



## COMMENT

# The use of gene expression as a disease stratification tool of neonatal encephalopathy

Constance Burgod<sup>1</sup>, Sudhin Thayyil<sup>1</sup> and Paolo Montaldo<sup>1,2</sup>*Pediatric Research* (2021) 89:12–13; <https://doi.org/10.1038/s41390-020-1104-2>

We congratulate Balada et al.<sup>1</sup> for pursuing important transcriptomic research in neonatal encephalopathy (NE). They examined the messenger RNA (mRNA) expression of 23 candidate genes in whole blood using real-time polymerase chain reaction from 24 babies with NE and 34 control babies (1 healthy; 33 with mild polycythaemia), and also at ages <6 h ( $n = 15$  NE vs. 2 controls), 12 h ( $n = 17$  NE vs. 1 control), 24 h ( $n = 18$  NE vs. 2 controls), 48 h ( $n = 14$  NE vs. 13 controls), 72 h ( $n = 13$  NE vs. 9 controls), and 96 h ( $n = 4$  NE vs. 7 controls). Six genes (*MMP9*, *PPARG*, *IL8*, *HSPA1A*, *TLR8*, and *CCR5*) had a discriminant graphical pattern between babies with NE and control infants, and hence were selected for subsequent statistical modeling.

While this is useful information, we would like to highlight some limitations of these observations and methodology. First, the use of a preselected panel of just 23 genes may lead to erroneous conclusions, particularly in preliminary studies. Often, the results of microarray-based gene expression studies are not reproducible, and when the analysis is limited to a small set of preselected genes, the significance of the observed differences is even more difficult to interpret.<sup>2</sup> Second, two controls were recruited within 6 h and only one within 12 h, which limited the contribution of the first two time points to the identification of the differentially expressed genes to be used for the analysis. While we acknowledge that recruitment of control infants represents a significant challenge for neonatal studies, the identification of a gene expression profile within 6 h of birth is pivotal since the therapeutic window for neonatal neuroprotective interventions is narrow, and treatments need to be initiated within hours after birth to have a chance of success. Finally, this is not the first report on use of gene expression in NE, as authors have claimed. We had previously reported differences in gene expression profiles of babies with NE, healthy control, and those with sepsis, using next-generation sequencing approach.<sup>3,4</sup>

To illustrate this issue, we examined the differential expression of the six genes reported by Balada et al.<sup>1</sup> in our previously published dataset of 12 encephalopathic babies and 6 time-matched healthy term controls.<sup>3</sup> We used linear mixed-effect models to compare the expression of the six genes (*MMP9*, *IL8*, *HSPA1A*, *CCR5*, *PPARG*, *TLR8*) in the two groups during the first 11 h of age. The linear mixed-effect models were adjusted by gender and mothers' age and were individually developed for each of the genes. In agreement with Balada et al.,<sup>1</sup> our results confirmed a higher expression of

matrix metalloproteinase 9 (*MMP9*) and peroxisome proliferator activated receptor (*PPARG*) in neonatal encephalopathic patients during the first 11 h of age compared to the control group (+1.07 log (FPKM (fragments per kilobase of transcript per million mapped)),  $p = 0.0511$ , 95% confidence interval (CI) [0.16, 1.99]; +1.65 log (FPKM),  $p = 0.0278$ , 95% CI [0.37, 2.87]). However, the variations in expression of the remaining four genes were not validated. Figure 1 depicts expression of the six genes over the first 11 h of life in each group.

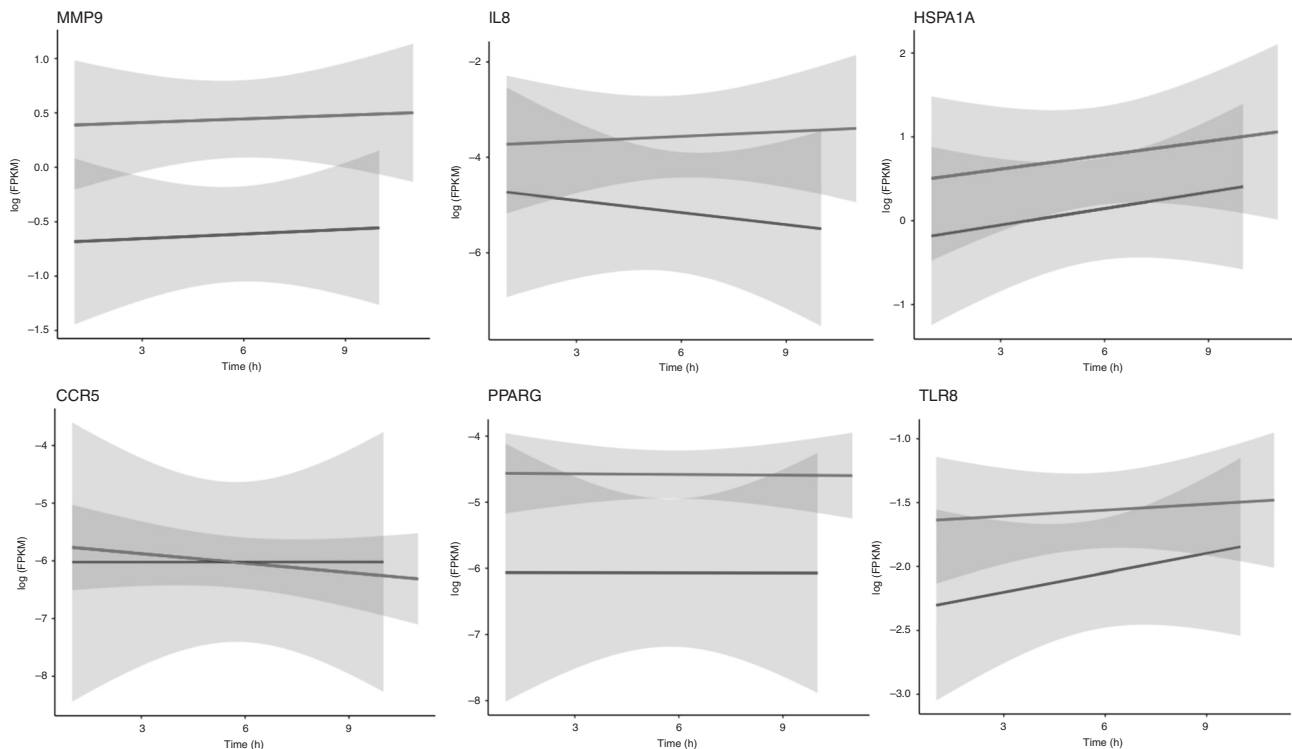
We must emphasize that this analysis does not suggest any inaccuracy in the analysis reported by Balada et al.,<sup>1</sup> and merely represents a pitfall of their approach to biomarker discovery. Gene preselection is a controversial issue in biological experiments and may result in precipitous conclusions. Although transcriptomic signatures have great potential for developing personalized neuroprotection, these are still in an early stage of development with only few studies in NE so far.<sup>3,5</sup> Hence, we recommend a cautious and rigorous approach before basing biomarker identification on preselected genes, which can lead to bias, and to type 1 error. Moving away from hypothesis testing or setting it as a validation step following genome-wide research may allow us to select an appropriate and reproducible combination of genes involved in NE.<sup>6</sup>

The importance of choosing the most appropriate approach for biomarker discovery has also been recently highlighted in microRNA (miRNA) studies. Circulating miRNAs, the regulators of gene expression, are often identified as candidate biomarkers. However, small sample sizes and use of low vs. high-throughput sequencing technologies may lead to results that are not subsequently validated.<sup>7,8</sup> This issue has been highlighted in a cord blood miRNA study, which assessed miRNA profiles in blood plasma from 38 mothers and their newborns. The authors found low correlation between cord and maternal samples based on the identified miRNA signature. However, the sample size was too small for both the training and validation of these results in a robust predictive model.<sup>9</sup>

Genome-wide profiling of gene expression levels can be performed without a prior hypothesis, ensuring all genes are considered equally. This would also allow potential discoveries of novel genes involved in disease onset and outcome prediction and would prevent from drawing conclusions solely based on a narrowed-down set of genes. The paradigm of hypothesis-generating research should complement and facil-

<sup>1</sup>Centre for Perinatal Neuroscience, Imperial College London, London, UK and <sup>2</sup>Neonatal Unit, Università degli Studi della Campania "Luigi Vanvitelli", Naples, Italy  
Correspondence: Paolo Montaldo (p.montaldo@imperial.ac.uk)

Received: 19 June 2020 Revised: 21 July 2020 Accepted: 23 July 2020  
Published online: 3 August 2020



**Fig. 1** Expression of the six selected genes analyzed by LMM in controls and NE patients during the first 11 h of life. The LMMs exhibit the relationship between hours and gene count values expressed as logarithm of fragments per kilobase of transcript per million mapped reads. 95% Confidence intervals are represented on each side of the regression lines by a gray area. Blue lines indicate control and red lines indicate NE infants.

itate discoveries that have not been possible through the exclusive use of hypothesis-driven methodologies.

#### AUTHOR CONTRIBUTIONS

P.M. and S.T. conceptualized and designed the study. P.M. and C.B. collected the data. P.M. and C.B. analyzed the data. C.B. and S.T. drafted the initial manuscript. P.M. supervised the study. S.T. and P.M. reviewed and revised the manuscript critically. All authors approved the final manuscript as submitted.

#### ADDITIONAL INFORMATION

**Competing interests:** The authors declare no competing interests.

**Patient consent:** All the parents of the neonates who were enrolled in the study provided their written informed consent.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### REFERENCES

- Balada, R. et al. Enquiring beneath the surface: can a gene expression assay shed light into the heterogeneity among newborns with neonatal encephalopathy? *Pediatr. Res.* <https://doi.org/10.1038/s41390-020-0764-2> (2020).
- Ioannidis, J. P. et al. Repeatability of published microarray gene expression analyses. *Nat. Genet.* **41**, 149–155 (2009).
- Montaldo, P. et al. Whole blood gene expression reveals specific transcriptome changes in neonatal encephalopathy. *Neonatology* **115**, 68–76 (2019).
- Montaldo, P., Lally, P. J., Oliveira, V. & Thayyil, S. Hypothermic neuroprotection for neonatal encephalopathy in low- and middle-income countries: a new approach to an old problem. *NeoReviews* **19**, e735–e741 (2018).
- Montaldo, P. et al. Transcriptomic profile of adverse neurodevelopmental outcomes after neonatal encephalopathy. *Nat. Scientific Rep.* <https://doi.org/10.1038/s41598-020-70131-w> (2020) (in press).
- Peng, R. D. Reproducible research in computational science. *Science (New York, N. Y.)* **334**, 1226–1227 (2011).
- Kok, M. G. M. et al. Small sample sizes in high-throughput miRNA screens: a common pitfall for the identification of miRNA biomarkers. *Biomol. Detect. Quantif.* **15**, 1–5 (2017).
- Pritchard, C. C., Cheng, H. H. & Tewari, M. MicroRNA profiling: approaches and considerations. *Nat. Rev. Genet.* **13**, 358–369 (2012).
- Dypås, L. B., Gützow, K. B., Olsen, A.-K. & Duale, N. MiRNA profiles in blood plasma from mother–child duos in human biobanks and the implication of sample quality: circulating miRNAs as potential early markers of child health. *PLoS ONE* **15**, e0231040-e (2020).