

Original Article

Identification of prognostic genes in paediatric medulloblastoma from mRNA expression profiles

Changjun Cao, Wei Wang, Pucha Jiang

Department of Neurosurgery, Zhongnan Hospital of Wuhan University, China

Received October 15, 2015; Accepted March 1, 2016; Epub October 15, 2016; Published October 30, 2016

Abstract: Medulloblastoma is the most common malignant brain tumour of childhood. The identification of prognostic biomarkers correlated with overall survival remains a crucial step towards the refinement of medulloblastoma treatment. A total of 100 medulloblastoma samples from two independent cohorts were included in this study. The statistical modelling approach, Bayesian Model Averaging algorithm, was used to discover the prognostic biomarkers. Six genes including BICD2, CD300LG, RAB21, RAD18, SYNRG and TNFSF13 were identified to be related to medulloblastoma overall survival. We demonstrated this six-gene signature could successfully discriminate low-risk group from high-risk groups in two independent medulloblastoma cohorts. We have successfully identified a six-gene medulloblastoma prognostic signature. We anticipate that these genes could serve as biomarkers or drug targets in personalised therapy of medulloblastoma.

Keywords: Prognosis, medulloblastoma, mRNA expression, bayesian model averaging

Introduction

Medulloblastoma is the most common malignant brain tumour of childhood and accounts for 15%-20% of all paediatric primary brain tumours [1]. Although there is approximately 70% improvement in 5-year survival rates of standard-risk patients as a result of advances in treatment regimens, many survival children suffered from long-term neurologic and endocrinologic adverse effects [2, 3]. Despite the improvements in survival rates, approximately 30% of patients are incurable [4].

The identification of prognostic biomarkers correlated with overall survival remains a crucial step towards the refinement of medulloblastoma treatment. The elevated expression level of NTRK3 (neurotrophic tyrosine kinase, receptor, type 3) indicated good prognosis [5-7], while the presence of high expression of CDK6 has been associated with poorer prognosis [8]. With the advances of high-throughput mRNA expression profiling technologies, known as microarray, genes related to cell proliferation, transcription and mitosis showed promising results in predicting medulloblastoma outcome

[9]. The ability of measuring mRNA expression levels of thousands of genes at one time has made the microarray technology a promising direction in cancer research. Based on microarray data, many gene signatures have also been identified for the prediction of medulloblastoma prognosis [10, 11].

One major aim in microarray-based survival analysis is to subset a few predictive candidates from large amount of measured genes in the analysed samples. In this way, the significantly smaller set of biomarkers would make clinical use more affordable in terms of time and cost. However, most existing prognostic biomarker identification methods rely on univariate analysis [12-14], which considers the expression profile of each gene individually. The multivariate analysis, on the other hand, evaluates multiple genes simultaneously and identifies predictive genes in combinations.

In the present study, we investigated the Bayesian Model Averaging (BMA) based approach on mRNA expression profiles from a large cohort of primary medulloblastoma samples to generate novel prognostic genes that

Identification of prognostic genes in medulloblastoma

Table 1. Characteristic of clinical samples

	Dataset 1, n=61	Dataset 2, n=39
NCBI GEO accession number	GSE10327	GSE12992
Age under 3		
Yes	15	4
No	46	35
Histology		
Anaplastic	1	2
Classic	44	31
Desmoplastic	13	3
NA	3	3

NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>).

could be used to predict patient overall survival. The BMA algorithm is among multivariate gene selection strategies and has many advantages [15, 16]. It is computationally efficient and systematically determines the number of predictive genes and models. The final selected models usually consists only a few genes. We anticipate that these genes could serve as biomarkers or drug targets in personalised therapy of medulloblastoma.

Materials and methods

Patient samples

As shown in **Table 1**, a total of 100 samples from two independent cohorts were selected for the present study. Tumours were classified as classic (n=75; 75%), anaplastic (n=3; 3%) or desmoplastic (n=16; 16%), while no histological information was provided with 6 samples. 19% of the patients at diagnosis of medulloblastoma were under the age of three. Median survival was 41 months (range: 2-277 months).

Microarray experiments

Although the microarray experiments were conducted in two different institutes, the researchers literally following the same protocol to generate mRNA expression datasets [17, 18]. In summary, 4 µg total RNA was used for cRNA synthesis and fragmented. Labelling was performed with One-Cycle cDNA Synthesis Kit (Affymetrix, Santa Clara, California) according to manufacturer's protocol. Sample quality was checked on a Bioanalyzer prior and after fragmentation. 10 µg of labelled cRNA was hybridized to Affymetrix Human Genome U133 Plus 2.0 arrays according to manufacturer's protocol (Affymetrix, Santa Clara, California). Arrays

were scanned with a GeneChip Scanner 3000 (Affymetrix, Santa Clara, California) according to manufacturer's protocol.

Combination of independent medulloblastoma cohorts

Expression datasets from the two independent cohorts were obtained from the National Centre for Biotechnical Information Gene Expression Omnibus (<http://www.ncbi.nlm.nih.gov/geo/>) with series accession numbers (dataset 1: GSE10327; dataset 2: GSE12992). The Ethical Committee of the Zhongnan Hospital, Wuhan University, China, approved the research. Since the same experimental platform was used datasets from two cohorts were pre-processed together. Data from 2 patients (GEO accession numbers: GSM324138 and GSM260981) were excluded to avoid surgery related complications because they died within one month after surgery. The expression datasets were normalized using RMA (Robust Multi-array Average) algorithm [19] in the Bioconductor package (<http://www.bioconductor.org/>) of R statistical environment (<http://www.r-project.org/>). Finally, the matrix consisting of mRNA expression datasets from 100 medulloblastoma patients was generated for BMA modelling.

Bayesian model averaging for survival analysis

We implemented the BMA algorithm for medulloblastoma survival analysis by using a Bioconductor package, iterativeBMASurv [20]. Firstly, the genes were ranked in the descending order of their log likelihood using the Cox Proportional Hazards Model [21]. Then the BMA algorithm was applied on the 10 top log-ranked genes. In the next step, genes to which the BMA assigns low posterior probabilities (less than 1%) of being in the predictive model are removed. If *n* genes are removed, the next *n* genes from the previously ranked list are added back to the set of genes and the BMA algorithm will be applied again. These steps continue until all genes are considered.

The BMA training process was applied on all the 100 samples. After the prognostic model was established, it was tested on the two independent datasets respectively.

Identification of prognostic genes in medulloblastoma

Table 2. List of the identified six prognostic genes in medulloblastoma

Official symbol	Description	Gene ID
BICD2	Bicaudal D homolog 2 (Drosophila)	23299
CD300LG	CD300 molecule-like family member g	146894
RAB21	RAB21, member RAS oncogene family	23011
RAD18	RAD18 E3 ubiquitin protein ligase	56852
SYNRG	Synergina, gamma	11276
TNFSF13	Tumor necrosis factor (ligand) superfamily, member 13	8741

RAB21 were significantly enriched in one the cellular component category: Golgi apparatus (GO term accession: GO: 00-05794). However, the prognostic gene signatures were not significantly enriched in any KEGG pathways.

Characterisation of selected medulloblastoma prognostic genes

The selected medulloblastoma prognostic genes were categorized in Gene Ontology groups and mapped to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways using the Database for Annotation, Visualization and Integrated Discovery (DAVID, <http://david.abcc.ncifcrf.gov/>) software [22, 23].

Results

Identification of six genes associated with medulloblastoma overall survival

In the clinical practice of medulloblastoma therapy, using a small set of genes (biomarkers) to predict overall survival reduces the costs associated with high throughput time-consuming data analysis. We managed to identify six genes (**Table 2**) to distinguish between low-risk and high-risk medulloblastoma samples. Next, the BMA classifier was applied in two individual datasets (**Table 1**). From **Figure 1**, the built classifier successfully yielded high confidence (Dataset 1: Log-rank test P -value = $1.06e-06$; Dataset 2: Log_r test P -value = $1.11e-10$) in dividing the test samples into two groups: high-risk and low-risk. It is evident that the BMA based approach outperformed standard microarray analysis.

Characterisation of the six prognostic genes

The identified six prognostic genes include BICD2 (bicaudal D homolog 2 (Drosophila)), CD300LG (CD300 molecule-like family member g), RAB21 (RAB21, member RAS oncogene family), RAD18 (RAD18 E3 ubiquitin protein ligase), SYNRG (synergina, gamma) and TNFSF13 (tumour necrosis factor (ligand) superfamily, member 13). From the Gene Ontology analysis, three genes including SYNRG, BICD2 and

Discussion

In the present study, we have successfully applied machine-learning approaches to identify six overall survival-associated genes in medulloblastoma to meet the pressing challenge in clinical treatment. Because this gene panel consists of rather smaller number of genes, mRNA expression could be easily assessed by quantitative PCR (qPCR) experiments; we believe that this could establish a practical and inexpensive molecular diagnostic tool for clinical use in the near future.

To our knowledge, none of the six genes has been reported in previous medulloblastoma prognosis related researches. Only one gene, TNFSF13, was reported to be associated with prognostic prediction in other disease. The protein encoded by this gene, also known as APRIL (A proliferation-inducing ligand), belongs to the tumour necrosis factor (TNF) ligand family. It is reported that TNFSF13 could serve as a prognostic biomarker in non-small cell lung cancer [24], pancreatic cancer [25], B-Cell Chronic lymphocytic leukemia [26-28] and neuromyelitis optica [29]. It is also could serve as a clinical chemo-resistance biomarker in 5FU-treated colorectal adenocarcinoma patients [30].

For the future research, we plan to collect more datasets for training to improve the prognostic classifier's performance and robustness. On the other hand, further experiments are also needed to investigate these genes' biological behaviours in medulloblastoma.

Conclusion

In conclusion, we have successfully identified a six-gene signature that could distinguish high- and low-risk medulloblastoma patients. However, further validation is required before these predictive genes could serve as biomarker in medulloblastoma clinical treatments.

Identification of prognostic genes in medulloblastoma

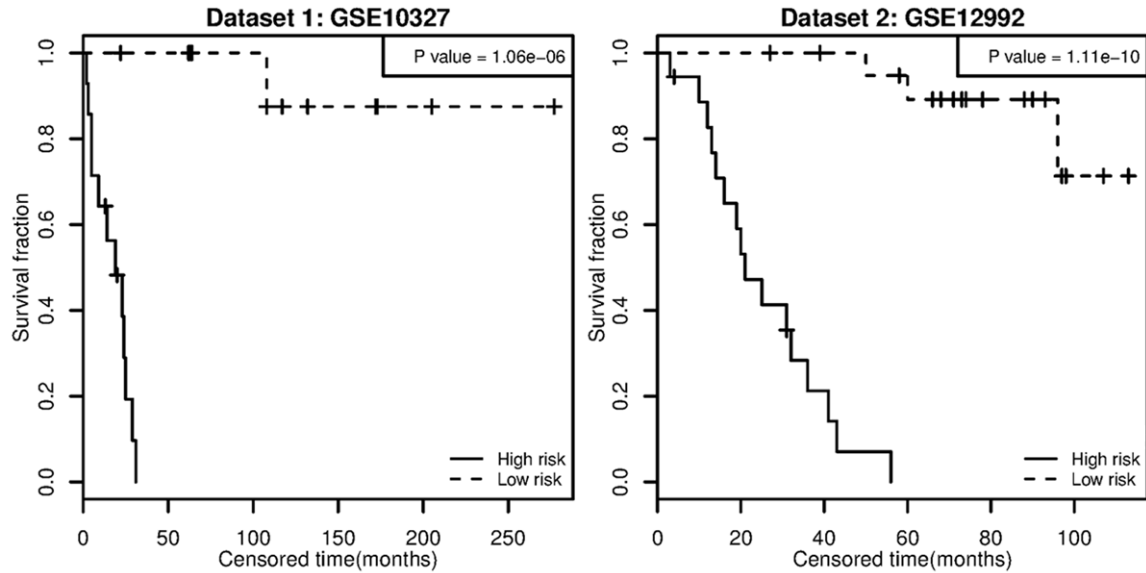


Figure 1. Survival analysis of two independent datasets using the survival models.

Disclosure of conflict of interest

None.

Address correspondence to: Changjun Cao, Department of Neurosurgery, Zhongnan Hospital of Wuhan University, 169 Donghu Road, Wuhan 430071, Hubei, China. E-mail: caocj2030@163.com

References

- [1] Crawford JR, MacDonald TJ and Packer RJ. Medulloblastoma in childhood: new biological advances. *Lancet Neurol* 2007; 6: 1073-1085.
- [2] Mulhern RK, Merchant TE, Gajjar A, Reddick WE and Kun LE. Late neurocognitive sequelae in survivors of brain tumours in childhood. *Lancet Oncol* 2004; 5: 399-408.
- [3] Rood BR, Macdonald TJ and Packer RJ. Current treatment of medulloblastoma: recent advances and future challenges. *Semin Oncol* 2004; 31: 666-675.
- [4] Pizer BL and Clifford SC. The potential impact of tumour biology on improved clinical practice for medulloblastoma: progress towards biologically driven clinical trials. *Br J Neurosurg* 2009; 23: 364-375.
- [5] Kim JY, Sutton ME, Lu DJ, Cho TA, Goumnerova LC, Goritchenko L, Kaufman JR, Lam KK, Billet AL, Tarbell NJ, Wu J, Allen JC, Stiles CD, Segal RA and Pomeroy SL. Activation of neurotrophin-3 receptor TrkC induces apoptosis in medulloblastomas. *Cancer Res* 1999; 59: 711-719.
- [6] Grotzer MA, Janss AJ, Fung K, Biegel JA, Sutton LN, Rorke LB, Zhao H, Cnaan A, Phillips PC, Lee VM and Trojanowski JQ. TrkC expression predicts good clinical outcome in primitive neuroectodermal brain tumors. *J Clin Oncol* 2000; 18: 1027-1035.
- [7] Castellino RC, De Bortoli M, Lin LL, Skapura DG, Rajan JA, Adesina AM, Perlaky L, Irwin MS and Kim JY. Overexpressed TP73 induces apoptosis in medulloblastoma. *BMC Cancer* 2007; 7: 127.
- [8] Mendrzyk F, Radlwimmer B, Joos S, Kokocinski F, Benner A, Stange DE, Neben K, Fiegler H, Carter NP, Reifenberger G, Korshunov A and Lichter P. Genomic and protein expression profiling identifies CDK6 as novel independent prognostic marker in medulloblastoma. *J Clin Oncol* 2005; 23: 8853-8862.
- [9] Neben K, Korshunov A, Benner A, Wrobel G, Hahn M, Kokocinski F, Golanov A, Joos S and Lichter P. Microarray-based screening for molecular markers in medulloblastoma revealed STK15 as independent predictor for survival. *Cancer Res* 2004; 64: 3103-3111.
- [10] Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES and Golub TR. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature* 2002; 415: 436-442.
- [11] MacDonald TJ, Brown KM, LaFleur B, Peterson K, Lawlor C, Chen Y, Packer RJ, Cogen P and Stephan DA. Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK path-

Identification of prognostic genes in medulloblastoma

- way as therapeutic targets for metastatic disease. *Nat Genet* 2001; 29: 143-152.
- [12] Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD and Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; 286: 531-537.
- [13] Nguyen DV and Rocke DM. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* 2002; 18: 39-50.
- [14] Ben-Dor A, Bruhn L, Friedman N, Nachman I, Schummer M and Yakhini Z. Tissue classification with gene expression profiles. *J Comput Biol* 2000; 7: 559-583.
- [15] Hoeting JA, Madigan D, Raftery AE and Volinsky CT. Bayesian model averaging: a tutorial (with comments by M. Clyde, David Draper and E. I. George, and a rejoinder by the authors. 1999; 382-417.
- [16] Yeung KY, Bumgarner RE and Raftery AE. Bayesian model averaging: development of an improved multi-class, gene selection and classification tool for microarray data. *Bioinformatics* 2005; 21: 2394-2402.
- [17] Kool M, Koster J, Bunt J, Hasselt NE, Lakeman A, van Sluis P, Troost D, Meeteren NS, Caron HN, Cloos J, Mrsic A, Ylstra B, Grajkowska W, Hartmann W, Pietsch T, Ellison D, Clifford SC and Versteeg R. Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS One* 2008; 3: e3088.
- [18] Fattet S, Haberler C, Legoux P, Varlet P, Lelouch-Tubiana A, Lair S, Manie E, Raquin MA, Bours D, Carpentier S, Barillot E, Grill J, Doz F, Puget S, Janoueix-Lerosey I and Delattre O. Beta-catenin status in paediatric medulloblastomas: correlation of immunohistochemical expression with mutational status, genetic profiles, and clinical characteristics. *J Pathol* 2009; 218: 86-94.
- [19] Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U and Speed TP. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 2003; 4: 249-264.
- [20] Annest A, Bumgarner RE, Raftery AE and Yeung KY. Iterative Bayesian Model Averaging: a method for the application of survival analysis to high-dimensional microarray data. *BMC Bioinformatics* 2009; 10: 72.
- [21] Prentice R. Introduction to Cox (1972) Regression Models and Life-Tables. In: Kotz S, Johnson N, editors. *Breakthroughs in Statistics*. New York: Springer; 1992. pp. 519-526.
- [22] Huang da W, Sherman BT and Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009; 4: 44-57.
- [23] Huang DW, Sherman BT and Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009; 37: 1-13.
- [24] Qian Z, Qingshan C, Chun J, Huijun Z, Feng L, Qiang W, Qiang X and Min Z. High expression of TNFSF13 in tumor cells and fibroblasts is associated with poor prognosis in non-small cell lung cancer. *Am J Clin Pathol* 2014; 141: 226-233.
- [25] Wang F, Chen L, Ding W, Wang G, Wu Y, Wang J, Luo L, Cong H, Wang Y, Ju S, Shao J and Wang H. Serum APRIL, a potential tumor marker in pancreatic cancer. *Clin Chem Lab Med* 2011; 49: 1715-1719.
- [26] Tecchio C, Nichele I, Mosna F, Zampieri F, Leso A, Al-Khaffaf A, Veneri D, Andreini A, Pizzolo G and Ambrosetti A. A proliferation-inducing ligand (APRIL) serum levels predict time to first treatment in patients affected by B-cell chronic lymphocytic leukemia. *Eur J Haematol* 2011; 87: 228-234.
- [27] Bojarska-Junak A, Hus I, Chocholska S, Wasik-Szczepanek E, Sieklucka M, Dmoszynska A and Rolinski J. BAFF and APRIL expression in B-cell chronic lymphocytic leukemia: correlation with biological and clinical features. *Leuk Res* 2009; 33: 1319-1327.
- [28] Kern C, Cornuel JF, Billard C, Tang R, Rouillard D, Stenou V, Defrance T, Ajchenbaum-Cymbalista F, Simonin PY, Feldblum S and Kolb JP. Involvement of BAFF and APRIL in the resistance to apoptosis of B-CLL through an auto-crine pathway. *Blood* 2004; 103: 679-688.
- [29] Pellkofer HL, Krumbholz M, Berthele A, Hemmer B, Gerdes LA, Havla J, Bittner R, Canis M, Meinel E, Hohlfeld R and Kuempfel T. Long-term follow-up of patients with neuromyelitis optica after repeated therapy with rituximab. *Neurology* 2011; 76: 1310-1315.
- [30] Petty RD, Samuel LM, Murray GI, MacDonald G, O'Kelly T, Loudon M, Binnie N, Aly E, McKinlay A, Wang W, Gilbert F, Semple S and Collie-Duguid ES. APRIL is a novel clinical chemoresistance biomarker in colorectal adenocarcinoma identified by gene expression profiling. *BMC Cancer* 2009; 9: 434.