TP TRANSLATIONAL PEDIATRICS
AN OPEN ACCESS JOURNAL COVERING ALL ASPECTS OF PEDIATRICS RESEARCH
IMPACT FACTOR 2.286

Peer Review File

Article information: https://dx.doi.org/10.21037/tp-21-193

Reviewer A

Comment 1. Add more on the limitations of this research paper: as example the limited number of patients enrolled.

Reply 1: Thanks for your kindly suggestions! According to your helpful suggestions, we have added the limitations of this study in the revised manuscript.

Changes in the text: We have modified our text as advised (see page 13, line 265-273).

Comment 2. Some parts of results section seem discussion.

Reply 2: Thanks for your suggestions. We apologized for inappropriate forms. According to your helpful suggestions, we have merged the results with the discussion section in the revised manuscript.

Changes in the text: We have modified our text as advised (see page 8-13, line 157-273).

Comment 3. It is not very clear if the difference in proteins seen regards a decrease in concentration in autistic children.

Reply 3: Thanks for your great question. Please allow us to explain.

In this research, mass spectrometry- based relative quantitation proteomics was adopted. Specific, 1μg peptide of each urine sample was performed mass spectrometry acquisition and analysis. The peptide intensity was calculated by summing the peak areas of their respective fragment ions for MS2, and then the protein intensity was calculated by summing the intensity of their respective peptides. The differential proteins were screened relied on the relative quantitative information for each protein between the two groups. Therefore, the differential proteins have no direct relationship with the concentration of urine.

Changes in the text: We have modified our text as advised (see material and methods).

Reviewer B

Summary

Autism spectrum disorder (ASD) is complex developmental disability, typically appearing during childhood and affecting a person's ability to communicate and interact with others. This study aims to identify differential proteins in the urinary

proteome of autistic and non-autistic children (between 3 to 7 years) utilizing a quantitative strategy called data-independent acquisition (DIA). Specifically urinary proteome has been employed to detect proteins that can be used to diagnose ASD given that urine is an ideal source of biomarkers because it is abundant, easily acquired and enriched with proteins and metabolites. A data set of 24 children (18 with autism 6 non-autistic) children were analyzed using t-test, resampling and classification capacities for proteins. The study identified 118 differentiated proteins and 13 proteins with successful clustering performance based on the statistical analysis.

Reply: Thanks for your great and helpful comments.

Major Issues:

Comment 1: Response to "To our knowledge, no previous studies have been reported on urinary proteome for autism." I do not think this is true, there are several studies some of which are listed below in the references.

Reply 1: Thanks for your kindly suggestions! We apologize for the error, and we omitted these previous urine proteomic studies your mentioned above. According to your helpful suggestions, we deleted this sentence in the revised manuscript.

Changes in the text: We have modified our text as advised (see page 3, line 60-62).

Comment 2: Given your sample size being small and possibly having non-normal distributed variables, I suggest using a non-parametric test instead of t- test.

Reply 2: Thanks for your kindly suggestions. Please allow us to explain.

The purpose of this study was to discovery differential proteins between the two groups of urine samples from autism children and healthy control. We screened 118 differential proteins between the two groups. To confirm whether these differences can be generated randomly, we conducted randomization grouping statistical analysis, which is a strict statistical strategy for clinical omics research (details of randomization grouping analysis are presented on reply 5). The results of randomization grouping analysis showed that the average number of differential proteins in all combinations was 10, which means that only 10 proteins may be randomly generated, further indicating that 91.5% of differential proteins were credible. Moreover, some of these differential proteins have been reported to be related to autism in previous studies, and some important pathways enriched by these differential proteins were related to autism. In summary, our results show that the current sample size in this study can demonstrate that urine can show significant difference between autistic and non-autistic children.

Therefore, it can be proved that there is clear difference in urine between the two groups, the type of test we use is not very critical in this stage. For the understanding and recognition of peers in the field, the widely used t-test was used to screen the differences. Considering the time and economic cost of mass spectrometry, we apologized that the limited size of samples was used in this study, but this pilot study will provide a basis for future research on large sample sizes.

Changes in the text: We have modified our text as advised (see statistical analysis).

Comment 3: Could authors explain further what they mean by "*fold change in increasing group ≥1.5 and in decreasing group ≤0.67*," and how these thresholds are determined?

Reply 3: Thanks for your questions. Please allow us to explain.

In this research, we screened differential proteins between the autistic group and the non-autistic group based on quantitative information of proteins. We used traditional proteomics strategies to screen differential proteins, including fold change (FC) and *P* value. Specifically, we first calculated the average quantitative value of each protein in the autistic and non-autistic group respectively, and the fold change is obtained by dividing two values. Therefore, "fold change in increasing group ≥ 1.5 and in decreasing group ≤ 0.67," mean average quantitative value of proteins in the autistic group divide by non-autistic group. If the fold change ≥ 1.5 of the protein means that this protein upregulated in the autistic group, and if the fold change ≤ 0.67 of the protein means that this protein downregulated in the autistic group. Finally, differential proteins were screened with the following criteria: fold change in increasing group ≥ 1.5 and in decreasing group ≤ 0.67, $P < 0.01$.

The threshold value of 1.5 for FC is generally adopted for criteria in most clinical omics studies, since this value could ensure that the difference between biological duplications is greater than technical duplications, so the difference between two groups was significant enough. Thus, the threshold value of 1.5 for FC is widely recognized in the field. In this study, we adopted the threshold value of 1.5 for FC.

Changes in the text: We have modified our text as advised (see statistical analysis).

Comment 4: Which classification method is used to obtain the ROC curves obtained by Metaboanalyst software?

Reply 4: Thanks for your questions. Please allow us to explain.

MetaboAnalyst (http://www.metaboanalyst.ca) is a comprehensive Web application for omics data analysis and interpretation, including ROC curve analysis. Binary classification method was used to perform ROC curve analysis in MetaboAnalyst

software (*Xia et al. 2013*). Briefly, the assessment of biomarker performance is to consider the frequency with which the test produces: true positives, true negatives, false positives and false negatives. One then summarizes these values into the proportion of actual positives that are correctly classified as positive (sensitivity) and the proportion of actual negatives that are correctly classified as negative (specificity).

References

Xia, J., D. I. Broadhurst, M. Wilson and D. S. Wishart (2013). "Translational biomarker discovery in clinical metabolomics: an introductory tutorial." Metabolomics 9(2): 280-299.

Changes in the text: We have modified our text as advised (see statistical analysis).

Comment 5. Could authors explain details of the randomized grouping statistical analysis strategy used? Further I do not understand the statement "the average number of differential proteins in all random combinations was 10, which indicated that reliability of the differential proteins was 91.5%."

Reply 5: Thanks for your great questions. Please allow us to explain.

This is a very strict analysis strategy for omics clinical research.

When using omics technology to study disease biomarkers, the differences are usually screened between disease group and control group. Because omics data is huge but sample size is limited, the differences between two groups may be randomly generated. To this end, we have proposed a randomized grouping statistical analysis strategy, which is suitable for the study of clinical omics disease biomarkers with limited sample size, and to determine whether the differences between two groups are randomly generated (Meng W, Gao Y. Randomized grouping statistical analysis in clinical omics biomarker discovery. *MOJ Proteomics Bioinformatics.* 2020; 9(3): 73-75. DOI:10.1 5406/MOJPB.2020.09.00283).

Specifically, screen the differences between the disease group and the control group, divide all samples into two groups randomly, screen the differences in each random combination, and calculate the average number of differences in all combinations. We compare it with the number of differences in the normal group to determine whether these are randomly generated. The workflow of randomized grouping statistical analysis is presented in Figure 1. Based on this strategy, some proteomic studies have performed randomized grouping statistical analysis (*Huan, Wei et al. 2021, Zhang, Gao et al. 2021*).

Therefore, in this study, 118 differential proteins were screened between the autistic and non-autistic groups. To confirm whether these differential proteins were randomly generated, we performed randomized grouping statistical analysis. Twenty-four samples from the autism (n=18) and control groups (n=6) were randomly divided

into two groups and the same criteria were used to screen the differential proteins. In a total of 134,596(C_{24}^{6}) random combinations, the average number of differential proteins in all combinations was 10. In other words, only 10 differential proteins might be generated randomly, which further demonstrated that 91.5% of 118 differential proteins were reliable.

The details of results and method of randomized grouping statistical analysis strategy were available in this link: https://pan.baidu.com/s/19ZSzJ2-I_jCp9wxtxQzheg Password: n9yh. If we have not answered satisfactorily, please let us know. We will try to explain in more detail.

Changes in the text: We have modified our text as advised (see page 13, lines 178-183).

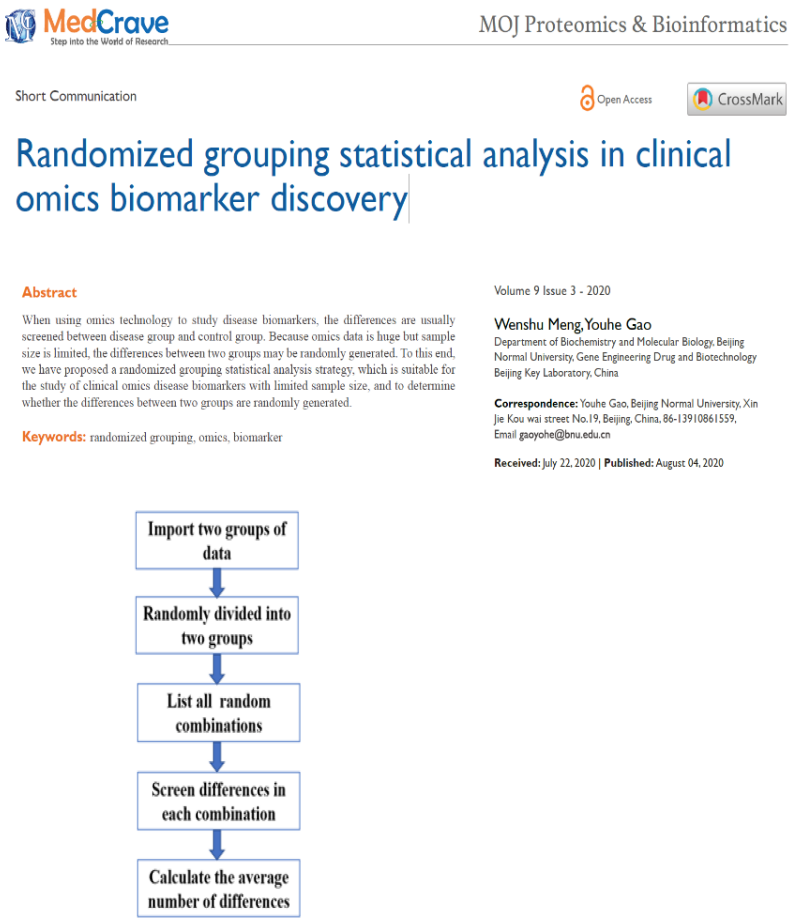

Figure 1 The workflow of randomized grouping statistical analysis program.

References

Meng W, Gao Y. Randomized grouping statistical analysis in clinical omics biomarker discovery. MOJ Proteomics Bioinformatics. 2020; 9(3): 73-75. DOI:10.15406/MOJPB.2020.09.00283

Zhang, Y., Y. Gao, J. Wei and Y. Gao (2021). Dynamic Changes of Urine Proteome in Rat Models Inoculated with Two Different Hepatoma Cell Lines. J Oncol 2021: 8895330.
Huan, Y., J. Wei, T. Su and Y. Gao (2021). Urine proteome changes in a chronic unpredictable mild stress (CUMS) mouse model of major depressive disorder. J Pharm Biomed Anal 199: 114064.

Comment 6: Related to my previous comment, I do not think the randomization results guarantee (as stated in the paper) that all 118 proteins are true positive findings. Further validations are needed such as using parallel reaction monitoring (PRM) in an independent validation set or some other way.

Reply 6: Thanks for your kindly suggestions. Please allow us to explain.
As mentioned on reply 5, randomized grouping statistical analysis was a strict strategy for omics clinical research. Our results showed that the average number of differential proteins was 10, showed that only 10 differential proteins could be generated randomly, further indicating that 91.5% of 118 differential proteins were reliable. Therefore, our findings indicated the current sample size suggested that urine can distinguish between autistic and non-autistic children. This preliminary study will provide clues for larger clinical studies for autism in this field.
We agree with your suggestions of further validations. Parallel reaction monitoring (PRM) is a classic targeted proteomic technique that is suitable for the validation of candidate specific proteins in large samples. In the early stage of urinary proteomics research, the targeted validations may miss valuable proteins. To confirm sufficient and reliable candidate biomarkers before validation is crucial for promoting the transformation of candidate biomarkers into clinical practice. As your helpful suggestions, we are collecting a large number of ASD patients with earlier-ages and multiple subtypes from multicenter for future research.
Changes in the text: We have modified our text as advised (see page 17, line 266-273).

Comment 7: It would be more clear if authors can use venn diagrams to show intersections and differences between these 118 findings, 21 given in the literature, 13 identified with ROC curve with AUC >0.9 (or a different cutoff).
Reply 7: Thanks for your kindly suggestions! According to your helpful suggestions, we added the Veen diagrams in *Figure 2B* to show the differences between these 118 differential proteins, 18 proteins (modified in revised manuscript) related to autism, and the top 13 differential proteins with AUC values (AUC >0.9).
Changes in the text: We have added Veen diagram in Figure 2B.

Comment 8: Minor Issues

• Page 1m line 27□An extra dot after "Results"

• There are missing spaces between the text and references in multiple places such as lines 47 (activities[1]), 56, 65,69, 72, ...please edit these throughout the paper carefully.

• Page 4, line 120□Please include a space after comma for the numbers listed in parentheses.

• Page 5, line 149□An extra space after 100 ms.

• Page 23, line 538□Dot is missing after Figure 4 caption • Supplemental Table 3, 4 and 5 should be written in bold.

Reply 8: Thanks for your kindly suggestions. We apologized for the error. According to your helpful suggestions, we have modified in the revised manuscript, including all details your mentioned above. Thanks for your suggestions.

Changes in the text: We have modified our text in all details mentioned above.

References:

Yan Wang, Jishui Zhang, Wenqi Song, Xiaoyi Tian, Ying Liu, Yanfei Wang, Jie Ma, Chengbin Wang, Guangtao Yan, A proteomic analysis of urine biomarkers in autism spectrum disorder, Journal of Proteomics, Volume 242, 2021, 104259, ISSN 1874-3919, https://doi.org/10.1016/j.jprot.2021.104259.

Suganya V, Geetha A, Sujatha S. Urine proteome analysis to evaluate protein biomarkers in children with autism. Clin Chim Acta. 2015 Oct 23;450:210-9. doi: 10.1016/j.cca.2015.08.015. Epub 2015 Aug 19. PMID: 26296899.

Citation: Yang L, Rudser K, Golnik A, Wey A, Higgins LA, et al. (2016) Urine Protein Biomarker Candidates for Autism. J Proteomics Bioinform S14: 004. doi:10.4172/jpb.S14-004

Reviewer C

Comment 1: The authors provide some important information on the proteome of autistic kids in urine samples. They should consider higher "n" in their future studies. The paper can be accepted after these minor comments.

Reply 1: Thanks for your kindly suggestions! We agree with your opinions.

We apologized that the limited size of samples was used in this study. Considered the limited samples, randomized grouping statistical analysis were conducted to confirm whether these differential proteins were disease differences. Our results showed that 91.5% of differential proteins were reliable, indicating that the urinary proteome could distinguish between autistic children and non-autistic children. This study

suggested that urine proteome is a promising approach for diagnosis of autism. A large number of autism patients with earlier-ages and various subtypes from multicenter should be considered in future studies.

Changes in the text: We have modified our text as advised (see page 17, line 266-273).

Comment 2: Change the title to only: Urinary proteome profiling 2 for children with autism using data-independent acquisition proteomics.

Reply 2: Thanks for your kindly suggestions.

There may be some problems with format display because the suggested title was the same as the original title. We apologize that we are a little confused about the title that the reviewer suggested us to modify. We would appreciate it if you could give us your advice again.

Changes in the text: We would appreciate it if you could give us your advice again.

Comment 3: The authors should cite other proteomics studies that measured proteins and PTMs in different regions in the body, here are some that need to be cited:
https://link.springer.com/chapter/10.1007/978-3-030-05542-4_12
https://pubmed.ncbi.nlm.nih.gov/29988084/

Reply 3: Thanks for your great suggestions! We apologized for omitting these previous proteomic studies for autism. According to your helpful suggestions, we cited these proteomics studies in the body in the revised manuscript.

Changes in the text: We have modified our text as advised (see page 3, line 60-61).

Comment 4: Does the non-autistic have other morbidities?

Reply 4: Thanks for your question. This is an important challenge for clinical researches. Please allow us to explain.

In theory, the healthy control group should be clinically healthy and free of any disease. In this study, urine samples of non-autistic children were collected from the children of Beijing Normal University staff, who showed no clinical manifestations of autism and had no obvious physical abnormalities. While a general physical examination would be preferable for the healthy control group, additional physical examinations for children for the purpose of this study were a challenge, mainly due to parental acceptance and high cost.

In future studies, we will consider comprehensive physical examination on the healthy control group according to your suggestion.

Changes in the text: We have modified our text as advised (see page 17, line 265-273).

Comment 5: Can the authors provide Venn diagram to visualize the difference between groups?

Reply 5: Thanks for your kindly suggestions. According to your helpful suggestions, we added the Veen diagrams in Figure 2B to show the differences between these 118 differential proteins, 18 proteins related to autism, and the top 13 differential proteins with AUC values (AUC >0.9).

Changes in the text: We have added Veen diagram in Figure 2B.

Comment 6: Figure 4 is related to which sets of proteins? Shared? Upregulated? Please clarify…

Reply 6: Thanks for your question and helpful suggestions. We apologized for the omission.

Figure 4 showed the functional annotations of all differential proteins in this study, we imported the118 differential proteins (all were down-regulated) to DAVID and IPA software to perform biological processes, molecular functions, cellular components and pathways. According to your suggestions, we have added more details of related proteins in the revised manuscript on Supplement Table 4 and 5.

Changes in the text: We have modified our text as advised (see page 15, line 222)