Journal of Affective Disorders 391 (2025) 120024

Contents lists available at ScienceDirect

Journal of Affective Disorders

journal homepage: www.elsevier.com/locate/jad

Correspondence

Reassessing lipid-mood disorder associations: The limitations of PCA for nonlinear patterns

ARTICLE INFO

Keywords:
Lipid patterns
Mood disorders
Principal component analysis
Nonlinear relationships
Statistical validation

ABSTRACT

Liu et al. (2025) analyzed UK Biobank data, using Principal Component Analysis (PCA) to identify lipid patterns associated with depression and bipolar disorder. Their work reported that the first principal component (PC1), reflecting Apolipoprotein B (ApoB), cholesterol, and low-density lipoprotein cholesterol (LDL-C), showed a protective effect against depression. However, their methodological approach warrants discussion. PCA is a linear dimensionality reduction technique. The authors noted nonlinear relationships between lipid profiles and mood disorder risk, contradicting PCA's inherent linearity assumption. Applying linear methods like PCA to nonlinear data can lead to significant distortions, systematic bias, and underfitting, failing to capture true data complexity. PC1 may have obscured genuine associations by forcing distinct biological features into a single linear equation, potentially diluting crucial signals. For future research, complementing PCA with unsupervised learning techniques like Feature Agglomeration (FA) and Highly Variable Gene Selection (HVGS) could offer a more robust approach. Additionally, using nonlinear nonparametric statistical methods such as Spearman's rho or Kendall's tau would be beneficial. These methods detect monotonic relationships without linearity assumptions, precisely capturing potentially nonlinear associations and enhancing interpretability in translational biomarker research.

Dear Editor,

Liu et al. (2025) analyzed UK Biobank data to uncover lipid patterns associated with depression and bipolar disorder. While their work provides valuable insights into the biological underpinnings of mood disorders, their methodological approach requires further discussion. At the core of their analysis, principal component analysis (PCA) was used to reduce the dimensionality of seven blood lipids into interpretable components, with the first three principal components (PCs) being derived. Specifically, the first principal component (PC1), primarily reflecting levels of Apolipoprotein B (ApoB), cholesterol, and low-density lipoprotein cholesterol (LDL-C), showed a protective effect against depression.

However, this paper raises a fundamental methodological concern. PCA is inherently a linear dimensionality reduction technique. It operates on the assumption that directions of maximum variance adequately capture the underlying structure of the data in a linear fashion. Yet, Liu et al. themselves acknowledge discovering nonlinear relationships between lipid profiles and mood disorder risk, a finding that directly contradicts the suitability of PCA for this particular dataset. When linear methods like PCA are applied to data with intrinsically nonlinear patterns, significant distortions can emerge in the resulting model. This methodological mismatch typically leads to systematic bias in parameter estimates and problematic underfitting, where the model fails to capture the true complexity of relationships present in the data (Cristian et al., 2024; Dyer and Kording, 2023; Elhaik, 2022; Yao and Ochoa, 2023; Mohseni and Elhaik, 2024). Such limitations can compromise both the validity and interpretability of the study's conclusions regarding lipid-mood disorder associations. Most critically, the first principal component (PC1) likely obscured genuine associations between individual lipid markers and mood disorders by forcing these biologically distinct features into a single linear equation, effectively diluting signals that might have been detected through direct analysis of the original variables.

Given that PCA inherently assumes linear relationships and prioritizes global variance capture over the preservation of local structure, its application to inherently nonlinear biological data risks obscuring clinically significant associations. More methodologically robust, multifaceted approaches would involve complementing PCA with unsupervised learning techniques like Feature Agglomeration (FA) and, where applicable, Highly Variable Gene Selection (HVGS) (Zhang et al., 2020; Arora et al., 2023). Additionally, the use of nonlinear nonparametric statistical methods such as Spearman's rho or Kendall's tau would be beneficial (Okoye and Hosseini, 2024; Yu and Hutson, 2024). These methods specifically detect monotonic relationships without imposing linearity assumptions, thereby capturing potentially nonlinear associations between individual lipid markers and mood disorder risk with greater precision and reliability. Beyond their statistical appropriateness, such nonparametric approaches offer enhanced interpretability—a critical consideration in translational biomarker research where findings must guide clinical decision-making. This interpretability advantage proves particularly valuable when communicating complex relationships to diverse stakeholders across the healthcare continuum, facilitating more effective translation from statistical findings to actionable clinical insights.

This correspondence offers a timely and critical evaluation of methodological choices in lipid biomarker research, with particular emphasis on the limitations of applying linear techniques such as PCA to biologically complex and nonlinear data. One of the main strengths of our analysis is the proposal of alternative analytical approaches,

https://doi.org/10.1016/j.jad.2025.120024
Received 1 July 2025; Received in revised form 27 July 2025; Accepted 4 August 2025
Available online 5 August 2025

including unsupervised learning and nonparametric methods, which may be better suited to capturing the true structure of associations between lipid profiles and mood disorders. As is the case with any commentary based on previously published findings, the generalizability of the observations may depend on the specific context and may not apply universally to all datasets or study designs. Nevertheless, this discussion provides a useful framework that can help refine future research methodologies in the field of psychiatric biomarker discovery.

While Liu et al. (2025) have made a valuable contribution to our understanding of lipid profiles in mood disorders through their analysis of UK Biobank data, their methodological approach warrants critical examination. The application of PCA to biological data exhibiting nonlinear relationships fundamentally contradicts the technique's underlying assumptions, potentially compromising the validity of their findings. This mismatch between the analytical method and data characteristics likely resulted in systematic bias and underfitting of the true relationships between lipids and mood disorders. Future research in this domain would substantially benefit from employing alternative analytical techniques better suited to capturing nonlinear associations. Such approaches would not only offer statistical advantages but also enhance interpretability—a crucial consideration for translating biomarker research into clinical applications. Addressing these methodological concerns would strengthen the foundation for identifying reliable lipid biomarkers in psychiatric conditions.

CRediT authorship contribution statement

Soki Ogawa: Writing – original draft, Investigation. **Souichi Oka:** Writing – review & editing, Conceptualization. **Yoshiyasu Takefuji:** Supervision, Project administration.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

None.

Data availability

No new data was generated or analyzed in support of this research.

References

Arora, J.K., Opasawatchai, A., Teichmann, S.A., Matangkasombut, P., Charoensawan, V., 2023. Computational workflow for investigating highly variable genes in single-cell RNA-seq across multiple time points and cell types. STAR Protocols 4 (3), 102387. https://doi.org/10.1016/j.xpro.2023.102387.

Cristian, P.M., Aarón, V.J., Armando, E.D., Jesús, E.S.V., José, L.G.H., 2024. Diffusion on PCA-UMAP manifold: the impact of data structure preservation to denoise high-dimensional single-cell RNA sequencing data. Biology 13 (7), 512. https://doi.org/10.3390/biology13070512.

Dyer, E.L., Kording, K., 2023. Why the simplest explanation isn't always the best. Proc. Natl. Acad. Sci. 120 (52), e2319169120. https://doi.org/10.1073/pnas.2319169120.

Elhaik, E., 2022. Principal component analyses (PCA)-based findings in population genetic studies are highly biased and must be reevaluated. Sci. Rep. 12 (1), 14683. https://doi.org/10.1038/s41598-022-14395-4.

Liu, X., Yang, B., Zhou, X., Smith, D.J., Li, M., Zhang, C., 2025. Blood lipid profiles and mood disorders: a principal component analysis of UK biobank data reveals distinct associations with depression and bipolar disorder. J. Affect. Disord. 377, 23–34. https://doi.org/10.1016/j.jad.2025.02.040.

Mohseni, N., Elhaik, E., 2024. Biases of principal component analysis (PCA) in physical anthropology studies require a reevaluation of evolutionary insights. eLife 13, RP94685. https://doi.org/10.7554/eLife.94685.2.

Okoye, K., Hosseini, S., 2024. Correlation tests in R: Pearson Cor, Kendall's tau, and spearman's rho. In: Okoye, K., Hosseini, S. (Eds.), R Programming: Statistical Data Analysis in Research. Springer Nature, pp. 247–277. https://doi.org/10.1007/978-981-97-3385-9_12.

Yao, Y., Ochoa, A., 2023. Limitations of principal components in quantitative genetic association models for human studies. eLife 12, e79238. https://doi.org/10.7554/eLife.79238.

Yu, H., Hutson, A.D., 2024. A robust spearman correlation coefficient permutation test. Commun. Stat. - Theory Methods 53 (9), 2141–2153. https://doi.org/10.1080/03610926.2022.2121144.

Zhang, J., Wu, X., Hoi, S.C.H., Zhu, J., 2020. Feature agglomeration networks for single stage face detection. Neurocomputing 380, 180–189. https://doi.org/10.1016/j.neucom.2019.10.087.

Soki Ogawa[a,*], Souichi Oka[b], Yoshiyasu Takefuji[c]

[a] *Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan*

[b] *Science Park Corporation, 3-24-9 Iriya-Nishi Zama-shi, Kanagawa 252-0029, Japan*

[c] *Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan*

[*] Corresponding author at: Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan.

E-mail addresses: g2550001@stu.musashino-u.ac.jp (S. Ogawa), souichi.oka@sciencepark.co.jp (S. Oka), takefuji@keio.jp (Y. Takefuji).