Molecular Plant
Correspondence

CellPress
Partner Journal

Reevaluating statistical methods in metabolomic studies: A case for Spearman's correlation

Dear Editor,

Recently, Liang et al. (2024) conducted comprehensive genome resequencing and metabolomic analyses of 299 representative accessions from extant citrus species, offering an extensive overview of genetic diversity and metabolic profiles across citrus populations. Their study included a correlation analysis that assessed the relationship between metabolite content and bioactivities, resulting in the selection of the top 100 annotated metabolites exhibiting the highest correlation coefficients (Pearson's r) across various chemical assays, cell lines, and CYP450 enzymes (Liang et al., 2024).

While this paper acknowledges the substantial efforts made by Liang et al. (2024), it raises important concerns regarding the use of Pearson's r for feature selection, primarily due to its inherent biases arising from linear assumptions that may not accurately reflect the complexities of biological data. Pearson's correlation is limited in scenarios where relationships between variables are non-linear or involve complex interactions, which are often the case in metabolomic and genomic studies. As such, reliance on Pearson's r could lead to oversimplified interpretations and potentially misleading conclusions.

Consequently, this discussion strongly advocates for the adoption of more robust, bias-free statistical methods, such as Spearman's correlation (Liu et al., 2018; Rovetta, 2020; Eden et al., 2022; Yu and Hutson, 2024), which offer significant advantages in the analysis of complex datasets commonly encountered in biological research. Spearman's correlation is a non-parametric, non-linear analysis tool that evaluates the rank order of data rather than relying on their actual values. This characteristic makes it particularly well suited for biological datasets, which frequently do not meet the assumptions of normality required for traditional parametric tests.

By circumventing these assumptions, Spearman's correlation significantly reduces the risk of skewed results and provides a more reliable assessment of relationships between variables. This enhancement is crucial in metabolomic and genomic studies, where understanding the intricate associations between metabolites and their bioactivities can yield deeper insights into biological processes. Moreover, Spearman's correlation includes p values, which help establish statistical significance and validate the robustness of the analysis. The ability to detect true associations, combined with a clearer interpretation of variable interactions, strengthens the overall reliability of the findings. Thus, incorporating Spearman's correlation not only enriches the analytical framework but also allows researchers to draw more accurate and meaningful conclusions from their data. This

methodological shift is essential for advancing the rigor of research and fostering a deeper understanding of the complexities inherent in biological systems.

It is also worth noting that while Liang et al. (2024) possess expertise in molecular plant studies, they may not have a comprehensive background in algorithmic computing and the potential biases that can arise from various statistical analyses. Researchers must be vigilant in understanding the underlying assumptions of the statistical methods they employ, as well as the biases and distortions that can emerge from these assumptions. This awareness is crucial for ensuring accurate data interpretation and robust conclusions in any scientific investigation. In light of these considerations, the incorporation of Spearman's correlation, along with a critical examination of statistical methodologies, emerges as essential for advancing the rigor and reliability of research outcomes in the field.

*Yoshiyasu Takefuji**
Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan
Correspondence: Yoshiyasu Takefuji (takefuji@keio.jp)
https://doi.org/10.1016/j.molp.2024.12.014

REFERENCES

Eden, S.K., Li, C., and Shepherd, B.E. (2022). Nonparametric estimation of Spearman's rank correlation with bivariate survival data. Biometrics **78**:421–434. https://doi.org/10.1111/biom.13453.

Liang, X., Wang, Y., Shen, W., Liao, B., Liu, X., Yang, Z., Chen, J., Zhao, C., Liao, Z., Cao, J., et al. (2024). Genomic and metabolomic insights into the selection and differentiation of bioactive compounds in citrus. Mol. Plant **17**:1753–1772. https://doi.org/10.1016/j.molp.2024.10.009.

Liu, Q., Li, C., Wanga, V., and Shepherd, B.E. (2018). Covariate-adjusted Spearman's rank correlation with probability-scale residuals. Biometrics **74**:595–605. https://doi.org/10.1111/biom.12812.

Rovetta, A. (2020). Raiders of the Lost Correlation: A Guide on Using Pearson and Spearman Coefficients to Detect Hidden Correlations in Medical Sciences. Cureus **12**:e11794. https://doi.org/10.7759/cureus.11794.

Yu, H., and Hutson, A.D. (2024). A robust Spearman correlation coefficient permutation test. Commun. Stat. Theor. Methods **53**:2141–2153. https://doi.org/10.1080/03610926.2022.2121144.