MATTERS ARISING   OPEN

Reply to: Addressing racial and ethnic disparities in AACR project GENIE

Alexander T. M. Cheung [1,2], Andrzej Niemierko[3], Eliezer Van Allen[2,4], Neha Vapiwala[5] and Sophia C. Kamran [2,3,6✉]

We thank Sweeney et al. for their comments regarding our recent article evaluating the representation of racial/ethnic minority groups in the AACR Project GENIE for precision oncology research. Our findings[1] suggest that this real-world biorepository—while providing a wealth of clinical-genomic data—may not accurately reflect the actual distribution of various cancer types in the general United States patient population.

Sweeney et al. disagree with our "emphasis on powering comparisons for a Cohen's h of less than 0.2." However, these benchmark values of 0.2, 0.5, and 0.8 for "small," "medium," and "large" effect sizes, respectively, are arbitrary and should not be interpreted rigidly[2], with different thresholds depending on the scientific study[3]; Cohen in fact warns against inflexibility with respect to these values[4,5]. Sawilowsky expanded the benchmarks and definitions to range from 0.01 (very small) to 2.0 (huge), based on updated research findings in the applied literature[4]. We explicitly defined a small effect size as a Cohen's h less than or equal to 0.2 and used this definition to identify, in comparison to white samples, which racial/ethnic groups within the GENIE database have insufficient sample sizes for studying small genomic differences. Moreover, since genomic differences between cancer phenotypes often occur on a long-tail distribution[6], we anticipated that researchers might be interested in even smaller and subtler effect sizes. Regardless of the 0.2 vs. <0.2 cutoff, there would be no change expected in the overarching conclusions, which are (i) that the dataset is underpowered for comparisons of Black, Asian, and Hispanic primary and metastatic tumor samples versus white samples in two of five common cancers (prostate and pancreatic), and (ii) the dataset is also underpowered for comparisons of Native American and Pacific Islander samples versus white samples for all five common cancers that we evaluated in the primary and metastatic setting. Furthermore, for less common cancers, the lack of adequate representation within AACR Genie is more striking, as demonstrated in a recent analysis of pancreatic neuroendocrine tumors[7].

We agree that it's important to note that GENIE is a dynamic database and that analyses in our study were based on the version cited (v9.1-public release, January 2021). We were pleased to note that the latest release of GENIE (v13.0-public) contains more diverse and representative patient samples. However, our findings remain relevant and novel for the literature for several reasons. First, our study provides a snapshot of the GENIE database at a certain point in time, and our expectation is that future researchers can likewise audit the data and assess its evolution over successive iterations. Second, while the repository changes every six months, it is unlikely that all groups utilizing GENIE will

be able to download, process, and analyze the data—not to mention the time required for peer review and actual publication—within that same timeframe. Indeed, we felt it was important to perform this analysis given several publications that used earlier versions of the GENIE database to draw conclusions about race and cancer genomics in poorly represented populations[8–10] (occasionally with conflicting results)[11,12], while also demonstrating its strengths in other patient subgroups and cancers for researchers interested in exploring questions in those contexts. We hope our findings provide a frame of reference for readers interpreting and evaluating published studies that employed this version and also encourage continued contributions from centers to help strengthen the registry in areas of need.

Lastly, we agree with Sweeney et al.'s point that biorepositories "*may not represent the broader population of patients that are diagnosed and treated for cancer*" due to barriers like institutional and participation biases. This also pertains to the GENIE registry, which currently reflects patient populations and practice patterns only at participating institutions. In fact, this was the impetus for our manuscript. We commend AACR for their dedicated efforts to improve this representation through a variety of initiatives and look forward to seeing this already valuable public biorepository grow into a more diverse and representative database for future generations of scientists, clinicians, and, most of all, patients.

DATA AVAILABILITY

The data underlying the original article[1] are available via the publicly available Project GENIE at genie.cbioportal.com, with additional information regarding its initial release found here: https://doi.org/10.1158/2159-8290.CD-17-0151.

REFERENCES

1. Cheung, A. T. M. et al. Racial and ethnic disparities in a real-world precision oncology data registry. *NPJ Precis. Oncol.* **7**, 7 (2023).
2. Thompson, B. Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychol. Sch.* **44**, 423–432 (2007).
3. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* **4**, 863 (2013).
4. Sawilowsky, S. S. New effect size rules of thumb. *J. Mod. Appl. Stat. Methods* **8**, 597–599 (2009).
5. Cohen, J. *Statistical power analysis for the behavioral sciences.* 2nd edn (Erlbaum, 1988).

[1]NYU Grossman School of Medicine, New York, NY, USA. [2]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [3]Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [4]Dana-Farber Cancer Institute, Boston, MA, USA. [5]Department of Radiation Oncology, Hospital of the University of Pennsylvania, Perelman School of Medicine, Philadelphia, PA, USA. [6]Twitter: @sophia_kamran ✉email: skamran@mgh.harvard.edu

6. Armenia, J. et al. The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).

7. Herring, B. R. et al. Under-representation of racial groups in genomics studies of gastroenteropancreatic neuroendocrine neoplasms. *Cancer Res. Commun.* **2**, 1162–1173 (2022).

8. Goel, N. et al. Racial differences in genomic profiles of breast cancer. *JAMA Netw. Open* **5**, e220573 (2022).

9. Mahal, B. A. et al. Racial differences in genomic profiling of prostate cancer. *N. Engl. J. Med.* **383**, 1083–1085 (2020).

10. Nassar, A. H., Adib, E. & Kwiatkowski, D. J. Distribution of KRAS (G12C) somatic mutations across race, sex, and cancer type. *N. Engl. J. Med.* **384**, 185–187 (2021).

11. Kamran, S. C. et al. Tumor mutations across racial groups in a real-world data registry. *JCO Precis. Oncol.* **5**, 1654–1658 (2021).

12. Schumacher, F. R. et al. Race and genetic alterations in prostate cancer. *JCO Precis. Oncol.* **5**, PO.21.00324 (2021).

AUTHOR CONTRIBUTIONS

Concept and design: All authors. Acquisition, analysis, or interpretation of data: All authors. Drafting of the manuscript: All authors. Critical revision of the manuscript for important intellectual content: All authors. Final approval of the completed version of the manuscript: All authors. Statistical analysis: A.N. Administrative, technical, or material support: E.V.A., N.V., S.C.K. Supervision: N.V., E.V.A., S.C.K. Accountability for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved: All authors

COMPETING INTERESTS

Dr. Kamran reported having a spouse who is employed by Sanofi Genzyme and receipt of grants from the Prostate Cancer Foundation outside the submitted work. Dr. Van Allen reported advisory and consulting work for Tango Therapeutics, Genome Medical, Invitae, Enara Bio, Janssen, Manifold Bio, and Monte Rosa; he reported research support from Novartis, BMS; he reported equity in Tango Therapeutics, Genome Medical, Syapse, Enara Bio, Manifold Bio, Microsoft, and Monte Rosa; he received travel reimbursement from Roche/Genentech; and he filed institutional patents on chromatin mutations and immunotherapy response and methods for clinical interpretation outside the submitted work. No other disclosures were reported.

ADDITIONAL INFORMATION

Correspondence and requests for materials should be addressed to Sophia C. Kamran.

Reprints and permission information is available at http://www.nature.com/reprints

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