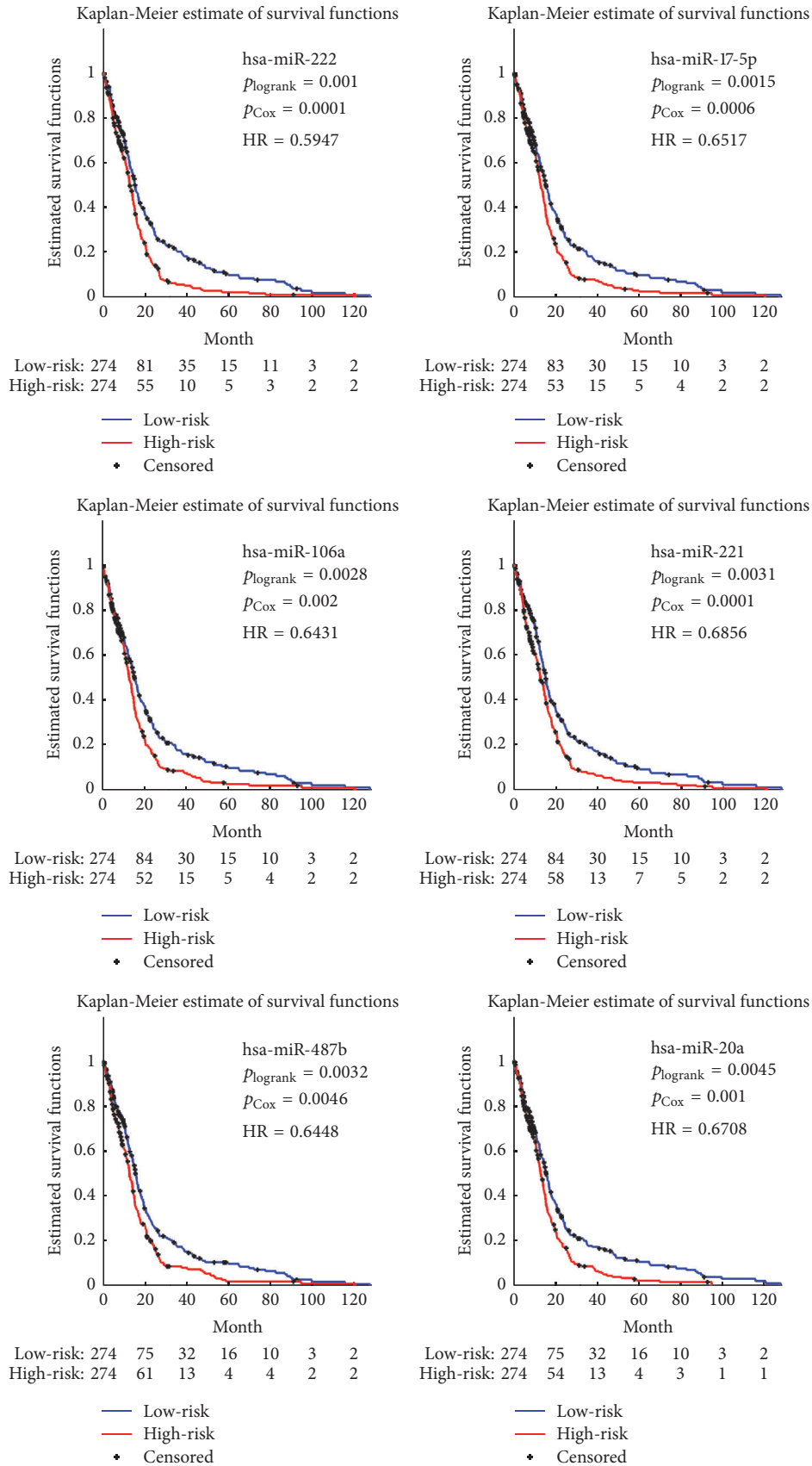


FIGURE 1: Kaplan-Meier survival analysis using significant individual miRNAs, each of which is jointly selected to Cox regression and log-rank test with  $p$  values  $\leq 0.01$ .

TABLE 1: Individual results using joint covariate detection.

miRNA probe	Z (log-rank)	p (log-rank)	Coef (Cox)	Z (Cox)	p (Cox)	Expressions in high-risk group
hsa-miR-222	3.042262	0.0012	0.245557	6.33205	0.0001	High
hsa-miR-17-5p	2.878053	0.0015	-0.22319	-3.36403	0.0006	Low
hsa-miR-106a	2.841924	0.0028	-0.18956	-3.00119	0.002	Low
hsa-miR-221	2.792194	0.0031	0.283759	5.395549	0.0001	High
hsa-miR-487b	2.711448	0.0032	0.207506	2.849673	0.0046	High
hsa-miR-20a	2.688864	0.0045	-0.1768	-3.16297	0.001	Low

TABLE 2: Pair results using joint covariate detection.

miRNA probe	miRNA probe	p (log-rank)	p (Cox)	p (Cox)	Expressions in high-risk group	Expressions in high-risk group
hsa-miR-10b	hsa-miR-222	0.0002	0.0004	0.0001	High	High
hsa-miR-196a	hsa-miR-20a	0.0003	0.0007	0.0002	High	Low
hsa-miR-222	hsa-miR-422b	0.0003	0.0001	0.0007	High	Low
hsa-miR-140	hsa-miR-148a	0.0007	0.0004	0.0001	Low	High
hsa-miR-196a	hsa-miR-340	0.0007	0.001	0.0003	High	Low
hsa-miR-340	hsa-miR-765	0.0009	0.0001	0.0006	Low	Low

miRNAs (see Table 2), of which survival analyses were shown in Figure 2.

Figure 2 illustrated the experimental results of the significant pairs. Kaplan-Meier survival analysis was made between the high-risk and low-risk groups of patients on each significant pair. *p* values of each component and that corresponding to log-rank test were also listed in Table 2. By making a careful comparison between the Kaplan-Meier curves associated with miRNA pairs shown in Figure 3 and those related to individually significant miRNAs illustrated in Figure 1, we discovered that the selected miRNA pairs contributed to an easier stratification of patients with survival months less than 10 months, as illustrated in Figure 3.

Next, we tried to validate that covariates most associated with survival outcomes were not equal to the set of individually significant variables. In order to demonstrate it, we enumerated all possible combinations of the six miRNAs that were individually significant as listed in Table 1 and illustrated in Figure 1 and performed joint covariate detection on each combination. We made 10000 rounds of permutations and set the threshold to be 0.05 for significant detection. Of all the 57 combinations except six individually significant miRNAs, two significant combinations were obtained, as illustrated in Figure 4. *p* values of each miRNA and that corresponding to log-rank test were listed for each combination. After carefully comparing parameters in Figure 2 with those in Figure 4, we made a conclusion that covariates selected for discrimination of GBM prognosis could not only consist of individually significant variables. In other words, significant covariates possibly consisted of different variables, each of which is either individually significant or not.

3.2. *Verification of miRNAs Associated with GBM Prognosis.* According to the small *p* values and small HRs illustrated in Figure 2, we selected the most significant miRNA pair (i.e., miR-10b and miR-222). In order to validate the chosen

pair’s close association with prognosis of GBM instead of significant or insignificant individuals, we used DIANA miRPath [20] and TarBase [21] which provide miRNA/gene interactions with high quality experimental validations to identify KEGG pathways related to both miR-222 and miR-10b. Pathways corresponding to miR-222, miR-10b and their combination are listed in Tables 3, 4, and 5, respectively.

Comparisons from Tables 3–5 show that pathways including glioma and melanoma may have direct relations with both miR-222 and miR-10b, which might indirectly support our option about the need of joint covariate detection. The glioma pathway is illustrated in Figure 5.

#### 4. Discussion

In this paper, a joint covariate detection strategy is proposed for selecting candidates that not only correspond to close association with survival outcomes but also help to make a clear stratification of patients. We choose GBM data and testify the effectiveness of our method on it for three reasons presented as follows. First, GBM data has a large sample size containing 548 patients. Such a large sample size ensures the reliability of statistical results, and that is also the reason why we keep the whole samples for training the model. Second, GBM data has a very long follow-up time, the longest of which has reached a length of over ten years. Third, right censored observations are keeping in a smaller sample size, which now has 98 cases compared to 450 having passed away. Less censored samples make more robust fitting result of Cox proportional hazards regression.

Joint covariate detection contains the concept that makes a combination between selecting covariates most associated with survival outcomes and seeking covariates which is capable of risk stratification. To the best of our knowledge, it is the first model using bottom-up enumerations of variable pairs other than combination of individually significant variables,

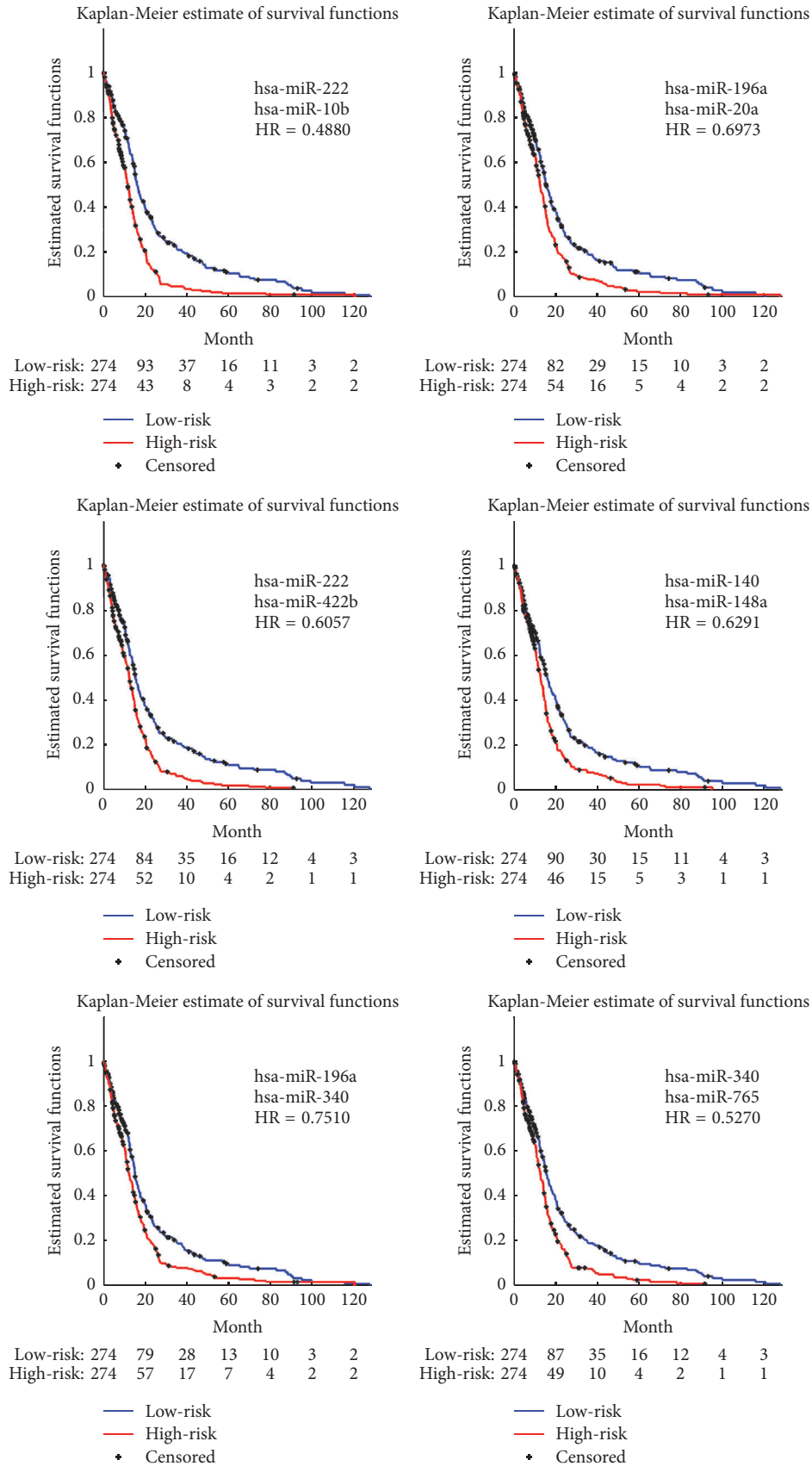


FIGURE 2: Kaplan-Meier survival analysis using significant miRNA pairs, each of which is jointly selected to Cox regression and log-rank test with  $p$  values  $\leq 0.001$ .

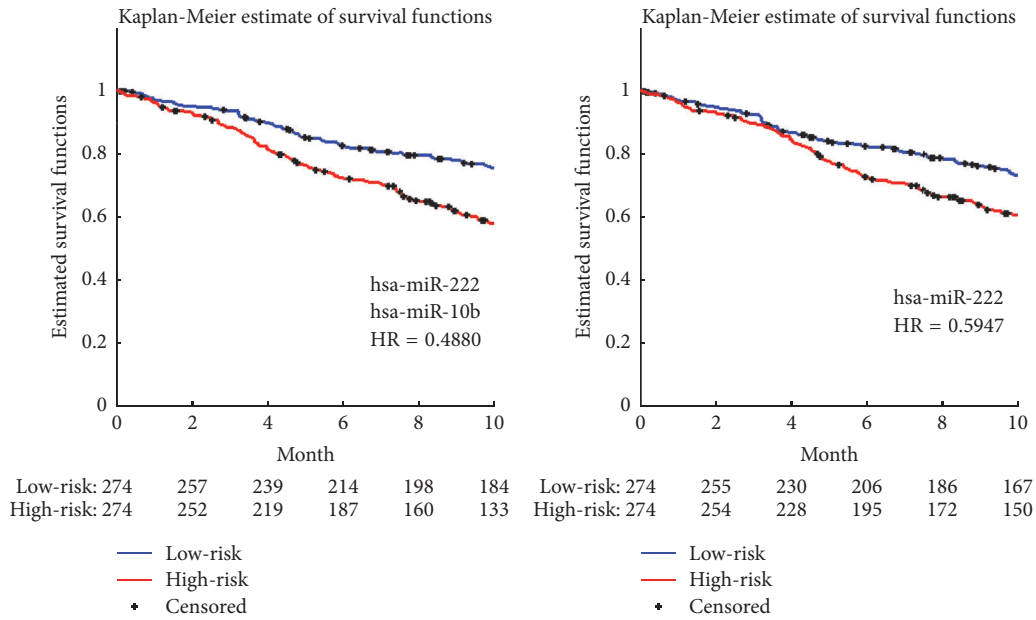


FIGURE 3: Comparisons between the most significant pair and the most significant individual miRNA.

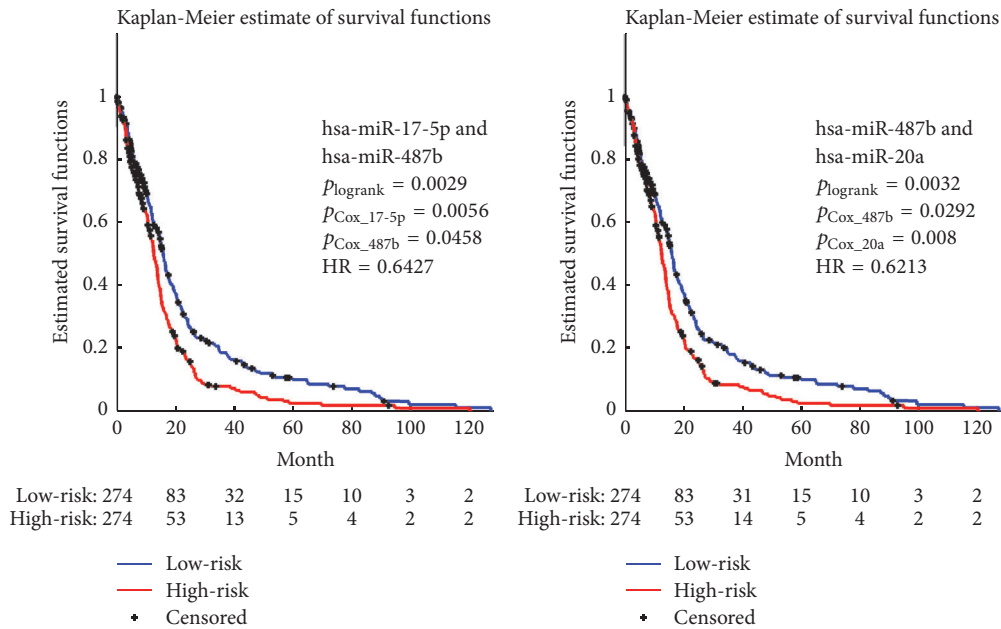


FIGURE 4: Representative survival analysis using combinations of the six significant individual miRNAs, each of which is jointly selected according to Cox regression and log-rank test with  $p$  values  $\leq 0.05$ .

which has been widely provided in practice. Considering that the expression profiles commonly have large dimension and small sample size, permutation tests are made by reordering survival outcomes to enlarge sample size. Besides, log-rank tests may also help to solve the overfitting problem. Model development such as covariate interaction [22] can be further introduced in.

However, several limitations of the proposed strategy have to be listed as follows. First, joint covariate detection terminates after finishing the pair enumeration due to the

high computational cost. In order to proof our inference, enumerations on multitudes of variables need to be done. Second, strategies which contain penalties or constraints are excluded considering a fast performing demand. Third, we keep Cox proportional hazards assumption that the hazard ratio is independent of survival time. In fact, it is the covariate but not the regression coefficient that does not depend on survival time. Covariates whose values are fixed at the beginning of observation remain unchanged throughout the follow-up time. On condition that the difference in log

TABLE 3: Pathways associated with miR-222 ( $p \leq 0.01$ ).

KEGG pathway	$p$ value	# genes	# miRNAs
Fatty acid biosynthesis	$1.64E - 25$	1	1
Fatty acid metabolism	$5.15E - 05$	2	1
Arrhythmogenic right ventricular cardiomyopathy (ARVC)	$5.15E - 05$	8	1
Viral carcinogenesis	0.000376	22	1
Protein processing in endoplasmic reticulum	0.006288	20	1
Lysine degradation	0.011993	5	1
RNA degradation	0.011993	11	1
p53 signaling pathway	0.011993	11	1
Ubiquitin mediated proteolysis	0.013637	17	1
RNA transport	0.02166	20	1
Cell cycle	0.02166	15	1
Spliceosome	0.024132	11	1
Endometrial cancer	0.038151	6	1
Adherens junction	0.041217	9	1
HTLV-I infection	0.046109	24	1
Central carbon metabolism in cancer	0.047174	6	1
Bacterial invasion of epithelial cells	0.049002	9	1

TABLE 4: Pathways associated with miR-10b ( $p \leq 0.01$ ).

KEGG pathway	$p$ value	# genes	# miRNAs
Fatty acid biosynthesis	$4.92E - 28$	1	1
Viral carcinogenesis	$1.11E - 06$	16	1
Fatty acid metabolism	$2.78E - 06$	1	1
Chronic myeloid leukemia	0.0005	9	1
Central carbon metabolism in cancer	0.002282	7	1
Non-small-cell lung cancer	0.007224	8	1
Glycosphingolipid biosynthesis, lacto- and neolactoseries	0.015745	2	1
Pyrimidine metabolism	0.021438	7	1
Cell cycle	0.022075	13	1
p53 signaling pathway	0.028173	9	1

TABLE 5: Pathways associated with both miR-10b and miR-222 ( $p \leq 0.01$ ).

KEGG pathway	$p$ value	# genes	# miRNAs
Fatty acid biosynthesis	$5.98E - 47$	1	2
Fatty acid metabolism	$6.60E - 22$	1	2
Viral carcinogenesis	$1.54E - 05$	5	2
Chronic myeloid leukemia	0.001129	4	2
<i>Glioma</i>	0.015536	4	2
Non-small-cell lung cancer	0.023173	4	2
<i>Melanoma</i>	0.030385	4	2
Cell cycle	0.043136	6	2

hazards depended on time, a nonparametric concordance measure [23] or an alternative selection of concordance regression and weighted Cox regression [24] were presented, respectively, instead of Cox regression. Once the problem

of computational cost is solved, these improvements can be added in. Fourth and most important, how to reasonably split samples with survival outcomes into training and testing ones is still a vital problem under discussion. Actually, splitting samples using a random dichotomy or by balancing survival outcomes between training and testing group will not work, especially on low-dimension feature space. And that has been experimentally demonstrated (not shown). All these limitations are to be settled in the future.

Using joint covariate detection, we chose one miRNA pair associated with GBM prognosis. In order to reveal the relationship between the chosen miRNA pair and the survival time of GBM patients, DIANA miRPath [20] and TarBase [21], which provided miRNA/gene interactions with high quality experimental validations, were utilized. As listed in Table 5, two pathways including glioma and melanoma were manifested, which indicated a joint action of the miRNA pair. With experimental validations, four common targets of each



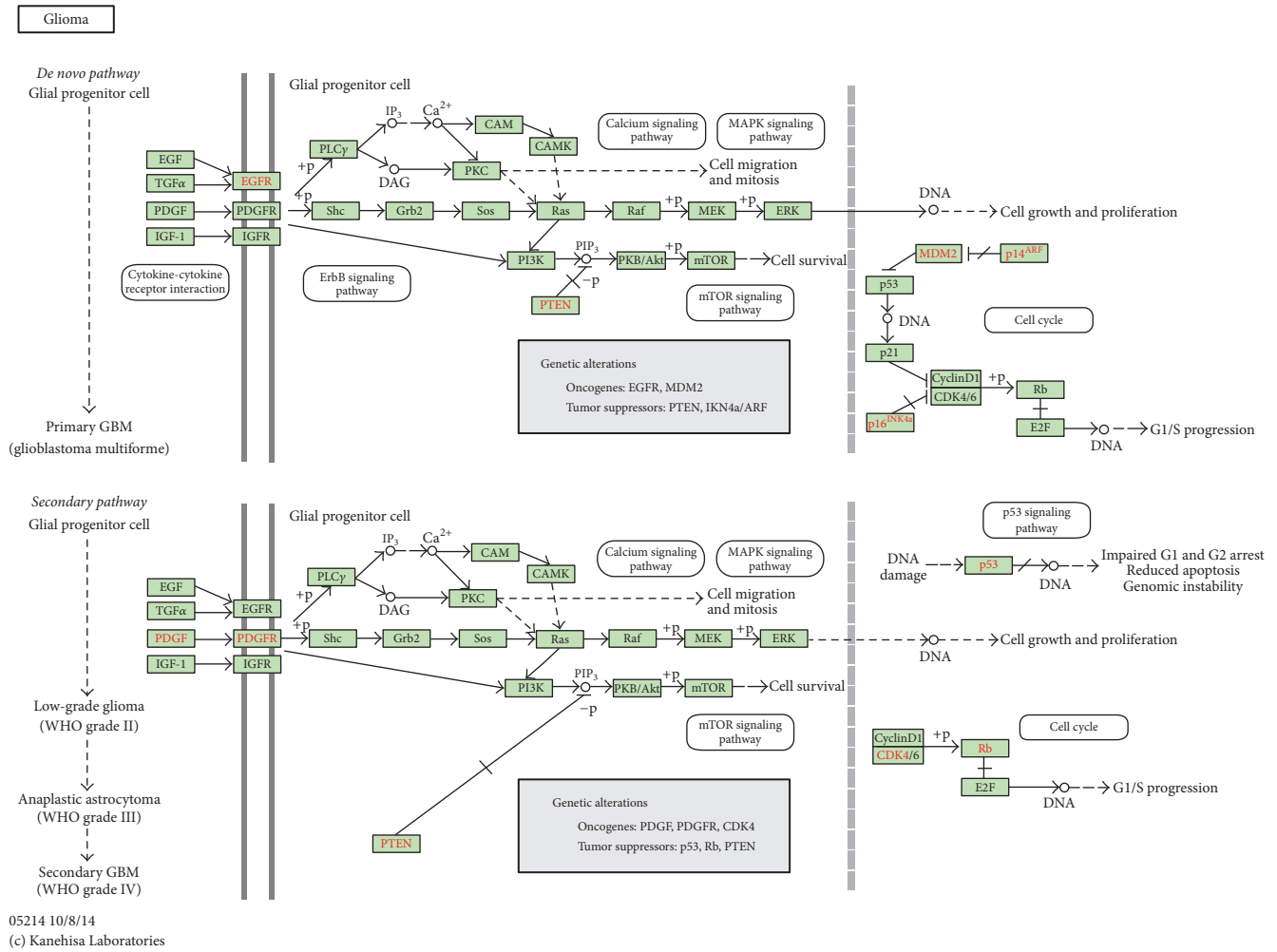


FIGURE 5: Glioma pathway that shows close association with miR-222 and miR-10b.

component in the miRNA pair (i.e., MDM2, TP53, CDK6, and E2F3) were focused on. In Figure 5, the four targets were included in cell cycle pathway, which was also the case in melanoma pathway. MDM2 and TP53 were reported to be directly associated with prognosis of GBM [25]. As illustrated in Figure 5, INK4a and ARF acted as tumor suppressors, which were upstream genes of the common targets. The loss of INK4a and ARF together with p53 gene mutation was reported to be mutually exclusive events in GBM [26].

**Conflicts of Interest**

The authors declare that they have no conflicts of interest.

**Acknowledgments**

The authors would like to thank the financial support of Specialized Personnel Start-Up Grant (no. 41112419) and Special Program for Applied Research on Super Computation of the NSFC-Guangdong Joint Fund (the second phase).

**References**

- [1] L. A. M. Gravendeel, M. C. M. Kouwenhoven, O. Gevaert et al., "Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology," *Cancer Research*, vol. 69, no. 23, pp. 9065–9072, 2009.
- [2] D. R. Cox, "Regression models and life tables (with discussion)," *Journal of Royal Statistical Society: Series B*, vol. 34, no. 2, pp. 187–220, 1972.
- [3] S. Volinia and C. M. Croce, "Prognostic microRNA/mRNA signature from the integrated analysis of patients with invasive breast cancer," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 18, pp. 7413–7417, 2013.
- [4] H.-Y. Chen, S.-L. Yu, C.-H. Chen et al., "A five-gene signature and clinical outcome in non-small-cell lung cancer," *The New England Journal of Medicine*, vol. 356, no. 1, pp. 11–20, 2007.
- [5] S. S. Dave, G. Wright, B. Tan et al., "Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells," *New England Journal of Medicine*, vol. 351, no. 21, pp. 2159–2169, 2004.

- [6] R. Li, K. Gao, H. Luo et al., "Identification of intrinsic subtype-specific prognostic microRNAs in primary glioblastoma," *Journal of Experimental and Clinical Cancer Research*, vol. 33, no. 1, article 9, 2014.
- [7] Z.-S. Bao, C.-B. Zhang, H.-J. Wang et al., "Whole-genome mRNA expression profiling identifies functional and prognostic signatures in patients with mesenchymal glioblastoma multiforme," *CNS Neuroscience and Therapeutics*, vol. 19, no. 9, pp. 714–720, 2013.
- [8] X.-Q. Zhang, S. Sun, K.-F. Lam et al., "A long non-coding RNA signature in glioblastoma multiforme predicts survival," *Neurobiology of Disease*, vol. 58, pp. 123–131, 2013.
- [9] H. Li and J. Gui, "Partial Cox regression analysis for high-dimensional microarray gene expression data," *Bioinformatics*, vol. 20, no. 1, pp. i208–i215, 2004.
- [10] L. Li and H. Li, "Dimension reduction methods for microarrays with applications to censored survival data," *Bioinformatics*, vol. 20, no. 18, pp. 3406–3412, 2004.
- [11] M. L. Wallace, "Time-dependent tree-structured survival analysis with unbiased variable selection through permutation tests," *Statistics in Medicine*, vol. 33, no. 27, pp. 4790–4804, 2014.
- [12] A. Kawaguchi, N. Yajima, N. Tsuchiya et al., "Gene expression signature-based prognostic risk score in patients with glioblastoma," *Cancer Science*, vol. 104, no. 9, pp. 1205–1210, 2013.
- [13] J. Gui and H. Li, "Penalized Cox regression analysis in the high-dimensional and low-sample size settings, with applications to microarray gene expression data," *Bioinformatics*, vol. 21, no. 13, pp. 3001–3008, 2005.
- [14] L. Evers and C.-M. Messow, "Sparse kernel methods for high-dimensional survival data," *Bioinformatics*, vol. 24, no. 14, pp. 1632–1638, 2008.
- [15] W. Cheng, X. Ren, J. Cai et al., "A five-miRNA signature with prognostic and predictive value for MGMT promoter-methylated glioblastoma patients," *Oncotarget*, vol. 6, no. 30, pp. 29285–29295, 2015.
- [16] J. Cai, W. Zhang, P. Yang et al., "Identification of a 6-cytokine prognostic signature in patients with primary glioblastoma harboring M2 microglia/macrophage phenotype relevance," *PLoS ONE*, vol. 10, no. 5, Article ID 0126022, 2015.
- [17] Z.-S. Bao, M.-Y. Li, J.-Y. Wang et al., "Prognostic value of a nine-gene signature in glioma patients based on mRNA expression profiling," *CNS Neuroscience and Therapeutics*, vol. 20, no. 2, pp. 112–118, 2014.
- [18] W. Zhang, J. Zhang, W. Yan et al., "Whole-genome microRNA expression profiling identifies a 5-microRNA signature as a prognostic biomarker in Chinese patients with primary glioblastoma multiforme," *Cancer*, vol. 119, no. 4, pp. 814–824, 2013.
- [19] L. Xu, "Bi-linear matrix-variate analyses, integrative hypothesis tests, and case-control studies," *Applied Informatics*, vol. 2, no. 1, pp. 1–39, 2015.
- [20] I. S. Vlachos, K. Zagganas, M. D. Paraskevopoulou et al., "DIANA-miRPath v3.0: deciphering microRNA function with experimental support," *Nucleic Acids Research*, vol. 43, no. 1, pp. W460–W466, 2015.
- [21] I. S. Vlachos, M. D. Paraskevopoulou, D. Karagkouni et al., "DIANA-TarBase v7.0: indexing more than half a million experimentally supported miRNA:mRNA interactions," *Nucleic Acids Research*, vol. 43, no. 1, pp. D153–D159, 2015.
- [22] D. W. Hosmer, S. Lemeshow, and S. May, *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*, Wiley-Interscience, A John Wiley & Sons, Inc., Hoboken, NJ, USA, 2008.
- [23] S. Ma and X. Song, "Ranking prognosis markers in cancer genomic studies," *Briefings in Bioinformatics*, vol. 12, no. 1, pp. 33–40, 2011.
- [24] D. Dunkler, M. Schemper, and G. Heinze, "Gene selection in microarray survival studies under possibly non-proportional hazards," *Bioinformatics*, vol. 26, no. 6, pp. 784–790, 2010.
- [25] E. C. Burton, K. R. Lamborn, P. Forsyth et al., "Aberrant p53, mdm2, and proliferation differ in glioblastomas from long-term compared with typical survivors," *Clinical Cancer Research*, vol. 8, no. 1, pp. 180–187, 2002.
- [26] G. Fulci, M. Labuhn, D. Maier et al., "p53 gene mutation and ink4a-arf-deletion appear to be two mutually exclusive events in human glioblastoma," *Oncogene*, vol. 19, no. 33, pp. 3816–3822, 2000.