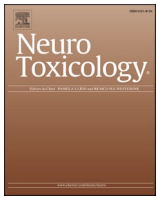
Neurotoxicology 114 (2026) 103456

Contents lists available at ScienceDirect

Neurotoxicology

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



Reconsidering principal component analysis in neurodevelopmental studies: A call for advanced frameworks

ARTICLE INFO

Edited by Dr. John O'Donoghue

Keywords:
Neurodevelopment
Principal component analysis
Nonlinear analysis
Non-parametric correlation
Statistical validation

ABSTRACT

Bodin et al. (2025) provide valuable insights into neurodevelopmental vulnerability by examining radiofrequency electromagnetic fields (RF-EMF) exposure during early life. Their integrative design, combining whole-body exposure with endpoints such as neonatal brain proteomics, BDNF expression, synaptogenesis, and oxidative stress, offers a comprehensive framework for developmental neurotoxicology. However, interpretation of proteomic clustering relies heavily on principal component analysis (PCA), a linear technique ill-suited for high-dimensional datasets dominated by non-linear dependencies and strong inter-feature correlations. PCA plots (Fig. 3) illustrate group separation, yet variance explained (55%) and clustering stability remain underreported, raising concerns about robustness and biological interpretability, particularly given only ten differentially expressed proteins. To enhance inference, future studies should adopt biologically meaningful feature selection and advanced frameworks such as Feature Agglomeration and Highly Variable Feature Selection, alongside non-parametric correlation measures such as Spearman's rho and Kendall's tau. These strategies will improve reproducibility, uncover mechanistic patterns, and strengthen translational relevance for neurodevelopmental research.

Bodin et al. (2025) deliver an important contribution to developmental neurotoxicology by investigating radiofrequency electromagnetic fields (RF-EMF) exposure during a critical neurodevelopmental window. Their multifaceted design, combining whole-body exposures with diverse endpoints such as neonatal brain proteomics, BDNF expression, synaptogenesis, and oxidative stress, offers a framework for assessing early-life vulnerability. For example, oxidative stress markers showed an age-related increase (Figure 7 C), while cortical BDNF expression declined at PND 17 under exposure (Figure 6 A), suggesting potential implications for synaptic development. However, the interpretation of global proteomic similarity and group separation warrants further discussion, because principal component analysis (PCA)–based exploratory visualizations are prone to unintentional overinterpretation in high-dimensional proteomic contexts.

While PCA provides a useful visualization of group separation, its linear assumptions may fail to capture non-linear dependencies inherent in neurodevelopmental proteomes, where interactions among neurotrophic signaling, oxidative stress, and synaptic organization are highly complex (Mohseni and Elhaik, 2024). These datasets exhibit compositional constraints and strong inter-feature correlations, making linear projections prone to distortion and potentially obscuring biologically relevant clusters. LC-MS proteomics yields relative abundance data with closed-sum constraints, which can induce spurious correlations and make PCA projections sensitive to preprocessing; this is a general methodological note rather than a critique of any specific dataset. Although PCA plots (Figure 3) illustrate separation between sham and exposed groups, critical metrics such as clustering stability and variance explained—reported as 55% for PC1 and PC2 combined— may be insufficient on its own for assessing robustness and reproducibility in high-dimensional settings, where additional stability checks can be informative. Furthermore, although ten proteins were identified as differentially expressed, even limited but consistent proteomic changes can be biologically meaningful; however, such situations increase the importance of analytical robustness and cautious interpretation. In the absence of public access to the dataset, we instead reference empirical work demonstrating that PCA can exhibit instability and reduced interpretive robustness in high-dimensional biological settings (Oka and Takefuji, 2026). This evidence does not critique the original study but highlights a general limitation of PCA when applied to complex biological data structures.

This reliance introduces two analytical limitations. First, PCA's linearity and orthogonality assumptions are ill-suited to high-dimensional proteomes dominated by non-linear dependencies. For instance, Bodin et al. applied PCA to infer group separation despite strong inter-feature correlations and compositional constraints. Second, downstream interpretations may implicitly inherit parametric assumptions from statistical tests applied to noisy data, conditions under which these assumptions frequently fail. Such oversimplification risks misrepresenting mechanistic pathways, particularly those involving synaptic organization and oxidative stress, thereby limiting translational insight.

To address these limitations, future studies should prioritize meaningful feature selection and adopt frameworks that better reflect neurodevelopmental complexity. High predictive accuracy alone does not guarantee reliable identification of exposure-related features, and interpretability remains a challenge. Robust unsupervised approaches such as Feature Agglomeration and Highly Variable Feature Selection offer strategies for capturing latent subpopulations and stage-dependent patterns without imposing linear constraints (Zhang et al., 2020; Xie et al., 2025). Additionally, non-parametric correlation measures (e.g.,

https://doi.org/10.1016/j.neuro.2026.103456
Received 24 December 2025; Received in revised form 14 March 2026; Accepted 16 April 2026
Available online 20 April 2026

Spearman's rho, Kendall's tau) are well-suited for detecting monotonic relationships among proteomic features, providing resilience to outliers and distributional assumptions (Okoye and Hosseini, 2024). Incorporating these strategies will enhance statistical rigor and uncover mechanistic patterns that inform translational research on RF-EMF-induced neurodevelopmental alterations, ultimately improving reproducibility and clinical relevance. Although a full methodological comparison lies beyond the scope of this correspondence, it is worth noting that several data-centric approaches have been historically underutilized despite demonstrated stability and interpretive advantages. Recent empirical evidence (Oka and Takefuji, 2026) indicates that these methods provide more reproducible feature prioritization than PCA, underscoring their relevance as robust alternatives for future neurodevelopmental proteomics. Network-based nonlinear approaches are also widely used in proteomics and transcriptomics; for example, network propagation has been shown to amplify biological signal by leveraging graph-structured molecular relationships (Cowen et al., 2017).

In conclusion, Bodin et al.'s work advances understanding of RF-EMF-induced neurodevelopmental alterations under ICNIRP thresholds. Moving beyond linear assumptions is essential for uncovering subtle proteomic and cellular dynamics. Incorporating advanced analytical frameworks will sharpen mechanistic inference, enhance reproducibility and contribute to more robust mechanistic insight in future neurodevelopmental studies, rather than implying direct therapeutic or clinical consequences. Ultimately, integrating computational tools with proteomics can bridge molecular insights and actionable interventions, supporting more informed risk assessment and guiding evidence-based strategies for chronic neurodevelopmental risk management.

CRediT authorship contribution statement

Takefuiji Yoshiyasu: Writing – review & editing, Supervision, Project administration. **Souichi Oka:** Writing – original draft, Conceptualization. **Ryota Ono:** Investigation.

Funding Statement

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.

Data availability

No data was used for the research described in the article.

References

Bodin, R., Godin, L., Mougin, C., Lecomte, A., Larrigaldie, V., Feat-Vetel, J., Méresse, S., Montécot-Dubourg, C., Marcelo, P., Mortaud, S., Villegier, A.-S., 2025. Altered development in rodent brain cells after 900 MHz radiofrequency exposure. NeuroToxicology 111, 103312. https://doi.org/10.1016/j.neuro.2025.103312.

Cowen, L., Ideker, T., Raphael, B.J., Sharan, R., 2017. Network propagation: a universal amplifier of genetic associations. Nat. Rev. Genet. 18, 551–562.

Mohseni, N., Elhaik, E., 2024. Biases of principal component analysis (PCA) in physical anthropology studies require a reevaluation of evolutionary insights. eLife 13, RP94685.

Oka, S., Takefuji, Y., 2026. Revisiting AI interpretability in precision oncology: why predictive accuracy does not ensure stable feature importance. Cancers 18, 593.

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.

Xie, Y., Jing, Z., Pan, H., Xu, X., Fang, Q., 2025. Redefining the high variable genes by optimized LOESS regression with positive ratio. BMC Bioinform. 26, 104.

Zhang, J., Wu, X., Hoi, S.C.H., Zhu, J., 2020. Feature agglomeration networks for single stage face detection. Neurocomputing 380, 180–189.

Souichi Oka[*,1], Ryota Ono[2]
Science Park Corporation, 3-24-9 Iriya-Nishi, Zama-shi, Kanagawa 252-0029, Japan

Yoshiyasu Takefuji[3]
Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan
E-mail address: takefuji@keio.jp.

[*] Corresponding author.
E-mail addresses: souichi.oka@sciencepark.co.jp (S. Oka), rono@sciencepark.co.jp (R. Ono).

[1] ORCID: 0009-0000-4840-5232.

[2] ORCID: 0009-0009-0905-8739.

[3] ORCID: 0000-0002-1826-742X.