

COMMENTARY

A commentary on evaluation of the evenness score in next-generation sequencing

Paul Horton

Journal of Human Genetics (2016) 61, 575; doi:10.1038/jhg.2016.29; published online 14 April 2016

The development and adoption of appropriate performance measures is critical to advancing science and engineering, because such measures allow the community to rationally choose the (current) best solution to a problem among numerous alternatives. In that sense, measures which help us make decisions by providing a single number criterion are especially useful. For example, the unfortunately named ‘area under the receiver operator characteristic’ (AUROC) is a classic measure that provides a well-principled way to rank the performance of binary classification; interpretable as the probability that a randomly chosen positive example achieves a higher score than a randomly chosen negative example.¹ Before the widespread adoption of AUROC it was often difficult to compare methods from different publications—how does one decide if a method reporting 50% sensitivity with 80% specificity is better or worse than one reporting 30% sensitivity with 95% specificity? Another popular measure which reduces a tradeoff into a single number is the so-called *h*-index,² which measures the scientific output of a researcher as the number of their published papers with at least *h* citations.

In this issue, Oexle³ presents a mathematical treatment of the ‘evenness’ score introduced by Mokry *et al.*⁴ as a single

number measure of the uniformity of DNA sequencing coverage. This measure provides an attractive alternative to simply listing the percentage of sites with coverage above an arbitrary threshold (for example, 30×), and has been used in some recent publications.^{5,6} However, to be widely adopted, a measure should not only be informative, but like the AUROC and *h*-index measures mentioned above also easy to compute and interpret. Oexle’s³ exposition goes a long way toward making this the case for the evenness score by (1) developing a simple and efficient formula for its computation and (2) giving mathematical insight regarding its behavior vis-à-vis the coefficient of variation, a more conventional statistical measure of the width of a distribution. Although his manuscript is by nature mathematical, Oexle’s³ main conclusions are easily understandable without diving deep into the derivations. Thus, regardless of whether the sequencing community ultimately chooses to widely adopt the evenness score or perhaps switches to the coefficient of variation or related measures; the choice can now be a well-informed one.

CONFLICT OF INTEREST

The author declares no conflict of interest.

- 1 Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**, 29–36 (1982).
- 2 Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl Acad. Sci. USA* **102**, 16569–16572 (2005).
- 3 Oexle, K. Evaluation of the evenness score in next-generation sequencing. *J. Hum. Genet.* **61**, 627–632 (2016).
- 4 Mokry, M., Feitsma, H., Nijman, I. J., de Bruijn, E., van der Zaag, P. J., Guryev, V. *et al.* Accurate SNP and mutation detection by targeted custom microarray-based genomic enrichment of short-fragment sequencing libraries. *Nucleic Acid Res.* **38**, e116 (2010).
- 5 Lelieveld, S. H., Spielmann, M., Mundlos, S., Veltman, J. A. & Gilissen, C. Comparison of exome and genome sequencing technologies for the complete capture of protein-coding regions. *Hum. Mutat.* **36**, 815–822 (2015).
- 6 García-García, G., Baux, D., Faugère, V., Moclyn, M., Koenig, M., Clausters, M. *et al.* Assessment of the latest NGS enrichment capture methods in clinical context. *Sci. Rep.* **6**, 20948 (2016).



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>