

CORRESPONDENCE

Open Access



A statistical and biological response to an informatics appraisal of healthy aging gene signatures

James A. Timmons^{1*} , Iain J. Gallagher^{5*}, Sanjana Sood¹, Bethan Phillips², Hannah Crossland², Robert Howard³, William E. Kraus⁴ and Philip J. Atherton²

Abstract

Jacob and Speed did not identify even a single example of a '150-gene-set' that was statistically significant at classifying Alzheimer's disease (AD) samples, or age in independent studies. We attempt to clarify the various misunderstandings, below.

In 2013 we discovered a 150 RNA 'gene-set' from healthy skeletal muscle (others named HAG) that proved to be a high performing, statistically significant, *non-linear* age classifier of multiple types of human tissue [1]. We profiled muscle samples from an independent birth cohort and found our HAG score correlated with 'successful aging' over two decades, using rank-order methods which incorporated the *direction* of gene expression change from the original model [1]. The following year, we found that the HAG score was regulated with age in human hippocampus and in human blood profiles from AD [1]. Since then over 50 HAG have been linked, by other laboratories, to the biology of age and dementia (see Additional file 1), robust confirmation of the biological validity of the HAG. We illustrated the potential utility of the HAG, by combining it with a set of AD regulated 'disease' genes [2] to yield a test that distinguished AD samples from *age-* and *gender-*matched controls (Sood et al Fig. 5 [1]). The original disease signature (the Lunnon et al AD signature [2]) was not statistically significant, on its own, in an independent AD cohort. Notably this 150 probe-set plus 48 gene prototype 'AD blood assay' (red-dot, Age+AD-disease) out performs all

150 gene-sets presented by Jacob and Speed (Figure 1a) and remains one of the only RNA assays validated in an independent AD cohort. We presented our transcriptional age model as a logical starting point for machine learning requiring independent data, to produce a tool to facilitate clinical AD research (i.e. screen for accelerated 'aging', *the* risk factor for AD).

Substantial misunderstandings of our article, by Jacob and Speed, led to them create a narrative through reference to Ein-Dor et al [Ref 2 in their letter], that our work was an example of a classification solution found using a small clinical cohort [3]. Typically such solutions do not transfer to independent clinical cohorts, because the model is over-fitted to *all* the characteristics of the first data set. Notably it was impossible for the HAG signature to belong to this category, as we did not select our gene list using an AD cohort. Importantly, our single HAG '150 gene list' was confirmed to be statistically significant in two independent AD cohorts [1]. In contrast, Jacob and Speed calculated thousands of ROC values; did not apply any statistical test for significance, and never confirmed the performance of any of their gene lists in any independent cohorts. For clarity on this complex topic, Table 1 contrasts some of the major differences in approach and rigor, taken by our team versus the approach presented by Jacob and Speed. It is noteworthy that the approach we took is far more laborious, and we understand why Jacob and Speed may have chosen to skip over most of the following time consuming and important steps.

* Correspondence: jamie.timmons@gmail.com; ij.gallagher@stir.ac.uk
Correspondence to: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0750-x>

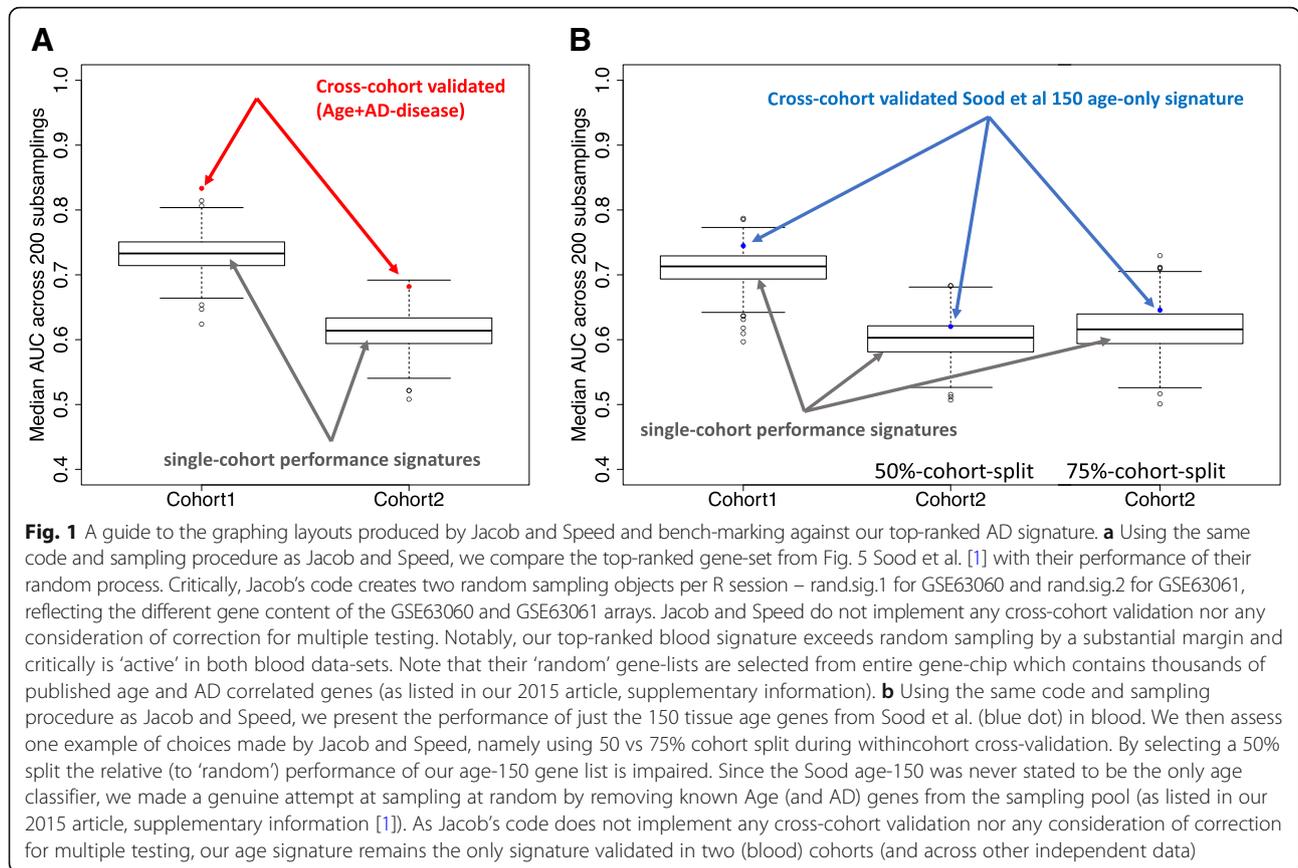
The Research to this article has been published in *Genome Biology* 2015 16:185: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0750-x>

¹Kings College London, London, UK

⁵University of Stirling, Stirling, UK

Full list of author information is available at the end of the article





To revisit Jacob and Speed analysis we used their computational code, and found that their *within-cohort* modeling of the AD data (an approach we did not use as it is unreliable [4] (Table 1)) yielded 150 gene-sets with performance that reflected laboratory batch-noise within the AD Illumina blood RNA arrays (AdNeuroMed consortium [2]). We illustrate this, by sampling using their code, but only from non-expressed ("background") genes (Additional file 1: Figure S1). This noise is unfortunate, but the illustration demonstrates that the performance of their random sampling protocol was not driven, as

claimed, to 'interconnected biology' or 'shared gene expression variance'. Furthermore, their claim that our 150-gene age signature (Fig. 1b, blue-dot) lay 'within' the performance range obtained via random sampling in blood is neither correct nor a fair like-for-like statistical comparison [4]. Our single 150 HAG test was statistically significant in two independent AD cohorts, and exceeded any background noise. Jacob and Speed random gene-sets require statistical correction for thousands of multiple-tests and notably they did not present any statistical significance values. The importance of our

Table 1 A comparison of the approaches of Sood et al. and Jacob and Speed

Performance criteria	Sood et al.	Jacob and Speed
Hypothesis driven gene selection in independent samples	Yes	No
Tissue age ROC performance externally validated	Yes	No
Multi-tissue age classification achieved	Yes	No
Longitudinal health association achieved	Yes	No
Top age gene-set statistically tested in blood	Yes	No
ROC performance assessed within an AD cohort	Yes	Yes
Top gene-set statistically significant across AD cohorts	Yes	No
Top gene-set independently linked to age and dementia	Yes	No
Gene-set trained in AD clinical cohort	No	Yes

approach, using hypothesis driven signatures and robust external validation (i.e. using independent data to validate a model) over their ‘easy to perform’ within data sampling is neither controversial or recent knowledge - See Konig et al. and cited articles [4].

We also found that their (less reliable) ‘within cohort’ ROC performance approach was enhanced by some of their choices e.g. 50% data-splits resulted in ~10% gain in ROC value, in their favor (Fig. 1b). Their random gene lists were not actually ‘random’ from a biological perspective. Sampling occurred from known age and AD correlated genes (see methods). Removal of age and AD genes known at the time of our study (2015) also reduces the performance of sampling at random (Additional file 1: Figure S2). Curiously, Jacob and Speed did not match AD cases with controls, for chronological age and gender (the two greatest risk factors for AD). Instead they combined AD with the Mild Cognitive Impairment (MCI) samples - a heterogenous population, comprising several clinical conditions and many who never develop AD [5]. This was clinically invalid but also unnecessary, as all required information was included in the GEO repository e.g. age, gender and clinical sample status (embedded in GSE63060 and GSE63061 as visualized in his files supplied with their letter (Additional file 1: Figure S3)). Combining the AD and MCI clinical samples also exaggerated the *within* cohort performance of their approach while it impaired the performance of our top-ranked gene-set (Additional file 1: Figure S4). Critically Jacob and Speed should have been able to replicate the design of our study analysis as all data necessary was at GEO, as illustrated in Additional file 1: Figure S3 - a screen shot taken directly from implementing the code provided by Jacob and Speed.

Finally, Jacob and Speed have claimed that our tissue age model was itself ‘unremarkable’. A clue to the exceptional performance of our age signature was the observation that when using muscle RNA profiles as the external validation data (classification space), the age of brain samples was classified correctly [1]. However, we formally revisited this issue by comparing our [1] HAG signature with 10,000 gene-sets chosen at random from the same muscle cohort. We calculated performance of each recorded ‘at random’ gene-set, using *cross-cohort* ‘gold standard’ external validation (four *independent* muscle data-sets; code and methods are included in the Additional file 1). Our 150 HAG signature was ranked better than all 10,000 ‘random’ gene-sets (Additional file 1: Figure S5A). We calculated the statistical significance of the average ‘random’ gene-set, and the performance of an ‘age model selected at random’ was not significant (*unadjusted* $p = 0.113$), despite having an ROC value > 0.6. Even gene-sets discovered and tested *within* a

single cohort were largely inferior to our age signature (Additional file 1: Figure S5B). Our observations lead to the conclusion that the claims made by Jacob and Speed are flawed, at least in part because they did not present a single example of a new 150-gene-set that significantly worked *across* independent data as a tissue age classifier, a muscle-based health prognostic or a blood AD classifier. Thus, while we agree with Ein-Dor et al, that classifiers built on small clinical cohorts should be treated with caution; reflecting the interactive network of tissue and cellular gene expression [6] and simple technical factors such as noise, as we did not build a disease classifier using AD or any other disease samples, we find Jacob and Speed's criticism [3] of our work unfounded.

Additional file

Additional file 1: Supplementary figures S1-S5, Supplementary Methods and code. (DOCX 3780 kb)

Acknowledgements

We thank a number of colleagues, including Professor Claes Wahlestedt, for their constructive comments, during the preparation of this response.

Authors' contributions

The computational code used was produced by Jacob and Speed and inspected by SS, JT and IG. The original Jacob and Speed codes were implemented by SS during 2016 and 2017. The most recent Jacob and Speed code was implemented by JT during 2018, and results inspected by IG and SS. The original Knn classification R code was designed by JT, SS and IG and implemented by SS. The muscle random sampling code with 4 study external validation was produced by SS, and the data analysis checked by JT and IG. All primary and secondary data analysis results were discussed by JT, PA, BP and HC at meetings during 2018, and the responses to Jacob and Speed were then drafted by JT, PA, SS and IG. Various iterations of the document were edited by RH and WK. The final response was reviewed and approved by all authors.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Kings College London, London, UK. ²University of Nottingham, Nottingham, UK. ³University College London, London, UK. ⁴Duke University, Durham, USA. ⁵University of Stirling, Stirling, UK.

Received: 27 November 2018 Accepted: 10 June 2019

Published online: 02 August 2019

References

1. Sood S, Gallagher IJ, Lunnon K, Rullman E, Keohane A, Crossland H, et al. A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol.* 2015;16:185 Available from: <http://genomebiology.com/2015/16/1/185>.
2. Lunnon K, Ibrahim Z, Proitsi P, Louridas A, Newhouse S, Sattler M, et al. Mitochondrial dysfunction and immune activation are detectable in early Alzheimer's disease blood. *J Alzheimers Dis.* 2012;30:685–710 [cited 2014 Oct 19]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22466004>.
3. Jacob L, Speed TP. The healthy ageing gene expression signature for Alzheimer's disease diagnosis: a random sampling perspective. *Genome Biol.* 2018;19:97 [cited 2019 Jun 9]. Available from: <https://genomebiology.biomedcentral.com/articles/10.1186/s13059-018-1481-6>.

4. König IR, Malley JD, Weimar C, Diener H-C, Ziegler A. Practical experiences on the necessity of external validation. *Stat Med.* 2007;26:5499–511 [cited 2019 May 3]. Available from: <http://doi.wiley.com/10.1002/sim.3069>.
5. Vos SJB, Verhey F, Frölich L, Kornhuber J, Wiltfang J, Maier W, et al. Prevalence and prognosis of Alzheimer's disease at the mild cognitive impairment stage. *Brain.* 2015;138:1327–38 [cited 2018 Sep 4]. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25693589>.
6. Zhang B, Gaiteri C, Bodea L-GG, Wang Z, McElwee J, Podtelezhnikov AA, et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell.* 2013;153:707–20 [cited 2017 Feb 3]. Available from: <https://doi.org/10.1016/j.cell.2013.03.030>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

