Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

# Answering Reviewers

**Name of journal:** World Journal of Gastroenterology
**Manuscript NO:** 38835
**Title:** Evaluation of the prognostic power of liver cancer by molecular marker
**Reviewer's code:** 00503516
**Reviewer's country:** Italy
**Science editor:** Xue-Jiao Wang
**Date sent for review:** 2018-03-22
**Date reviewed:** 2018-03-26
**Review time:** 4 Days

| CLASSIFICATION | LANGUAGE EVALUATION | SCIENTIFIC MISCONDUCT | CONCLUSION |
|---|---|---|---|
| [ ] Grade A: Excellent | [ Y] Grade A: Priority publishing | Google Search: | [ ] Accept |
| [ Y] Grade B: Very good | [ ] Grade B: Minor language polishing | [ ] The same title | [ ] High priority for publication |
| [ ] Grade C: Good | [ ] Grade C: A great deal of language polishing | [ ] Duplicate publication | [ ] Rejection |
| [ ] Grade D: Fair | [ ] Grade D: Rejected | [ ] Plagiarism | [ Y] Minor revision |
| [ ] Grade E: Poor | | [ ] No | [ ] Major revision |
| | | BPG Search: | |
| | | [ ] The same title | |
| | | [ ] Duplicate publication | |
| | | [ ] Plagiarism | |
| | | [ ] No | |

## COMMENTS TO AUTHORS

Yanzhou Song et. al systemically evaluated the prognostic power of different omics data for liver cancer using as source data the Cancer Genome Atlas. The work is well written and informative and can contribute to expand our knowledge with regard to the identification of HCC prognostic/diagnostic markers. I suggest the author to better specify which are, among the individuated pathways, (WNT/beta-catenin, E2F targets, mitotic spindle and G2M checkpoint) the specify variations in gene expression levels. For example, which are the E2F targets differently regulated? Do they include E2F family members? This piece of information would be very interesting as not all E2F family members are thought to play the same role in HCC.

**Answer**: Thanks so much for your kindly suggestion.
In our manuscript, the Gene Set Enrichment Analysis (GSEA) is a computational method

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

that determines whether an a priori defined set of genes shows statistically significant, concordant differences between two biological states. All the gene sets, including the E2F targets genes, could be found at The Molecular Signatures Database (MSigDB) (http://software.broadinstitute.org/gsea/msigdb/index.jsp).

For example, the significant the E2F targets, it means most of the highly expressed genes were enriched in the E2F targets gene sets. However, not all E2F targets were highly expressed. It indicated that in the view of the highly expressed genes, they might be involved in the E2F targets pathway. Thus the E2F targets pathway might be the main reason that result in the difference between the two groups.

It's really interesting as not all E2F family members are thought to play the same role in HCC. We would like to demonstrate these mechanisms by molecular experiments in the future research. Thanks so much!

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

# Answering Reviewers

**Name of journal:** World Journal of Gastroenterology
**Manuscript NO:** 38835
**Title:** Evaluation of the prognostic power of liver cancer by molecular marker
**Reviewer's code:** 01558002
**Reviewer's country:** Greece
**Science editor:** Xue-Jiao Wang
**Date sent for review:** 2018-03-28
**Date reviewed:** 2018-03-31
**Review time:** 3 Days

| CLASSIFICATION | LANGUAGE EVALUATION | SCIENTIFIC MISCONDUCT | CONCLUSION |
|---|---|---|---|
| [ ] Grade A: Excellent | [ ] Grade A: Priority publishing | Google Search: | [ ] Accept |
| [ ] Grade B: Very good | [ ] Grade B: Minor language polishing | [ ] The same title | [ ] High priority for publication |
| [Y] Grade C: Good | [Y] Grade C: A great deal of language polishing | [ ] Duplicate publication | [ ] Rejection |
| [ ] Grade D: Fair | [ ] Grade D: Rejected | [ ] Plagiarism | [Y] Minor revision |
| [ ] Grade E: Poor | | [ ] No | [ ] Major revision |
| | | BPG Search: | |
| | | [ ] The same title | |
| | | [ ] Duplicate publication | |
| | | [ ] Plagiarism | |
| | | [ ] No | |

## COMMENTS TO AUTHORS

For clinical researchers lacking bioinformatics expertise, extrapolating desired information from the copious amounts of data supplied by The Cancer Genome Atlas (TCGA) proves to be a difficult task. The authors are congratulated on accomplishing the task using molecular and clinical data on the TCGA, in which specific genes were shown to correlate with survival for patients with liver cancer. However, English writing, including the title, abstract and text, is suboptimal. The title may be changed to "Integrated genomic analysis for prediction of survival for patients with liver cancer using the Cancer Genome Atlas". Also, the abstract appears not to summarize the study results. The abbreviations, such as LASSO and CNP, should be defined on their first use. An additional figure, which explains the statistical process (algorithm), would improve the readability.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

**Answer**: Thanks so much for your kindly suggestion.

1. We have changed the title to "Integrated genomic analysis for prediction of survival for patients with liver cancer using the Cancer Genome Atlas".

2. We have revised the abstract part, to better summarize the study results. The revised part was highlighted in the manuscript, as follows:

ABSTRACT

AIM

To evaluate the prognostic power of different molecular data in liver cancer.

METHODS

Cox regression screen and least absolute shrinkage and selection operator (LASSO) were performed to select significant prognostic variables. Then the concordance index was calculated to evaluate the prognostic power. For the combination data, based on the clinical cox model, molecular features that better fit the model were combined to calculate the concordance index. Prognostic models were built based on the arithmetic summation of the significant variables. Kaplan-Meier survival curve and log-rank test were performed to compare the survival difference. Then the heatmap was constructed and gene set enrichment analysis was performed for pathway analysis.

RESULTS

mRNA data was the most informative prognostic variables in all kinds of omics data in liver cancer, with the highest C-index of 0.61. In the copy number variation (CNV), methylation and miRNA data, the combination of molecular data with clinical data could significantly boost the prediction accuracy of the molecular data alone ($P < 0.05$). On the other hand, the combination of clinical data with methylation, miRNA and mRNA data could significantly boost the prediction accuracy of the clinical data itself ($P < 0.05$). Based on the significant prognostic variables, different prognostic models were built. In addition, the heatmap, survival analysis, and gene set enrichment analysis validated the practicability of the prognostic models.

CONCLUSION

In all kinds of omics data in liver cancer, the mRNA data might be the most informative prognostic variables. The combination of clinical data with molecular data might be the future direction for cancer prognosis and prediction.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

3. We have defined the abbreviations, such as LASSO and CNP, on their first use.

4. We have also added an additional figure (Figure 1) to explain the statistical process (algorithm).

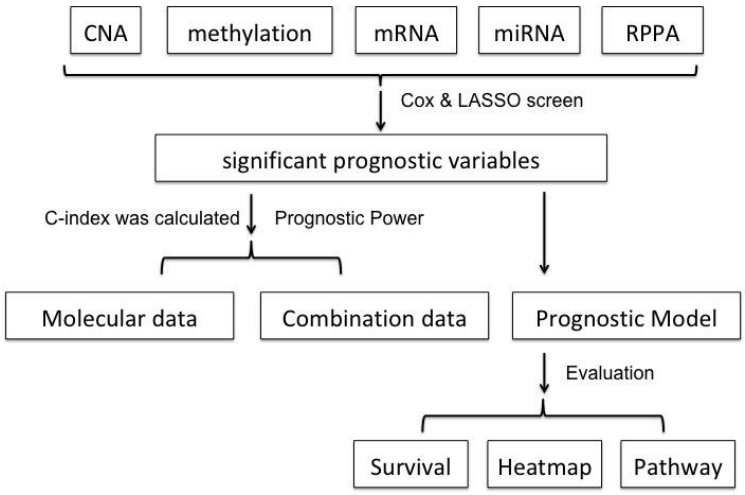


Figure 1 The statistical process (algorithm). Cox regression screen and least absolute shrinkage and selection operator (LASSO) were performed to select significant prognostic variables. Then the concordance index was calculated to evaluate the prognostic power. Prognostic models were built based on the arithmetic summation of the significant variables. Kaplan-Meier survival curve and log-rank test were performed to compare the survival difference. Then the heatmap was constructed and gene set enrichment analysis was performed for pathway analysis.

5. We have asked several doctors who have been studying abroad for several years to help improve the English writing of the paper. The revised part was highlighted in the manuscript.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

# Answering Reviewers

**Name of journal:** World Journal of Gastroenterology
**Manuscript NO:** 38835
**Title:** Evaluation of the prognostic power of liver cancer by molecular marker
**Reviewer's code:** 00053419
**Reviewer's country:** Spain
**Science editor:** Xue-Jiao Wang
**Date sent for review:** 2018-03-22
**Date reviewed:** 2018-04-03
**Review time:** 12 Days

| CLASSIFICATION | LANGUAGE EVALUATION | SCIENTIFIC MISCONDUCT | CONCLUSION |
|---|---|---|---|
| [ ] Grade A: Excellent | [ ] Grade A: Priority publishing | Google Search: | [ ] Accept |
| [Y] Grade B: Very good | [Y] Grade B: Minor language polishing | [ ] The same title | [ ] High priority for publication |
| [ ] Grade C: Good | [ ] Grade C: A great deal of language polishing | [ ] Duplicate publication | [ ] Rejection |
| [ ] Grade D: Fair | [ ] Grade D: Rejected | [ ] Plagiarism | [Y] Minor revision |
| [ ] Grade E: Poor | | [ ] No | [ ] Major revision |
| | | BPG Search: | |
| | | [ ] The same title | |
| | | [ ] Duplicate publication | |
| | | [ ] Plagiarism | |
| | | [ ] No | |

## COMMENTS TO AUTHORS

The authors report a meta analysis of high density molecular data from TCGA experiments on HCC. Genetic, epigenetic and expression data (miRNA and mRNA) have been considered in the analysis. Regrettably protein data did not reach statistical significance to support prognostic power. Perhaps the reasons explaining why protein profiling has no prognostic value in this case should be further discussed. Is it a consequence of the analytical methodology used or, alternatively is it associated to the complexity of the proteome? Minor points are as follows: Figure 1 legend should be extended to make the figure self-explanatory. Heatmaps in figures 2, 3, 4 and 5 are too small. Figure legend extension is also recommended in these cases.

**Answer**: Thanks so much for your kindly suggestion.
1. We've added the reasons why we did not include the protein model in the prognostic

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

models as follows:
With respect to the prognostic models based on different omics data, we did not include the protein model, since no significant prognostic variable passed through the cox screen and LASSO analysis. We suppose that one reason is the relative sample size of the patients with protein data. The other reason is due to the complexity of the proteome data.

2. We've extended Figure 1,2,3,4,5 legends to make the points more clear, which were highlighted in the manuscript.

3. We've also magnified the heatmaps in figures 2, 3, 4 and 5, which were highlighted in the manuscript.

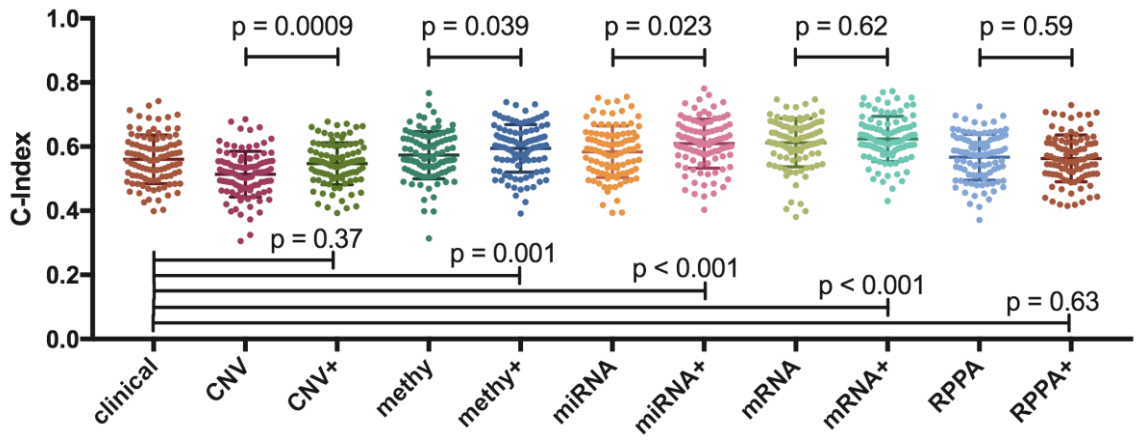


**Figure 2** The prognostic power of clinical data and different types of molecular data. The c-index value on the left Y-axis indicated the prognostic power of each data type. The p values on top half of the figure represented the comparisons between the molecular data alone and the combination data. The p values on lower half of the figure represented the comparisons between the clinical data alone and the combination data.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

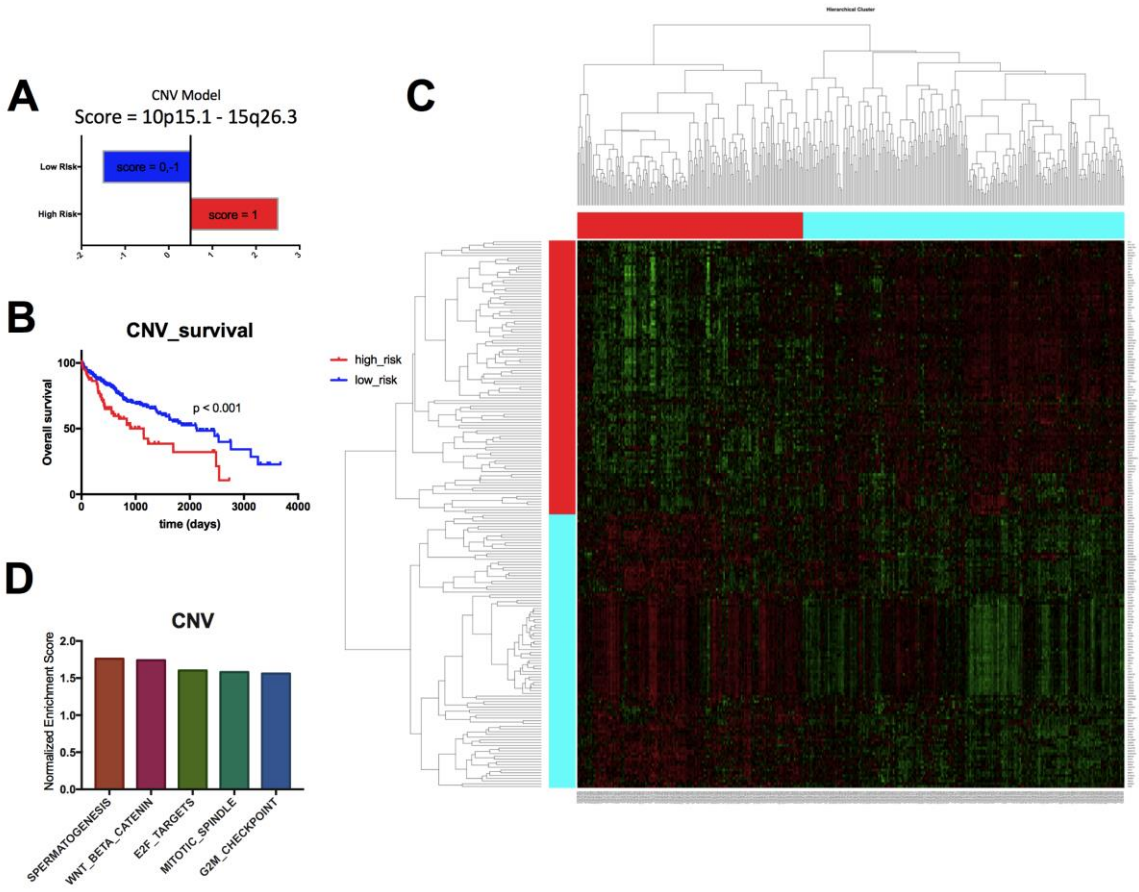


**Figure 3** Establishment of the prognostic model based on the copy number data.

(A) The high risk group and low risk group based on the prognostic score. The patients with the score of 1 were considered as high risk, and patients with the score of 0 and -1 were considered as low risk.

(B) The Kaplan-Meier survival curves of the high risk group and low risk group, which showed that there was significant difference of survival between the high risk patients and low risk patients.

(C) The heatmap showing the different gene expression patterns of the high risk group and low risk group. It showed that the gene expression pattern of high risk group and low risk group were obvious distinct.

(D) The top enriched pathways in the high risk group, as indicated by the gene set enrichment analysis. It showed that the spermatogenesis, WNT/beta-catenin, E2F targets, mitotic spindle and G2M checkpoint were the top 5 enriched pathways in the high risk group patients.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

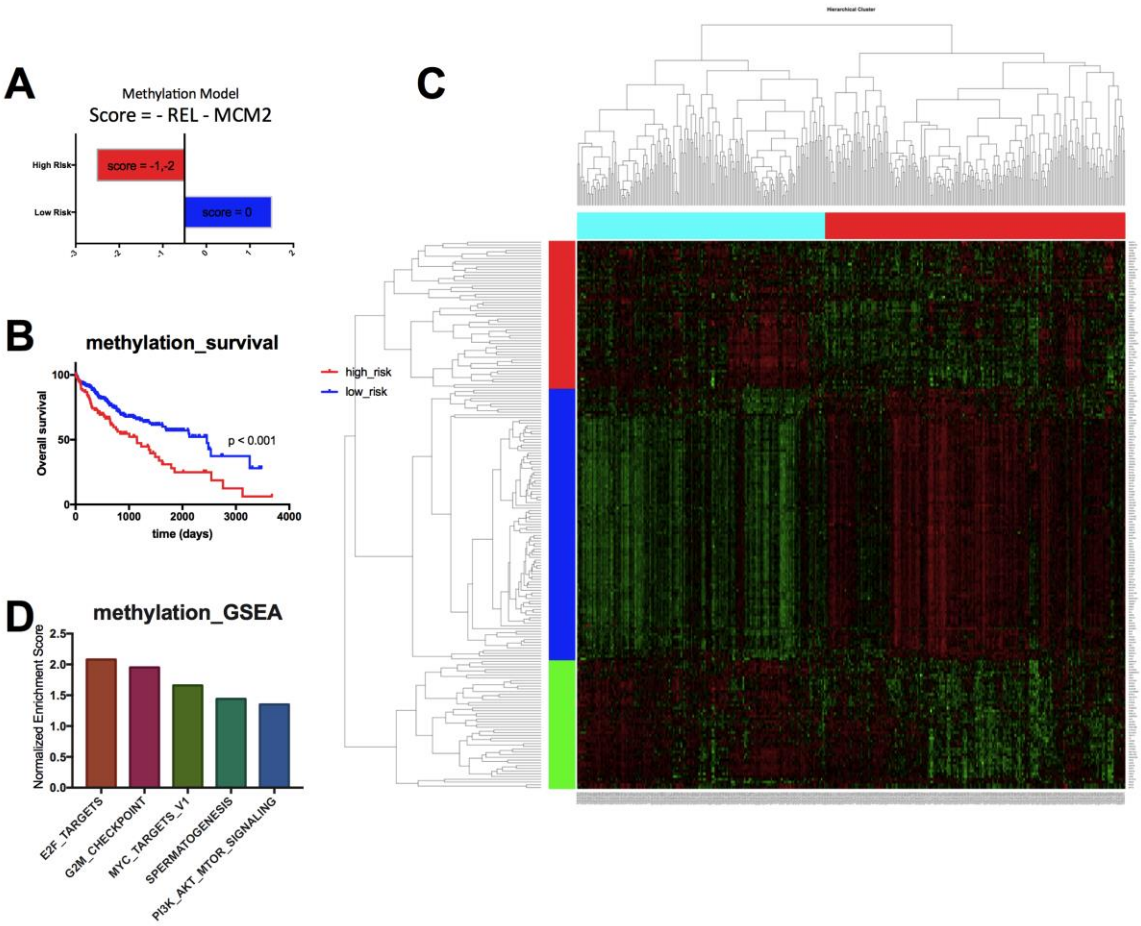


**Figure 4** Establishment of the prognostic model based on the methylation data.

(A) The high risk group and low risk group based on the prognostic score. The patients with the score of 0 were considered as high risk, and patients with the score of -1 and -2 were considered as low risk.

(B) The Kaplan-Meier survival curves of the high risk group and low risk group, which showed that there was significant difference of survival between the high risk patients and low risk patients.

(C) The heatmap showing the different gene expression patterns of the high risk group and low risk group. It showed that the gene expression pattern of high risk group and low risk group were obvious distinct.

(D) The top enriched pathways in the high risk group, as indicated by the gene set enrichment analysis. It showed that the E2F targets, G2M checkpoint, Myc targets V1, spermatogenesis and

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

PI3K/AKT/mTOR pathway signaling were among the top five enriched pathways in the high risk group of patients.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

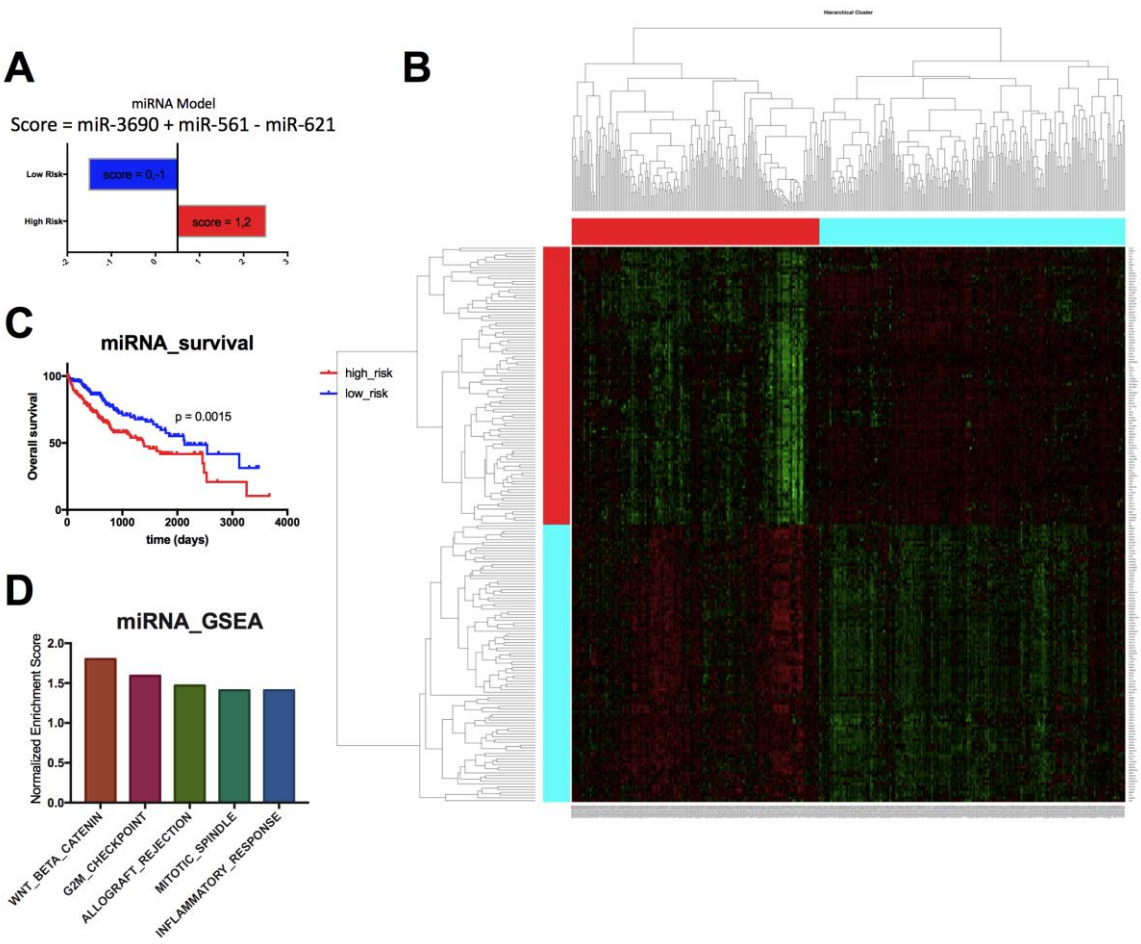


**Figure 5** Establishment of the prognostic model based on the miRNA data.

(A) The high risk group and low risk group based on the prognostic score. The patients with the score of 1 and 2 were considered as high risk, and patients with the score of 0 and -1 were considered as low risk.

(B) The Kaplan-Meier survival curves of the high risk group and low risk group, which showed that there was significant difference of survival between the high risk patients and low risk patients.

(C) The heatmap showing the different gene expression patterns of the high risk group and low risk group. It showed that the gene expression pattern of high risk group and low risk group were obvious distinct.

(D) The top enriched pathways in the high risk group, as indicated by the gene set enrichment analysis. It showed that the WNT/beta-catenin, G2M checkpoint, allograft rejection, mitotic spindle,

Baishideng Publishing Group
7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

and inflammatory response were among the top 5 enriched pathways in the high risk group patients.

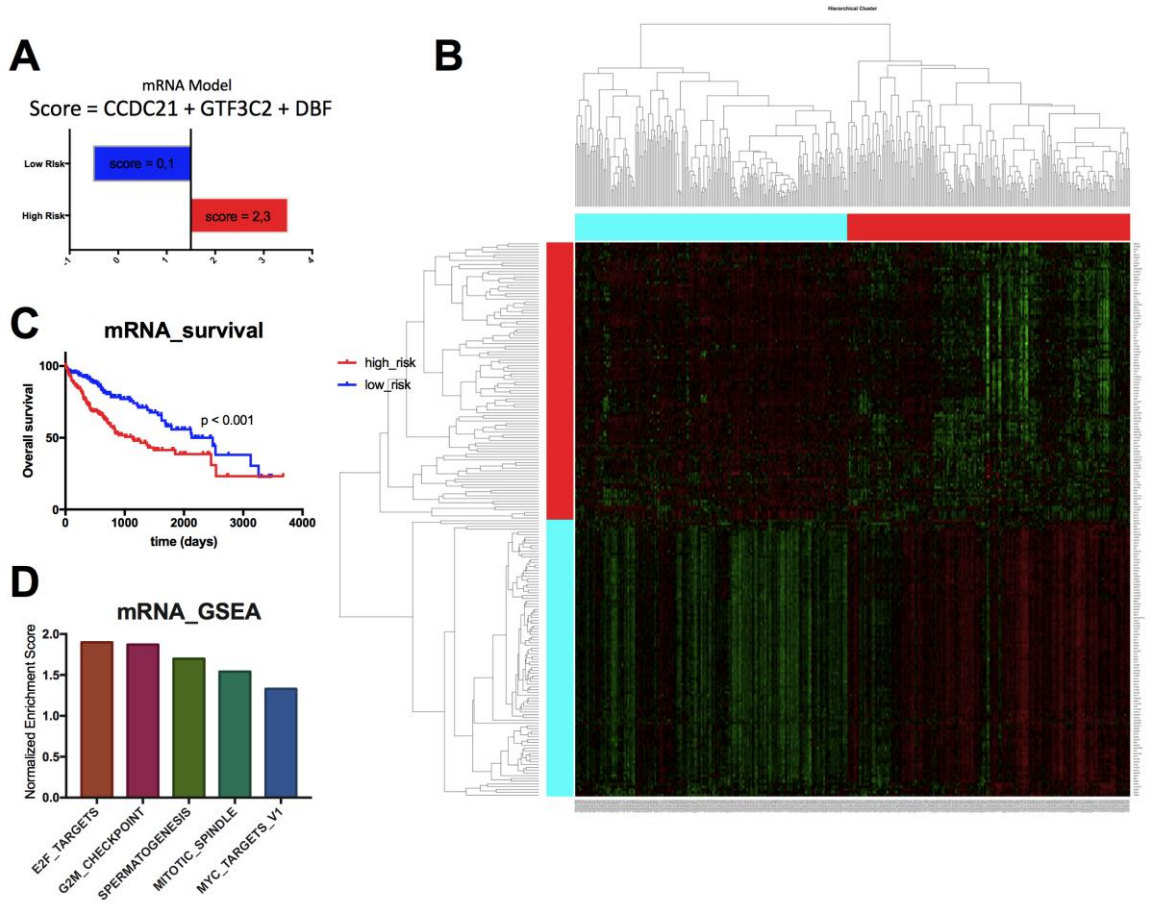


**Figure 6** Establishment of the prognostic model based on the mRNA data.

(A) The high risk group and low risk group based on the prognostic score. The patients with the score of 2 and 3 were considered as high risk, and patients with the score of 0 and 1 were considered as low risk.

(B) The Kaplan-Meier survival curves of the high risk group and low risk group, which showed that there was significant difference of survival between the high risk patients and low risk patients.

(C) The heatmap showing the different gene expression patterns of the high risk group and low risk group. It showed that the gene expression pattern of high risk group and low risk group were obvious distinct.

Baishideng Publishing Group

7901 Stoneridge Drive, Suite 501,
Pleasanton, CA 94588, USA
**Telephone:** +1-925-223-8242
**Fax:** +1-925-223-8243
**E-mail:** bpgoffice@wjgnet.com
**https**://www.wjgnet.com

(D) The top enriched pathways in the high risk group, as indicated by the gene set enrichment analysis. It showed that the E2F targets, G2M checkpoint, spermatogenesis, mitotic spindle, and Myc targets v1 were significant enriched in the high risk group patients.