Corrigendum: A Practical Guide to Sparse *k*-Means Clustering for Studying Molecular Development of the Human Brain

Justin L. Balsor[1], Keon Arbabi[1], Desmond Singh[2], Rachel Kwan[2], Jonathan Zaslavsky[2], Ewalina Jeyanesan[1] and Kathryn M. Murphy[1,2]*

[1] *McMaster Neuroscience Graduate Program, McMaster University, Hamilton, ON, Canada,* [2] *Department of Psychology, Neuroscience and Behavior, McMaster University, Hamilton, ON, Canada*

A Corrigendum on

A Practical Guide to Sparse *k*-Means Clustering for Studying Molecular Development of the Human Brain
by Balsor, J. L., Arbabi, K., Singh, D., Kwan, R., Zaslavsky, J., Jeyanesan, E., and Murphy, K. M. (2021). Front. Neurosci. 15:668293. doi: 10.3389/fnins.2021.668293

In the original article, there was an error in the section **Discussion, Paragraph Number 2 as published**. The following citations were missing (Murphy et al., 2005; Pinto et al., 2010, 2015; Williams et al., 2010; Witten and Tibshirani, 2010; Siu et al., 2015, 2017). The revised paragraph is included below:

"Many factors come into play when selecting an appropriate clustering algorithm for a study. Here, we considered the goal of the study (to resolve sometimes subtle age-related changes in molecular mechanism), the structure of the dataset ($p \sim n$ to $p >> n$), and the output of the algorithm (is it just the clusters or is feature selection included). Sparse K-means clustering was selected because it fit all of those considerations. We know from previous studies of the molecular development of the human brain that there can be subtle differences between age groups (Murphy et al., 2005; Pinto et al., 2010, 2015; Williams et al., 2010; Siu et al., 2015, 2017), and yet even small changes in protein or gene expression will alter neural function. Therefore, we looked for algorithms designed for omics datasets where subtle changes in a subset of the genes or proteins would identify important characteristics of the data. The development of sparse K-means clustering by Witten and Tibshirani (2010) was partially inspired by the need to better cluster a breast cancer dataset. In that dataset, subtle differences in gene expression significantly impacted patient outcomes, but standard clustering approaches did not pick those up. In addition, sparse clustering was developed to address datasets, like ours and the breast cancer data where the structure is $p \sim n$ to $p >> n$. Sparse K-means clustering is a good fit for those high dimensional structures because it minimizes the within-cluster sum of squares with a dissimilarity measure while maximizing the between-cluster sum of squares by iteratively reweighting the measures. Finally, and most importantly, sparse K-means clustering performs feature selection. The examples in this paper show the reweighted proteins and those distributions identifying how much each protein contributes to partitioning the samples into clusters. That matrix is sparse, with unimportant

proteins having near-zero weights and important ones having non-zero weights. Those weights are essential for cluster analysis to help with making neurobiologically relevant interpretations of brain development from the cluster analysis."

The authors apologize for this error and state that this does not change the scientific conclusions of the article in any way. The original article has been updated.

REFERENCES

Murphy, K. M., Beston, B. R., Boley, P. M., and Jones, D. G. (2005). Development of human visual cortex: a balance between excitatory and inhibitory plasticity mechanisms. *Dev. Psychobiol.* 46, 209–221. doi: 10.1002/dev.20053

Pinto, J. G. A., Hornby, K. R., Jones, D. G., and Murphy, K. M. (2010). Developmental changes in GABAergic mechanisms in human visual cortex across the lifespan. *Front. Cell Neurosci.* 4:16. doi: 10.3389/fncel.2010.00016

Pinto, J. G. A., Jones, D. G., Williams, C. K., and Murphy, K. M. (2015). Characterizing synaptic protein development in human visual cortex enables alignment of synaptic age with rat visual cortex. *Front. Neural. Circuit* 9:3. doi: 10.3389/fncir.2015.00003

Siu, C. R., Balsor, J. L., Jones, D. G., and Murphy, K. M. (2015). Classic and Golli Myelin Basic Protein have distinct developmental trajectories in human visual cortex. *Front. Neurosci-switz* 9:138. doi: 10.3389/fnins.2015.00138

Siu, C. R., Beshara, S. P., Jones, D. G., and Murphy, K. M. (2017). Development of Glutamatergic Proteins in Human Visual Cortex across the Lifespan. *J. Neurosci.* 37:6042. doi: 10.1523/jneurosci.2304-16.2017

Williams, K., Irwin, D. A., Jones, D. G., and Murphy, K. M. (2010). Dramatic Loss of Ube3A Expression during Aging of the Mammalian Cortex. *Front. Aging Neurosci.* 2:18. doi: 10.3389/fnagi.2010.00018

Witten, D. M., and Tibshirani, R. (2010). A framework for feature selection in clustering. *J. Am. Stat. Assoc.* 105, 713–726. doi: 10.1198/jasa.2010.tm09415