Baishideng Publishing Group

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

PEER-REVIEW REPORT

Name of journal: *World Journal of Clinical Oncology*

Manuscript NO: 87022

Title: Hub genes and their key effects on prognosis of Burkitt lymphoma

Provenance and peer review: Unsolicited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 05776245
Position: Peer Reviewer
Academic degree: BSc, MSc
Professional title: Academic Research, Research Scientist, Senior Scientist, Teaching Assistant
Reviewer's Country/Territory: Poland
Author's Country/Territory: China

Manuscript submission date: 2023-07-19
Reviewer chosen by: Geng-Long Liu
Reviewer accepted review: 2023-08-07 10:08
Reviewer performed review: 2023-08-10 17:58
Review time: 3 Days and 7 Hours

Scientific quality	[ ] Grade A: Excellent [ ] Grade B: Very good [ ] Grade C: Good [Y] Grade D: Fair [ ] Grade E: Do not publish
Novelty of this manuscript	[ ] Grade A: Excellent [Y] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No novelty
Creativity or innovation of this manuscript	[ ] Grade A: Excellent [ ] Grade B: Good [Y] Grade C: Fair [ ] Grade D: No creativity or innovation

Baishideng Publishing Group

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

Scientific significance of the conclusion in this manuscript	[ ] Grade A: Excellent [ ] Grade B: Good [Y] Grade C: Fair [ ] Grade D: No scientific significance
Language quality	[ ] Grade A: Priority publishing [Y] Grade B: Minor language polishing [ ] Grade C: A great deal of language polishing [ ] Grade D: Rejection
Conclusion	[ ] Accept (High priority) [ ] Accept (General priority) [ ] Minor revision [Y] Major revision [ ] Rejection
Re-review	[Y] Yes [ ] No
Peer-reviewer statements	Peer-Review: [Y] Anonymous [ ] Onymous
	Conflicts-of-Interest: [ ] Yes [Y] No

SPECIFIC COMMENTS TO AUTHORS

Dear Authors, your study on prognostic hub genes in Burkitt's lymphoma has its own merits, but major revisions need to be addressed before proceeding forward. Therefore, please answer or consider the following: (1) Abstract: most probably "CNS" abbreviation might need an explanation. It is used only two times throughout the paper, so alternatively you can put the full description. (2) Abstract: In "Chips and sequencing information", by "Chips" you meant microarray data? If yes, please change it. Same is on page 4 in the section "Background". (3) Abstract, Aims: you can add to the first sentence that you wanted to find hub genes in perform gene ontology specifically in BL. In the second sentence, most likely the "carry out" would be "carried out" and "construct" would be "constructed". Some parts of the manuscript might need to be reviewed by a native speaker; please double-check the language quality in your paper. (4) Abstract, Methods: The second sentence starting from a full name of WGCNA is not capitalized. (5) Abstract, and the entire paper: Did you mean "cytoHubba" instead of "cytoHub" by any chance? (6) Why hub genes were identified using a tool that is separated from WGCNA toolkit? WGCNA does provide Module Membership (MM)

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

values that can help in the investigation of hub genes, and combining it with the Gene Significance (GS) values can result in the identification of a so-called "driver genes". What was the reason of not using WGCNA in this step? Moreover, did you verified cytoHub/cytoHubba results using MM values in WGCNA, or were they not used at all in the workflow? (7) One of your aims was to perform a survival analysis, then why you only focused on overall survival and no other endpoints such as disease-free survival etc.? Events caused by disease recurrence occur earlier than death from the disease, so it may be beneficial to include such an endpoint instead of overall survival. Preferably, one could include more than one survival endpoint, especially if intending to perform a survival analysis. (8) Remember to italicize all gene symbols throughout the paper. (9) Abstract, results: Consider changing "Moreover, we found 2 hub genes associated with OS" to "Within these hubs, two genes were associated with OS", next in the part "we combined the two hub genes", "the" could be "these". Lastly, rewrite the sentence starting with "And" and try to avoid repetition of "we". (10) Abstract, results: Once I completed reading the paper, I am unsure if you "found several potential therapeutic targets for BL with poor prognosis". Therapeutic targetability of these findings were not investigated in your study, or at least not in a proper way. The best bet would be to focus on prognostic significance of investigated age-related biomarkers. Consequently, I would avoid "Therapeutic target" keyword. (11) Abstract, conclusion: add "that" before "might" (12) Throughout the text, some space marks are missing. Please double check the entire manuscript and checked whether words are separated from brackets, citations, etc. (13) Background: The part "while chronic EBV (Epstein-Barr virus, EBV) infection plays an important role in BL" could be "with chronic Epstein-Barr virus (EBV) infection playing an important role in BL". Moreover, you can specify in which clinical type of disease. (14) Background: The part "MYC regulates the expression of target genes which regulate a variety of cellular processes"

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

could be "MYC orchestrates the expression of target genes, regulating a variety of cellular processes". (15) Background: In the part "and other forms of high-throughput functional genomic data, which submitted by research communities", change "which submitted" to "that are submitted". (16) Background: whether the short synopsis about GEO and WGCNA is necessary in this section is a matter of debate. In my opinion you can shorten these descriptions and move what is left to the methodological section. (17) Background: In the part "was carried out to identify a mRNA signature which significant associated with prognosis. Finally, a prognostic nomogram was established based on the combination", you can change "which significant associated" to "that was significantly associated with" or "presenting a significant association with". If you would like to avoid two -ed in "established based", you can change it to the "established on the basis of". (18) Materials and Methods: you can consider deleting "s" and the end of "raw gene expressions". I think "expression" would be better. Double-check the entire paper. (19) What was the reason of selecting these two (GSE4475 and GSE69051) GEO datasets? Once I completed reading the paper, it turned out that GSE69051 was only used after GSE4475 because GSE4475 was too small to perform survival analysis. Alternatively, the entire workflow could have been performed only on GSE69051 since there was no verification of results from one dataset in the other. Moreover, if you aimed to perform survival analysis, why not searching for a dataset that included more than one survival endpoint? Were there some problems in finding such datasets? (20) What kind of tool was used to perform this step: "The mRNA sequencing data annotation information was used to match the probe with the corresponding gene to transform the gene name into gene symbol". I presume it was not done manually. Have you tried g:Profiler or SYNGO, or something similar? (21) Section 2.2: The part "The top 5,000 variant of expression profiles were used" could be "The top 5,000 most variable genes were used", assuming I am correct in understanding of your workflow. (22) Section 2.2:

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

In the descriptions of WGCNA, you did not specify if you selected unsigned, signed, or signed hybrid approach. (23) In methods, instead of adding links to websites, consider adding citations for each tool to help authors gain attention. If no preferred citation is available, URL links are okay. (24) Section 2.2: "Samples cluster analysis was performed using the hclust tool" means that you did not use a built-in clustering options from WGCNA? (25) Section 2.2: Beta-power used in WGCNA is mentioned in Results (it was 12 if I understood correctly), but please mention it also in the methodology. (26) Should section 2.3 be a part of WGCNA toolkit? Especially if you refer to MM and GS values at the end of this section. Moreover, the usefulness of GS/MM values in your study is not evident. Currently it appears that you omitted GS because you did not investigate driver genes but instead focused on hub genes, and secondly MM values from WGCNA were also not used because hub genes were indicated by a separate tool (cytoHub/cytoHubba) and not using values provided within WGCNA toolkit. Moreover, in the same section – what is "thermal mapping kit"? Did you mean module-trait relationship? If yes, then as I said at the beginning – all these descriptions should be a a part of WGCNA, and thus you could merge them with section 2.2. (27) Section 2.5: Most probably the GeneMania includes both PPI and GI data. Have you filtered out PPI data from the server so to obtain GI only? In the same section, the last sentence might be misleading ("The statistical significance was expressed as a collective score of >0.15"). I would change "statistical significance" to something else, or just write that the threshold of collective score of 0.15 was applied on the server, because in the next section you refer to the "real" statistical significance that was a typical $p<0.05$, which could introduce uncertainty among Readers. (28) Section 2.6: In the part "According to the 50th percentile cut-off value of each hub gene mRNA, patients were divided into the high-expression and low-expression groups" – that means that all hub genes were in fact protein-encoding? Were non-coding data included in the GEO datasets? Moreover, applying median cut-off

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

is not always a proper way from the biological and clinical point of view. Have you tried applying a cutpoint using relevant tools? Some genes in Figure 8 could be statistically significant (like in subfigure B, C, H) if other cut-off would be applied. (29) Section 2.6: The last sentence states that "Additionally, $P < 0.05$ was statistically significant unless otherwise indicated". I did not see any other part that indicates statistical significance and threshold, so maybe the part "unless otherwise indicated" is unnecessary? (30) Section 2.7: The name of the database "DSigDB" is not in line with the website to which a link is provided next to the name (URL link refers to DGIdb tool). Please double-check and correct whichever is wrong. Moreover, from what I understood, this part cannot be technically described with more details, because once the webtool is accessed, only gene symbol is provided and all results are automatically provided in the new tab? There are not filtering, thresholding, etc.? (31) Section 3.1: Change "In this study, we obtained the BL dataset in GSE4475, A total of 13, 514 gene expression values were derived from the raw file" to "In this study, we obtained the BL dataset from the GSE4475, resulting in a total of 13,514 gene expression values". (32) Section 3.1: In the part "Then, we selected a total of 5, 000 genes with the greatest average expression values for cluster analysis", should it be about most variable genes, not the ones with the greatest average expression? Focusing on greatest expression would be inappropriate since you can have a biologically meaningful change in the expression that is relatively small compared to others that are not so crucial. (33) Section 3.1: I would move this sentence to the figure's description "Red indicated more gene expression, white less, and gray indicated deletion (Fig. 1)" and moved reference to Fig1 to previous sentence in the text. Moreover, you should use other words than "indicated deletion" for gray in this context, it would be better to say that it represents an unknown status for some samples. This is equivalent to "unknown" you used in Table 1. (34) Figure S1 and S2 are unavailable for me to assess – I cannot identify them in the system or in the manuscript file. (35) Section 3.1: Please

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

standardize the use of the word "gray" or "grey". (36) Section 3.1: In the part "Genes in gray were not included in any module, then we analyzed", the word "thus" would fit better than "then". Right after this sentence, I would delete "After docking with clinical character data". (37) Section 3.1: Rationale for selecting the age as a clinical trait of interest is mentioned once in Discussion, but its relevance is not mentioned in Results. Moreover, why focusing only on age when there were also other traits that were found significantly correlated with gene modules, as shown in Figure 3? (38) Figure 1 could be better described (see my comment no. 33), with more details. Moreover, CCS and Ki67 are not mentioned in Table 1, while they constitute clinical data similar to age, sex, and stage. (39) Figure 2: The last sentence of figure description ("The yellow brightness of the middle part represented the strength of connections between modules") could mention about darker shades of yellow turning into brown or orange. (40) Figures in general would benefit much from increasing font size. (41) Figure 4 represents scatter plot with both GS and MM. If applying 0.6 threshold, a few driver genes would be identified but this threshold is not ideal. Have you at least tried to identify if genes most correlated to both trait and module were indicated as hub genes using cytoHub/cytoHubba? (42) It might be hard to increase font in Figure 5; thus, please consider moving it to the supplementary materials. In the same figure, if top hubs the ones that are in the center of the network? If yes, then I see 12 nodes, whereas you mentioned about 10 top hubs. How can we identify top hubs in this figure before you focused on them in further steps? (43) Figure 6: what is "with a common goal" or "that implemented common goals"? Moreover, in the same figure the slash is barely visible, please change it something more evident or increase the quality of the figure. The figure would generally benefit from small legends for each subfigure. For example, in subfigure A you can show that edges represent co-expression. Figure's description should be updated afterwards. (44) Section 3.3: the part "were shown in Fig. 7A" or the

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

equivalent for 7B could be put in brackets and moved earlier in sentences, next to the "top 10 GO/KEGG terms", so that the remaining part of the sentences would be only the name of terms. (45) Table 2 might need to be moved to the supplementary materials due to its size. Consequently, Table 3 might be moved too, even though its size it not that large. However, top 10 terms from GO or KEGG are visible in Figure 7, so it is not a big deal to make both tables as supplements. Another question, how "top" terms established? Based on p-value, number of annotated genes ("count") or what? While on the topic, you can also explain the meaning of "count" column. (46) Section 3.6: Enrichr website was first-time mentioned in this section; it is not present in methodology. Are you sure that association with drugs for IL2RA and CXCL10 was investigated using Enrichr built-in tool? Because methodology stated a specific URL link that is outside Enrichr. Please justify and correct. For the same section, description is rather weak, so is the part of Discussion related to it. You can mention that you focused only on IL2RA and CXCL10 because these were the only significant results from survival analysis. I also found that Discussion is lacking details on current drugs that are used in BL and maybe appeared in the results of this analysis. Some drugs common for IL2RA and CXCL10 might be further discussed. In general, current description is too short. Druggability of IL2RA and CXCL10 should be discussed thoroughly if you would like to leave some prospects for the future about therapeutic potential etc. Mentioning, e.g., that "This makes CXCL10 a 'key driver chemokine' and a valid target for therapy" entails a proper justification. (47) Discussion: In the part "between c-MYC and the gene for either the kappa or lambda light chain", would "of" be better than "for"? (48) Discussion: Is apoptosis really that high in BL, similar to proliferation? The part "The proliferation rate and apoptosis rate of BL tumor cells are extremely high" suggests so. (49) Discussion: "Ten hub genes (SRC, TLR4, CD40, STAT3, SELL, CXCL10, IL2RA, IL10RA, CCR7 and FCGR2B) and several pathways were identified by WGCNA". What pathways were identified using WGCNA?

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

(50) Discussion: In the part “And then , we used nomogram to find a new risk assessment system”, the part “And then” could be “Afterwards”. Later on, “What’s more” should be “What is more”. (51) Conclusion: mentioning about “driving genes” only in this part is not appropriate way of referring to “hub genes”. GS values were practically omitted in your WGCNA approach and MM values were probably not used because external tool was used for hubs identification. The part “that might be new therapeutic targets” could not conclude your findings. I would rather extensively enrich the paper with therapeutic methodology workflow and discuss it properly, or alternatively provide statements more related to prognostic significance of identified biomarkers. In the last sentence of Conclusion, “A” before “nomogram” must be lowercased. (52) Another study limitation is the use of only public datasets (you can mention it in the relevant part of Discussion). However, once properly presented and discussed, it is acceptable.

Baishideng Publishing Group

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

PEER-REVIEW REPORT

Name of journal: *World Journal of Clinical Oncology*

Manuscript NO: 87022

Title: Hub genes and their key effects on prognosis of Burkitt lymphoma

Provenance and peer review: Unsolicited Manuscript; Externally peer reviewed

Peer-review model: Single blind

Reviewer's code: 06125275
Position: Peer Reviewer
Academic degree: MD

Professional title: Doctor, Research Assistant Professor

Reviewer's Country/Territory: Algeria

Author's Country/Territory: China

Manuscript submission date: 2023-07-19

Reviewer chosen by: Geng-Long Liu

Reviewer accepted review: 2023-08-14 07:42

Reviewer performed review: 2023-08-15 10:06

Review time: 1 Day and 2 Hours

Scientific quality	[Y] Grade A: Excellent [ ] Grade B: Very good [ ] Grade C: Good [ ] Grade D: Fair [ ] Grade E: Do not publish
Novelty of this manuscript	[Y] Grade A: Excellent [ ] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No novelty
Creativity or innovation of this manuscript	[Y] Grade A: Excellent [ ] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No creativity or innovation

Baishideng Publishing Group

7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA
Telephone: +1-925-399-1568
E-mail: bpgoffice@wjgnet.com
https://www.wjgnet.com

Scientific significance of the conclusion in this manuscript	[Y] Grade A: Excellent [ ] Grade B: Good [ ] Grade C: Fair [ ] Grade D: No scientific significance
Language quality	[Y] Grade A: Priority publishing [ ] Grade B: Minor language polishing [ ] Grade C: A great deal of language polishing [ ] Grade D: Rejection
Conclusion	[ ] Accept (High priority) [Y] Accept (General priority) [ ] Minor revision [ ] Major revision [ ] Rejection
Re-review	[Y] Yes [ ] No
Peer-reviewer statements	Peer-Review: [Y] Anonymous [ ] Onymous
	Conflicts-of-Interest: [ ] Yes [Y] No

SPECIFIC COMMENTS TO AUTHORS

In the background section of your abstract, CNS? (explicate this abbreviation). In the manuscript background after your 3rd reference, please cite the rate of this incidence in development countries. Why the LDH rate which is a relevant biological marker have not been integrated in your nomogram? GS and MM do not appear in the final list of abbreviations. Your nomogram have not been discussed, despite your cites in you bibliography a reference about a nomogram used in BL, Lu, J., et al., Status and prognostic nomogram of patients with Burkitt lymphoma. Oncol Lett, 2020. 19(1): p. 972-984.)