Peer Review File

Article information: https://dx.doi.org/10.21037/cco-23-139

Reviewer A

The article "Identification of characteristic genes of cervical cancer via an artificial neural network" suggested that PHYHIP, CRISP3, CRNN, CDKN2A, MCM5, MCM2, C1orf112, HELLS, and KNTC1 were differentially expressed genes for cervical cancer which were identified by the protein-protein interaction network analysis. The topic is very novel by using interesting methods. The manuscript is concise and easy to read. However, the manuscript can be accepted after revision.

Comment #1: In the abstract, please only mention the significant information.

Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. We have taken all these comments and suggestion into account, and simplified the content of the abstract. We've changed [The differentially expressed genes (DEGs) were identified between the normal and CC tissues. The expression levels of the DEGs in the 2 groups were compared] to [The differentially expressed genes (DEGs) were identified and compared between the normal and CC tissues], [A neural network model was established using the characteristic genes of CC. A Cox regression analysis was used to analyze the experimental data and to examine the verification accuracy of the model] to [A neural network model was established using the characteristic genes of CC, while the verification accuracy of the model was examined by Cox regression]. (see Page2, line60-61, 63-64).

Comment #2: Mentioned description in the methodology should be in detail for the reproducibility of the study with proper citation.

Response: Thank you very much for your valuable comment. We gratefully appreciate you for reading our article carefully and giving the above comments. We have described the reproducibility of ANNs in comparison to traditional methods in the background. Meanwhile the accuracy of ANNS was assessed and validated in details. [Using the classical models require many assumptions that may not be true in some real applications. Violation of these assumptions may produce error in prediction and hypothesis testing. The inability to capture pattern complexity and inability to capture process dynamic are two major pitfalls of traditional methods (10).] (see Page4, line99-103) . We described the accuracy detailed. [The vertical and horizontal axes of a ROC curve define the true- and false-positive rates, respectively. The area under the curve (AUC) of the ROC curve was used to evaluate the accuracy of the model. The ROC curve describes the specificity and sensitivity of the classifier. The closer to 1 the AUC the higher quality of the classifier. The ideal classifier AUC is 1, the AUC =0.5 means it's random and useless (10,25).] (see Page6, line175-180) .

Comment #3: Grammatical error "…adjusted p-value < 0.05"

Response: Thank you very much for pointing out this problem in our manuscript. We

have thoroughly checked and corrected the Grammatical mistakes. We've changed [The following filter conditions were set: |log2 fold-change| ≥1 and an adjusted P value < 0.05] to [The following filter conditions were set: |log2 fold-change| ≥1 and an adjusted p-value < 0.05], [The cut-off criterion was an adjusted P value < 0.05] to [The cut-off criterion was an adjusted p-value of <0.05]. (see Page4, line130; see Page5, line136) .

Comment #4: Other related works on cancer must be cited in the manuscript, including https://doi.org/10.3390/ijms23073968, https://doi.org/10.1007/s12672-022-00546-6, https://doi.org/10.2174/1568009623666230214100159, 10.26599/TST.2022.9010035.
Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. We have taken all these comments into account and have cited these recommended articles in the manuscript. These works were very important and valuable in the Bioinformatics fields, which are great hints for our research ideas and further research, meanwhile, which were usable in our study to be understand easily.

Comment #5: Most of the figures are poorly presented in the manuscript. please consider rendering images at least 300 dpi, and increase the font size of labels in figures for better legibility.
Response: Thank you for pointing out this issue. According to the revised content, we have taken all these comments and suggestion into account, and have rearranged the Figs and increase the font size of labels in figures for better legibility. (see Page22,23,25).

Comment #6: Line 314-320: What is the connection between your findings is not clear?
Response: We agree with you and have incorporated this suggestion throughout our paper. As for this part, we really want to present the role of immunity in cervical cancer, immunotherapy is an effective treatment in the state of drug resistance. TME (tumor microenvironment) of CC must be an extremely complex one, including high-risk HPV infection, estrogen milieu, infammatory cell-associated, hypoxia, angiogenesis, the immunological host response and so on. So it's important to study how confirm novel biological markers to predict immunotherapy outcome. At the same time, we have made appropriate adjustments to this department. Thank you again for your advice sincerely. [The complexity of TIICs can affect the biological behavior and immune status of the host, thereby regulating the immunotherapy response (48). Thus, research on novel biological markers to predict immunotherapy outcome is of crucial importance.] (see Page11, line339-342).

Comment #7: The significant digits length problem (Check Table 1,2 and others)
Response: Thank you for pointing out this issue. According to the revised content, we have made appropriate adjustments of tables for better reading. (see Page19-21).

Comment #8: In the Conclusions section, needs more significant information.
Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. The comments are all valuable and very helpful for revising and improving our paper. We have rewrite a complete, comprehensive summary and

induction of the article detailly. [The current work used a comprehensive bioinformatics analysis to identified the biomarkers of CC by the use of ANNs. We identified a total of 10,360 DEGs with 35 upregulated and 72 downregulated genes. Metascape was used to conduct the function and pathway enrichment analyses of the DEGs. GO and KEGG revealing the biological characteristics of the target genes. PPI established a network comprised a subset of proteins that interact physically with other proteins. A neural network model was established using the characteristic genes of CC by the use of ANN, which was used to analyze the experimental data and to examine the verification by Cox regression analysis. The differences in TIICs were compared by CIBERSORT. Epigenetic biomarkers need to be considered within the context of differential diagnostic situations, and this is particularly true for cervical cancer biomarkers (56,57,58).

The relationship between viral immune escape and tumor immune escape, the reasons for further changes in the immune microenvironment in CC, and the specific mechanism of the decline of immune killing are not clear. Therefore, we have analysied the differences in TIICs to further understand the microenvironment of CC, to look for factors that can stabilize the cervical immune microenvironment, regulate the composition ratio of each immune cell, immune escape, may become a new direction of immunotherapy. In conclusion, ANNs was an intelligent method with relative good sensitivity and specificity in the diagnosis of CC and therefore explore biomarkers, functions, pathways, regulatory factors, and immunotherapy.] (see Page11, line371-372; Page12, line373-391) .

Comment #9: Please explain all abbreviations at last part.

Response: Thank you for pointing out this issue. We have taken this comment into account, and explain all abbreviations in the table at last part.

Abbreviations

ANN	An artificial neural network
CC	Cervical Cancer
AI	Artificial intelligence
CIBERSORT	Cell-type Identification By Estimating Relative Subsets Of RNA Transcripts
CDKN2A	Cyclin-dependent kinase inhibitor 2A
C1orf112	Chromosome 1 open reading frame 112
HELLS	Helicase, lymphoid-specific
MCM5	Mini-chromosome maintenance protein 5
MCM2	Mini-chromosome maintenance protein 2
KNTC1	Kinetochore associated 1
CRISP3	Cysteine-rich secretory protein 3
PHYHIP	Phytanoyl-CoA 2-hydroxylase interacting protein
CRNN	Cornulin
DEGs	Differentially Expressed Genes
HPV	Human Papillomavirus

TIICs	Tumor-infiltrating immune cells
TIMER	Tumor immune estimation resource
GEO	Gene Expression Omnibus
ROC	receiver operating characteristic
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
MFs	Molecular functions
CCs	Cellular components
BPs	Biological processes
PPI	Protein–Protein Interaction
STRING	Search Tool to Retrieve Interacting Genes
BioGrid	Biological General Repository for Interaction Datasets
MCC	maximal clique center
AUC	area under the curve
MCODE	Molecular Complex Detection

Reviewer B

1) First, the title needs to indicate the research design, i.e., a bioinformatics analysis.
Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. We have changed the title as your recommends. We've changed [Identification of characteristic genes of cervical cancer via an artificial neural network] to [Bioinformatics identification of characteristic genes of cervical cancer via an artificial neural network]. (see Page1, line3-4) .

2) Second, the abstract is inadequate. The background did not describe the potential strength of ANNs in comparison to traditional methods. The methods need to describe how the accuracy was assessed and validated. The results need corresponding findings on the accuracy indicators such as AUC and sensitivity values. The conclusion needs comments for the clinical implications of the findings.
Response: We gratefully appreciate you for reading our article carefully and giving these positive comments which are valuable and very helpful for revising and improving our paper. We have rewritten these contents as you recommends.
Firstly, Thank you for pointing out this issue. We have taken all these comments and suggestion into account, and decieded to simplified the content of the abstract for the concise and to supplement the content of background. We've changed [The differentially expressed genes (DEGs) were identified between the normal and CC tissues. The expression levels of the DEGs in the 2 groups were compared] to [The differentially expressed genes (DEGs) were identified and compared between the normal and CC tissues], [A neural network model was established using the

characteristic genes of CC. A Cox regression analysis was used to analyze the experimental data and to examine the verification accuracy of the model] to [A neural network model was established using the characteristic genes of CC, while the verification accuracy of the model was examined by Cox regression]. (see Page2, line60-61, 63-64).

Secondly, we have cited related articles to explain the potential strength of ANNs in comparison to traditional methods. [Using the classical models require many assumptions that may not be true in some real applications. Violation of these assumptions may produce error in prediction and hypothesis testing. The inability to capture pattern complexity and inability to capture process dynamic are two major pitfalls of traditional methods (10).] (see Page4, line99-103).

Thirdly, we described the accuracy detailed. [The vertical and horizontal axes of a ROC curve define the true- and false-positive rates, respectively. The area under the curve (AUC) of the ROC curve was used to evaluate the accuracy of the model. The ROC curve describes the specificity and sensitivity of the classifier. The closer to 1 the AUC the higher quality of the classifier. The ideal classifier AUC is 1, the AUC =0.5 means it's random and useless (10,25).] (see Page6, line175-180).

Finally, we have commented on the clinical significance of the study findings [The current work used a comprehensive bioinformatics analysis to identified the biomarkers of CC by the use of ANNs. We identified a total of 10,360 DEGs with 35 upregulated and 72 downregulated genes. Metascape was used to conduct the function and pathway enrichment analyses of the DEGs. GO and KEGG revealing the biological characteristics of the target genes. PPI established a network comprised a subset of proteins that interact physically with other proteins. A neural network model was established using the characteristic genes of CC by the use of ANN, which was used to analyze the experimental data and to examine the verification by Cox regression analysis. The differences in TIICs were compared by CIBERSORT. Epigenetic biomarkers need to be considered within the context of differential diagnostic situations, and this is particularly true for cervical cancer biomarkers (56,57,58).

The relationship between viral immune escape and tumor immune escape, the reasons for further changes in the immune microenvironment in CC, and the specific mechanism of the decline of immune killing are not clear. Therefore, we have analysied the differences in TIICs to further understand the microenvironment of CC, to look for factors that can stabilize the cervical immune microenvironment, regulate the composition ratio of each immune cell, immune escape, may become a new direction of immunotherapy. In conclusion, ANNs was an intelligent method with relative good sensitivity and specificity in the diagnosis of CC and therefore explore biomarkers, functions, pathways, regulatory factors, and immunotherapy.] (see Page11, line371-372; Page12, line373-391).

3) Third, in the introduction of the main text, the authors need to analyze why ANNs could address the limitations of traditional methods and have comments on the limitations of

traditional methods. Please have comments on the potential clinical significance of this study, not summarize the methodology of this research in the last paragraph of this part.

Response: We agree with you and have incorporated this suggestion throughout our paper. Thank you again for your positive comments on our article which are valuable and very helpful for revising and improving our paper.

Firstly, we have analyzed why ANNs could address the limitations of traditional methods and have comments on the limitations of traditional methods. [In fact, ANNs do not require the assumption of data normality and can determine a functional relationship in which the relationship between the independent and dependent variables is not necessarily linear (8). Moreover, since ANN has no limitation regarding its formulated function, it is more flexible and has more strength in mimicking complicated patterns than logistic regression. Another advantage of ANNs is that their ability to find patterns despite of missing data (9). Thus, a correct answer may still be obtained if certain cells are removed or exhibit a false function within the network (10). However, using the classical models require many assumptions that may not be true in some real applications. Violation of these assumptions may produce error in prediction and hypothesis testing. The inability to capture pattern complexity and inability to capture process dynamic are two major pitfalls of traditional methods (9).]. [However, using the classical models require many assumptions that may not be true in some real applications. Violation of these assumptions may produce error in prediction and hypothesis testing. The inability to capture pattern complexity and inability to capture process dynamic are two major pitfalls of traditional methods (9).] (see Page3, line94-100;Page1, line3-4) .

Secondly, we revised the conclusion by your suggestion. [The current work used a comprehensive bioinformatics analysis to identified the biomarkers of CC by the use of ANNs. We identified a total of 10,360 DEGs with 35 upregulated and 72 downregulated genes. Metascape was used to conduct the function and pathway enrichment analyses of the DEGs. GO and KEGG revealing the biological characteristics of the target genes. PPI established a network comprised a subset of proteins that interact physically with other proteins. A neural network model was established using the characteristic genes of CC by the use of ANN, which was used to analyze the experimental data and to examine the verification by Cox regression analysis. The differences in TIICs were compared by CIBERSORT. Epigenetic biomarkers need to be considered within the context of differential diagnostic situations, and this is particularly true for cervical cancer biomarkers (56,57,58).
The relationship between viral immune escape and tumor immune escape, the reasons for further changes in the immune microenvironment in CC, and the specific mechanism of the decline of immune killing are not clear. Therefore, we have analysied the differences in TIICs to further understand the microenvironment of CC, to look for factors that can stabilize the cervical immune microenvironment, regulate the composition ratio of each immune cell, immune escape, may become a new direction of immunotherapy. In conclusion, ANNs was an intelligent method with relative good sensitivity and specificity in the diagnosis of CC and therefore explore biomarkers, functions, pathways, regulatory factors, and immunotherapy.] (see Page11, line371-372; Page12, line373-391) .

4) Fourth, in the methodology, please briefly describe the datasets used, the calculation of AUC and its threshold values for an accurate classification tool, and details for the test of external validity in validation datasets.

Response: Thank you for pointing out this issue. According to the revised content, we have taken all these comments and suggestion into account and described the details of validation datasets. [The vertical and horizontal axes of a ROC curve define the true- and false-positive rates, respectively. The area under the curve (AUC) of the ROC curve was used to evaluate the accuracy of the model. The ROC curve describes the specificity and sensitivity of the classifier. The closer to 1 the AUC the higher quality of the classifier. The ideal classifier AUC is 1, the AUC =0.5 means it's random and useless (10,25).] (see Page6, line175-180).

5) Finally, please consider to cite several related papers: 1. Yu R, Zhang L, Yu Q, Zhao H, Yang H, Wang Y. Effect of LHX2 gene methylation level and its function on radiotherapy of cervical cancer. Transl Cancer Res 2021;10(6):2944-2961. doi: 10.21037/tcr-21-739. 2. Lorenc A, Romaszko-Wojtowicz A, Jaśkiewicz Ł, Doboszyńska A, Buciński A. Exploring the efficacy of artificial neural networks in predicting lung cancer recurrence: a retrospective study based on patient records. Transl Lung Cancer Res 2023;12(10):2083-2097. doi: 10.21037/tlcr-23-350.

Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. We have taken all these comments into account and have cited these recommended articles. These articles provide similarities and differences with our research, meanwhile, which are great hints for our research ideas and further research, meanwhile, which were usable in our study to be understand easily.

Reviewer C

The paper titled "Identification of characteristic genes of cervical cancer via an artificial neural network" is interesting. ANN is a robust neural network model that can be used to potentially predict CC based on the gene score. It can provide novel insights into the pathogenesis and molecular mechanisms of CC. However, there are several minor issues that if addressed would significantly improve the manuscript.

1) Compared with other models, what are the advantages of the model in this study? It is recommended to add relevant content.

Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. We have taken all these comments and suggestion into account.

We have explained the potential strength of ANNs in comparison to traditional methods. [Using the classical models require many assumptions that may not be true in some real applications. Violation of these assumptions may produce error in prediction and hypothesis testing. The inability to capture pattern complexity and inability to capture process dynamic are two major pitfalls of traditional methods (10).] (see Page4, line99-103). Secondly, we described the accuracy detailed. [The vertical and horizontal axes of a ROC curve define the true- and false-positive rates, respectively. The area under the curve (AUC) of the ROC curve was used to evaluate the accuracy of the model. The ROC curve describes the specificity and sensitivity of the classifier. The closer to 1 the AUC the higher

quality of the classifier. The ideal classifier AUC is 1, the AUC =0.5 means it's random and useless (10,25).] (see Page6, line175-180).

2) Some fonts need to be enlarged, as shown in Figures

Response: Thank you for pointing out this issue. According to the revised content, we have increased the font size of labels in figures for better legibility.(see Page22,23,25) .

3) How to identify and verify the prognostic characteristics for predicting disease-free survival and overall survival of patients with cervical cancer by integrating multiple data sets? It is recommended to add relevant content.

Response: We agree with you and have incorporated this suggestion throughout our paper. Thank you again for your positive comments on our article which are valuable and very helpful for revising and improving our paper. We can get these messages by integrating the GEO and Timer database, however, which is not the focus of this study. We can put forward the contents as a important indicators in the following research. We also plan to carry out relevant clinical experiments to verify our findings.

4) All figures are not clear enough. It is recommended to provide clearer figure again.

Response: Thank you for pointing out this issue. We have rearranged the Figs and increase the font size of labels in figures for better legibility.(see Page22,23,25) .

5) The functional research on the main target genes should be increased, which may be more meaningful.

Response: We gratefully appreciate you for reading our article carefully and giving this significant comments. We indeed have planned to more functional research of the main target genes in different depths and in many ways, which will be the focus and purpose of our following research, and will be known in the near future. Thank you again for your positive comments.

6) The introduction part of this paper is not comprehensive enough, and the similar papers have not been cited, such as "Diagnostic test of bioimpedance-based neural network algorithm in early cervical cancer, PMID: 35571399". It is recommended to quote the articles.

Response: Thank you for pointing out this issue. According to the revised content, we have rewritten the introduction part and cited this recommended article as your suggestion. The recommended article provides fully explain of the ANNS, which make our study to be understand easily. (see Page1, line3-4).

7) How to provide clinical guidance for prognosis intervention of cervical cancer based on the results of this study? It is recommended to add relevant contents.

Response: We gratefully appreciate you for reading our article carefully and giving these positive comments. We have added the relevant contents. [The relationship between viral immune escape and tumor immune escape, the reasons for further changes in the immune microenvironment in CC, and the specific mechanism of the decline of immune killing are not clear. Therefore, we have analysied the differences

in TIICs to further understand the microenvironment of CC, to look for factors that can stabilize the cervical immune microenvironment, regulate the composition ratio of each immune cell, immune escape, may become a new direction of immunotherapy. In conclusion, ANNs was an intelligent method with relative good sensitivity and specificity in the diagnosis of CC and therefore explore biomarkers, functions, pathways, regulatory factors, and immunotherapy.] (see Page12, line383-391) .

8) It is recommended to add in vivo and in vitro experimental validation of the results of this study.

Response: We gratefully appreciate you for reading our article carefully and giving this significant comment. We indeed planned to carry out in vivo and in vitro experimental validation of this study in the following research, which included the functional research of the main target genes in different depths and in many ways, which will be the focus and purpose of our next research, and will be known in the near future. Thank you again for your positive comments.