ETH zürich

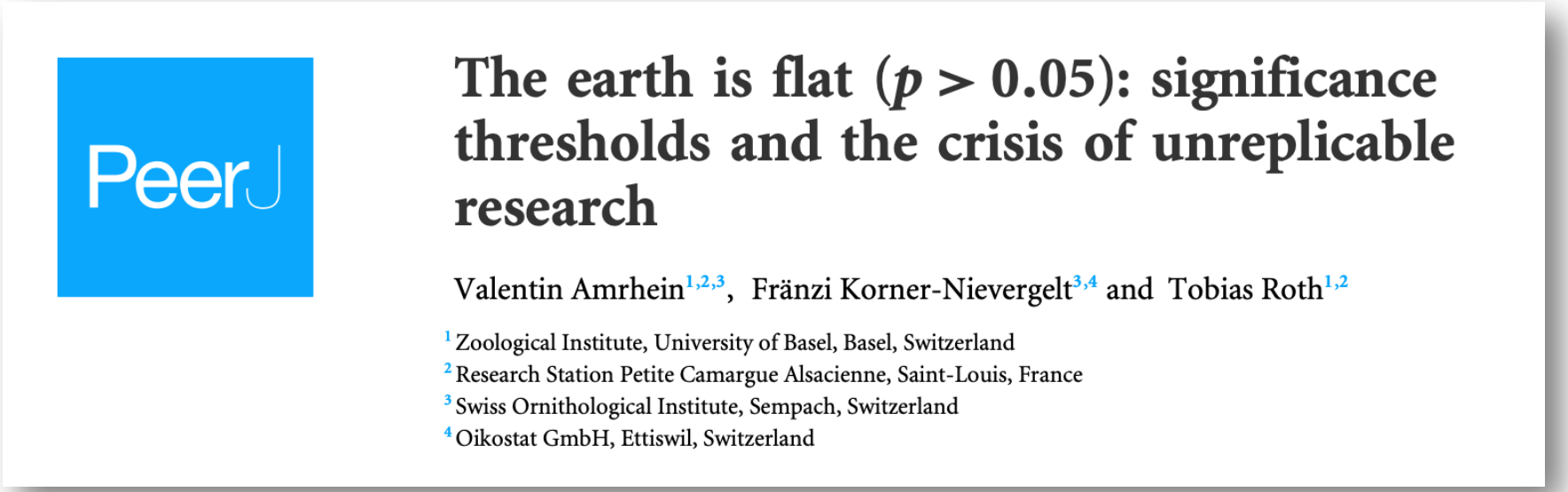
nexus
Personalized Health Technologies

Computational Reproducibility – a tale about tools and attitudes

Basel Biometric Society
12th March 2024

Daniel Stekhoven
ETH Zurich

Definitions



PeerJ

The earth is flat ($p > 0.05$): significance thresholds and the crisis of unreplicable research

Valentin Amrhein[1,2,3], Fränzi Korner-Nievergelt[3,4] and Tobias Roth[1,2]

[1] Zoological Institute, University of Basel, Basel, Switzerland
[2] Research Station Petite Camargue Alsacienne, Saint-Louis, France
[3] Swiss Ornithological Institute, Sempach, Switzerland
[4] Oikostat GmbH, Ettiswil, Switzerland

1. Amrhein, V et al. (2017). PeerJ 5:e3544; DOI: 10.7717/peerj.3544
2. Goodman SN et al. Sci. Transl. Med.8,341ps12-341ps12(2016). DOI:10.1126/scitranslmed.aaf5027

Transparency

scientific reports

Check for updates

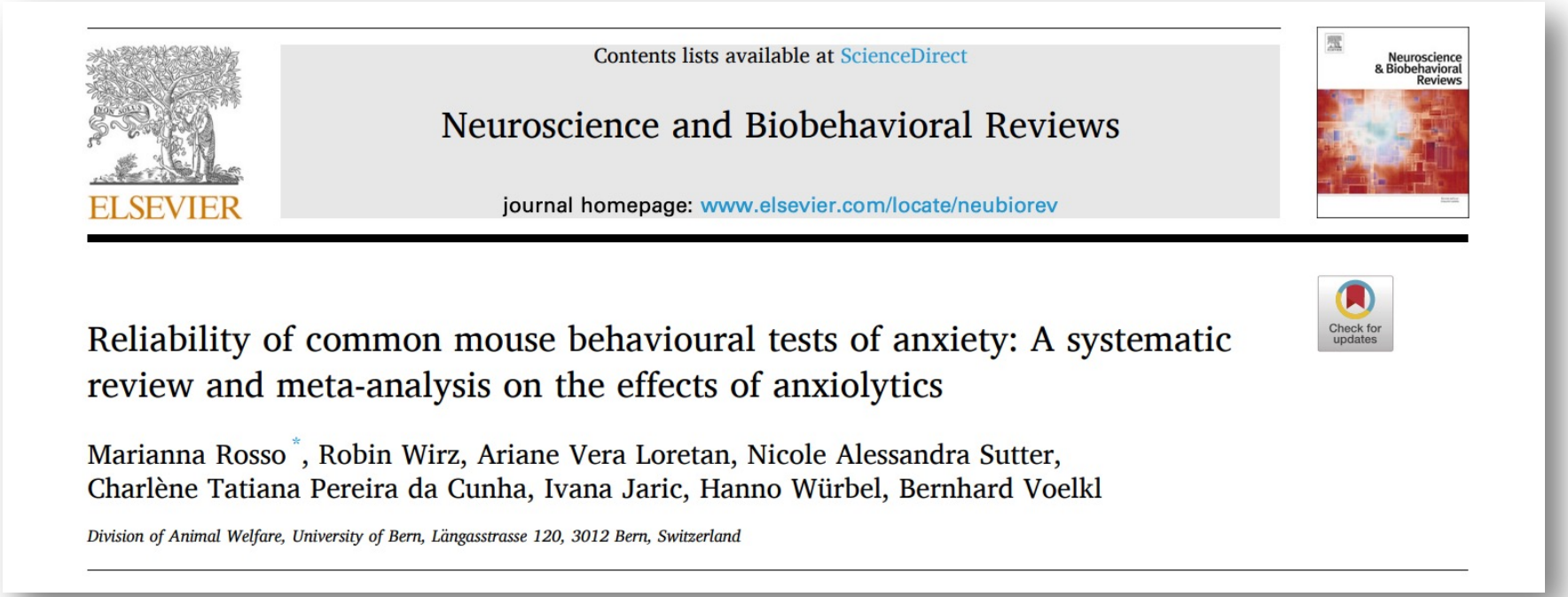
OPEN

Insufficient transparency of statistical reporting in preclinical research: a scoping review

Romain-Daniel Gosselin

1. Gosselin, RD. Sci Rep 11, 3335 (2021). DOI: 10.1038/s41598-021-83006-5

Systematic Reviews



Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

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

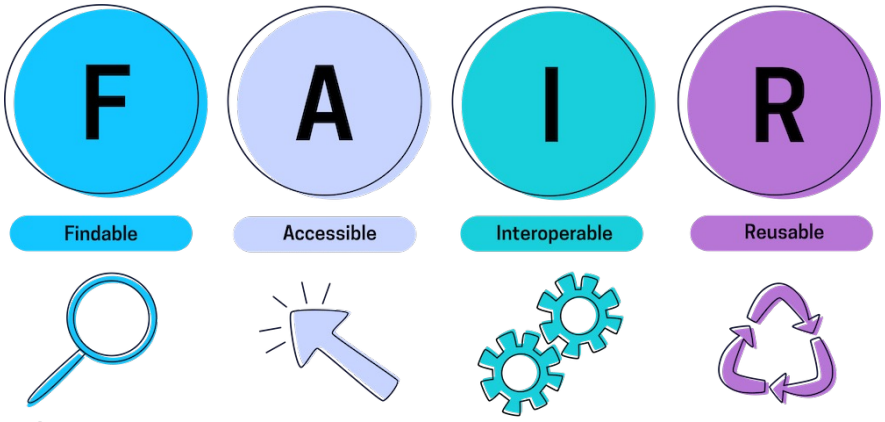
Reliability of common mouse behavioural tests of anxiety: A systematic review and meta-analysis on the effects of anxiolytics

Marianna Rosso *, Robin Wirz, Ariane Vera Loretan, Nicole Alessandra Sutter, Charlène Tatiana Pereira da Cunha, Ivana Jaric, Hanno Würbel, Bernhard Voelkl

Division of Animal Welfare, University of Bern, Längasstrasse 120, 3012 Bern, Switzerland

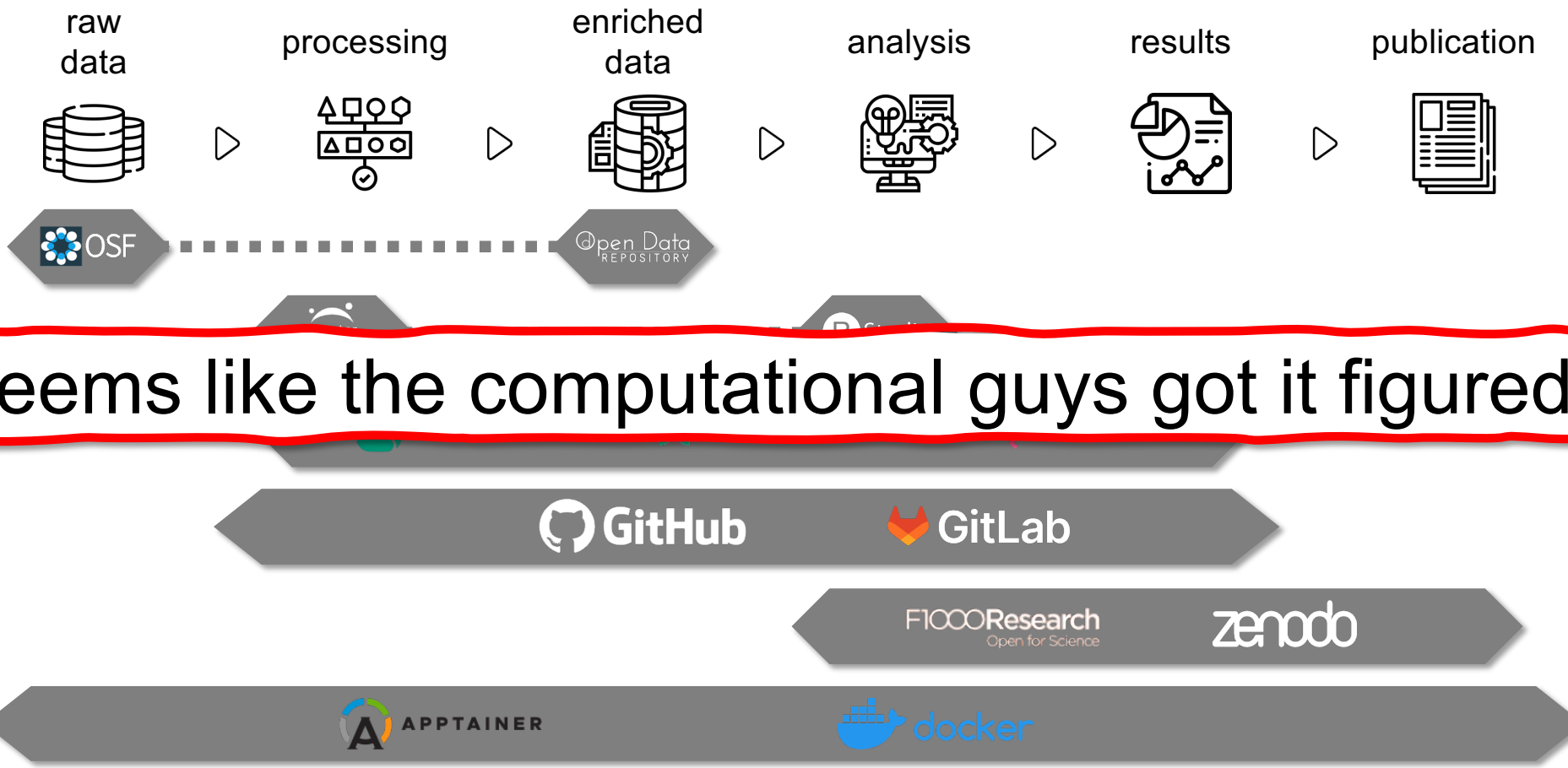
1. Rosso, M et al. J Neurosci Bio Rev 143, 104928 (2022), DOI: 10.1016/j.neubiorev.2022.104928

The role of tools in computational reproducibility



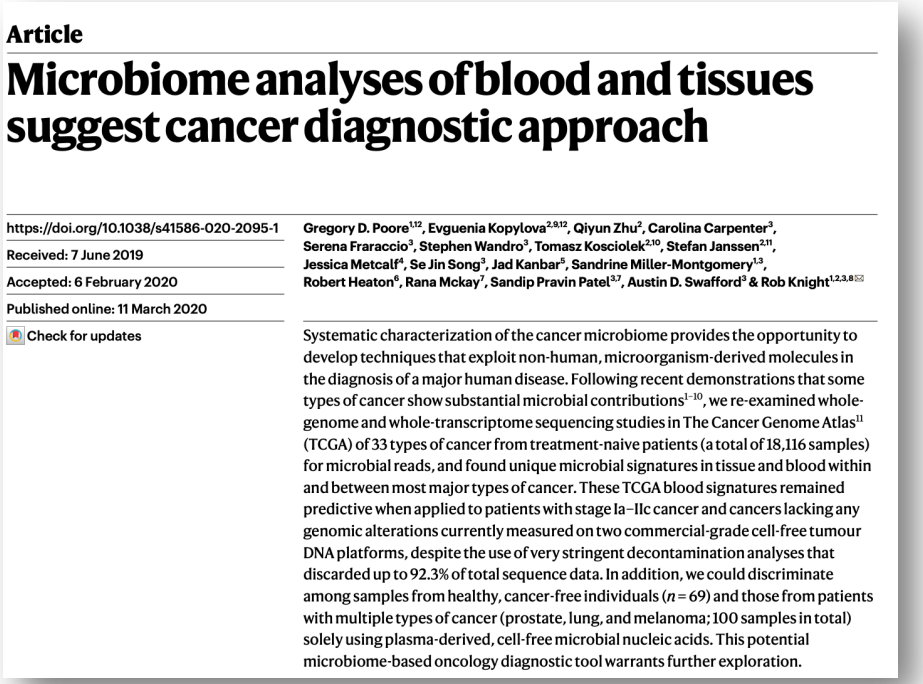
1. Making data FAIR, https://flur.ee/fluree-blog/making-data-fair/
2. Smoby Bob the Builder 3 Tools Set 360137

Essential tools



1. Icons from Freepik Company S.L., https://www.flaticon.com/

Well, no...



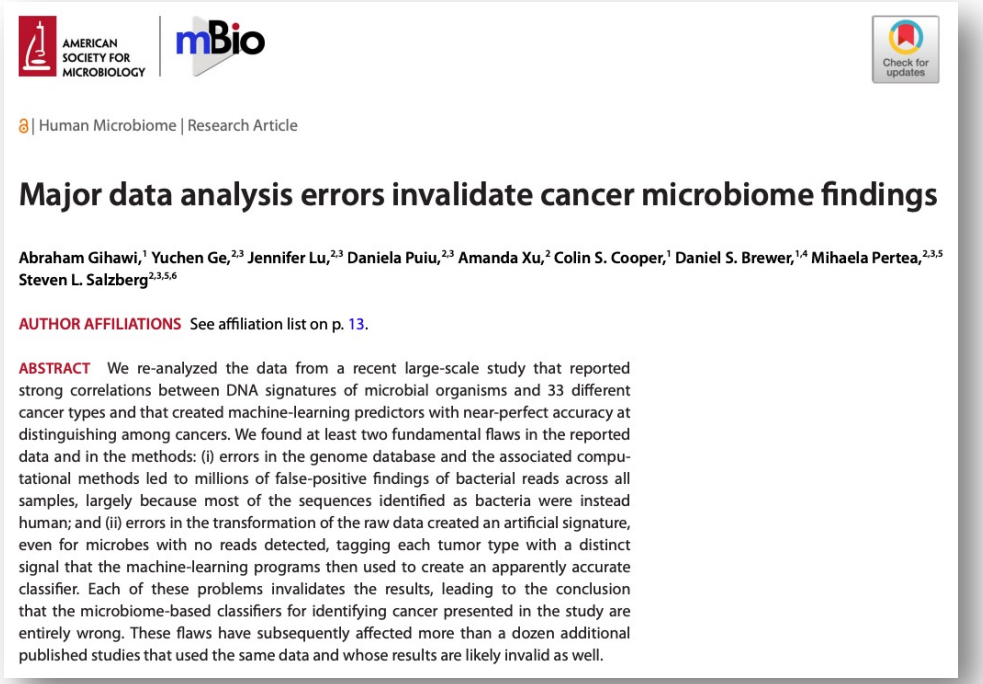
Article

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

https://doi.org/10.1038/s41586-020-2095-1
Received: 7 June 2019
Accepted: 6 February 2020
Published online: 11 March 2020
Check for updates

Gregory D. Poore[1,12], Evguenia Kopylova[2,9,12], Qiyun Zhu[2], Carolina Carpenter[3], Serena Fraraccio[3], Stephen Wandro[3], Tomasz Kosciolek[2,10], Stefan Janssen[2,11], Jessica Metcalf[4], Se Jin Song[3], Jad Kanbar[5], Sandrine Miller-Montgomery[1,3], Robert Heaton[6], Rana Mckay[7], Sandip Pravin Patel[3,7], Austin D. Swafford[3] & Rob Knight[1,2,3,8]

Systematic characterization of the cancer microbiome provides the opportunity to develop techniques that exploit non-human, microorganism-derived molecules in the diagnosis of a major human disease. Following recent demonstrations that some types of cancer show substantial microbial contributions[1–10], we re-examined whole-genome and whole-transcriptome sequencing studies in The Cancer Genome Atlas[11] (TCGA) of 33 types of cancer from treatment-naive patients (a total of 18,116 samples) for microbial reads, and found unique microbial signatures in tissue and blood within and between most major types of cancer. These TCGA blood signatures remained predictive when applied to patients with stage Ia–IIc cancer and cancers lacking any genomic alterations currently measured on two commercial-grade cell-free tumour DNA platforms, despite the use of very stringent decontamination analyses that discarded up to 92.3% of total sequence data. In addition, we could discriminate among samples from healthy, cancer-free individuals ($n = 69$) and those from patients with multiple types of cancer (prostate, lung, and melanoma; 100 samples in total) solely using plasma-derived, cell-free microbial nucleic acids. This potential microbiome-based oncology diagnostic tool warrants further exploration.



AMERICAN SOCIETY FOR MICROBIOLOGY | mBio
Check for updates
| Human Microbiome | Research Article

Major data analysis errors invalidate cancer microbiome findings

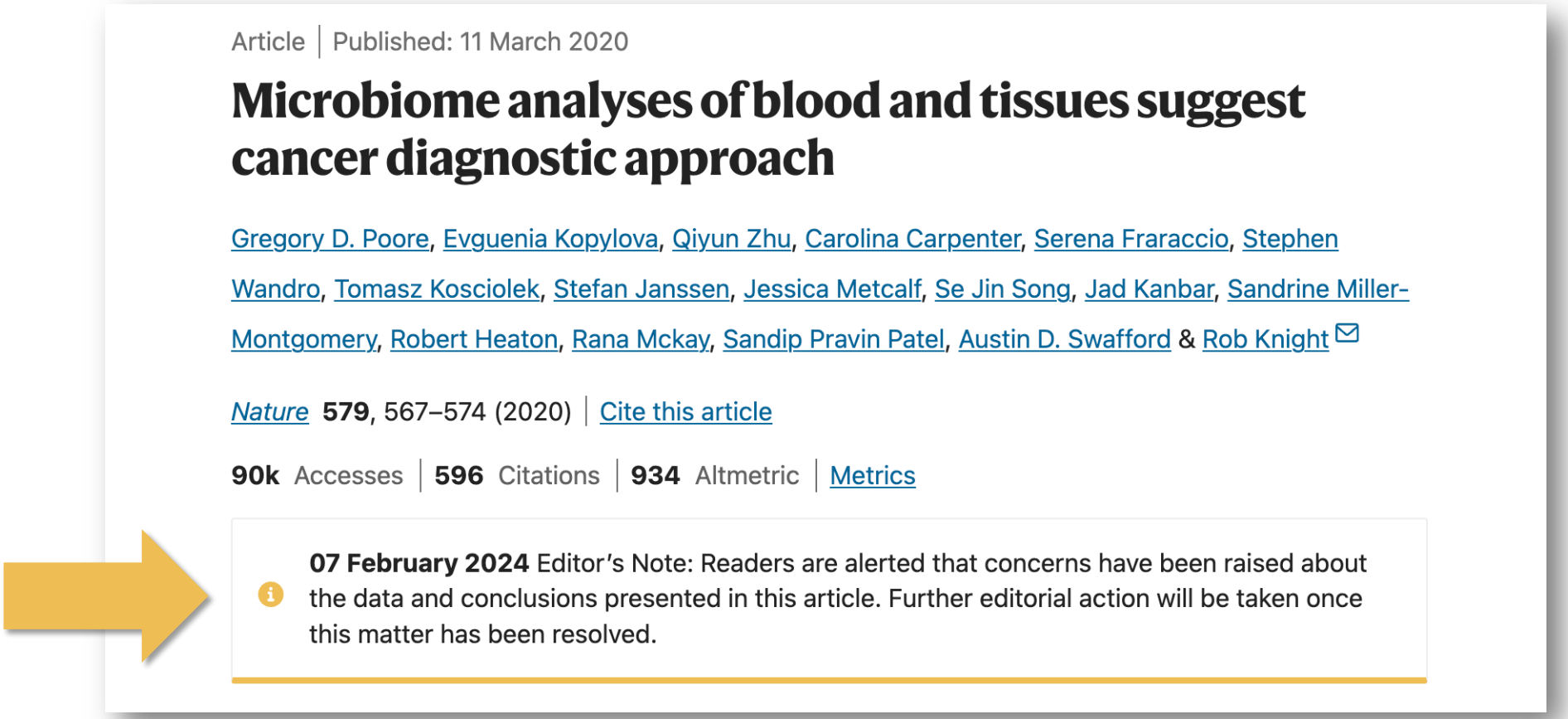
Abraham Gihawi,[1] Yuchen Ge,[2,3] Jennifer Lu,[2,3] Daniela Puiu,[2,3] Amanda Xu,[2] Colin S. Cooper,[1] Daniel S. Brewer,[1,4] Mihaela Pertea,[2,3,5] Steven L. Salzberg[2,3,5,6]

AUTHOR AFFILIATIONS See affiliation list on p. 13.

ABSTRACT We re-analyzed the data from a recent large-scale study that reported strong correlations between DNA signatures of microbial organisms and 33 different cancer types and that created machine-learning predictors with near-perfect accuracy at distinguishing among cancers. We found at least two fundamental flaws in the reported data and in the methods: (i) errors in the genome database and the associated computational methods led to millions of false-positive findings of bacterial reads across all samples, largely because most of the sequences identified as bacteria were instead human; and (ii) errors in the transformation of the raw data created an artificial signature, even for microbes with no reads detected, tagging each tumor type with a distinct signal that the machine-learning programs then used to create an apparently accurate classifier. Each of these problems invalidates the results, leading to the conclusion that the microbiome-based classifiers for identifying cancer presented in the study are entirely wrong. These flaws have subsequently affected more than a dozen additional published studies that used the same data and whose results are likely invalid as well.

1. Poore, G.D. et al. Nature 579, 567–574 (2020). https://doi.org/10.1038/s41586-020-2095-1
2. Giwahi A. et al. American Society for Microbiology (2023) 14 (5) e01607-23, https://doi.org/10.1128/mbio.01607-23

...the tools worked!



Article | Published: 11 March 2020

Microbiome analyses of blood and tissues suggest cancer diagnostic approach

Gregory D. Poore, Evguenia Kopylova, Qiyun Zhu, Carolina Carpenter, Serena Fraraccio, Stephen Wandro, Tomasz Kosciolek, Stefan Janssen, Jessica Metcalf, Se Jin Song, Jad Kanbar, Sandrine Miller-Montgomery, Robert Heaton, Rana Mckay, Sandip Pravin Patel, Austin D. Swafford & Rob Knight

Nature **579**, 567–574 (2020) | Cite this article

90k Accesses | **596** Citations | **934** Altmetric | Metrics

07 February 2024 Editor's Note: Readers are alerted that concerns have been raised about the data and conclusions presented in this article. Further editorial action will be taken once this matter has been resolved.

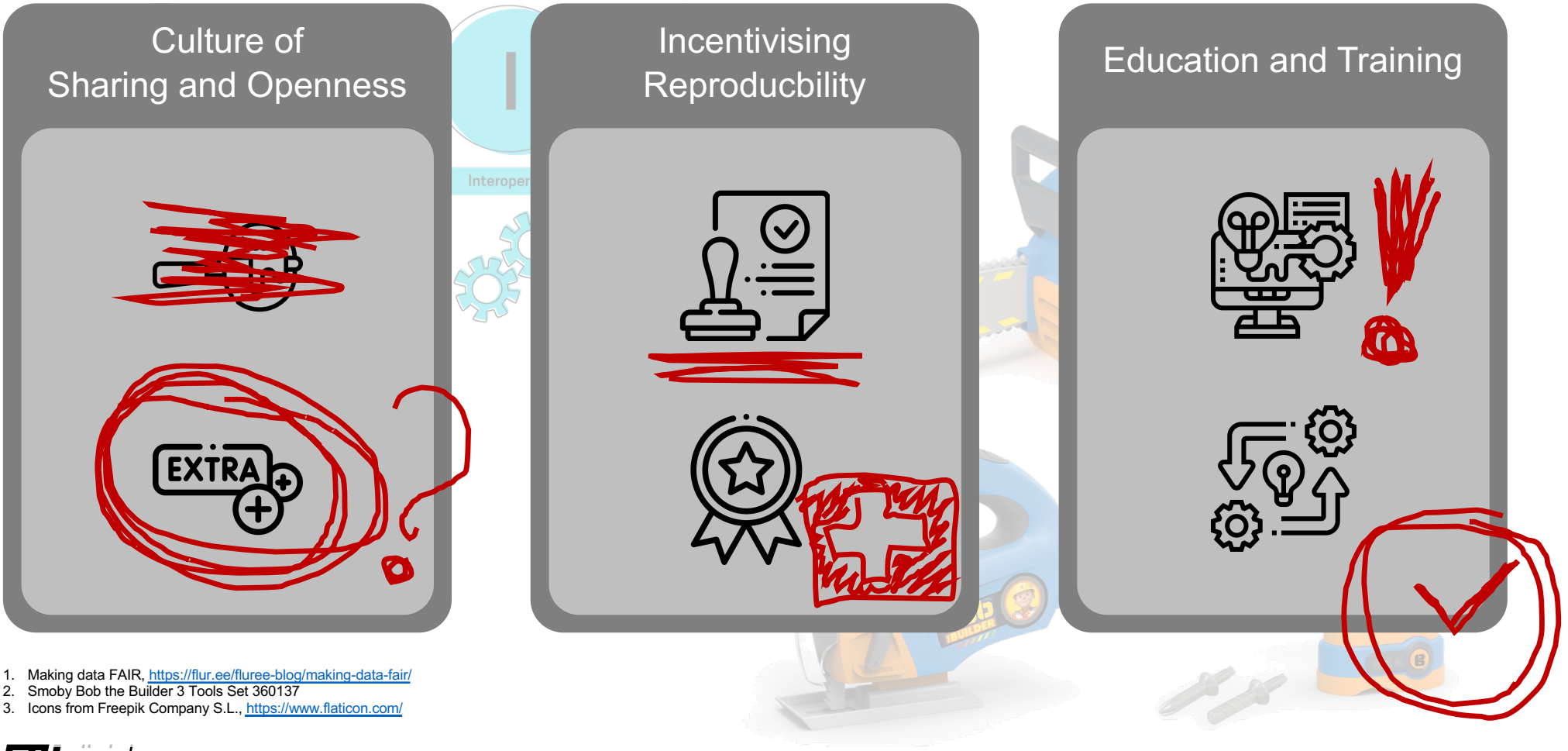
1. Poore, G.D. et al. Nature 579, 567–574 (2020). https://doi.org/10.1038/s41586-020-2095-1

...almost at least.

> For those interested in the code, all of our methods are based on free, open-source software. We ran our analysis with default KrakenUniq [3] parameters against a curated Kraken database [4] extended from EuPathDB [5], so it should be easy for others to replicate our results by following our Methods as described in the preprint.

1. Yu G on Github, https://github.com/yge15/Cancer_Microbiome_Reanalyzed

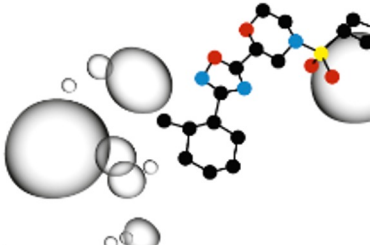
Beyond tools – the researcher's attitude



1. Making data FAIR, https://flur.ee/fluree-blog/making-data-fair/
2. Smoby Bob the Builder 3 Tools Set 360137
3. Icons from Freepik Company S.L., https://www.flaticon.com/

1. Vermutlich inspiriert durch Wiener Caricaturen, 24.05.1885, S. 2, Sp. 2; Puck: Illustrirtes humoristisches Wochenblatt, 1885, S. 630

Daniel Stekhoven
Director
stekhoven@nexus.ethz.ch

ETH Zürich
NEXUS Personalized Health Technologies
Wagistrasse 18
8952 Zürich (Schlieren)

www.nexus.ethz.ch