



**EDITORIAL**

# Are we inadvertently widening the disparity gap in pursuit of precision oncology?

*British Journal of Cancer* (2018) 119:783–784; <https://doi.org/10.1038/s41416-018-0223-6>

**SUMMARY**

The Cancer Genome Atlas (TCGA) is an excellent tool to interrogate over 30 different cancer types, containing DNA, RNA and protein from over 11,000 patients. Nevertheless, caution should be used when interpreting results from these studies, with careful consideration of their generalisability to the population, especially racial minority populations and the elderly.

**THE CANCER GENOME ATLAS**

The progress being made in oncology towards true personalised medicine is remarkable. Just a few decades ago the thought of performing whole exome sequencing, RNA sequencing, methylome analyses and broad coverage proteomic analysis on thousands of tumour samples would be beyond a scientist's or clinician's dream. Through technological advancements and backing from the US National Institute of Health, The Cancer Genome Atlas (TCGA) project was born. The TCGA has become a high-quality reference guide to the genetic fingerprint of most cancers, and targetable mutations identified in TCGA have become the focus of clinical exploration. With DNA, RNA and protein analysed from over 11,000 tumours across 33 cancer types, an unparalleled amount of data has been gathered that the global community is still unravelling. However, the TCGA will only be useful globally if the catalogued sequences are representative. A recent article published in this issue of BJC further explores this issue.<sup>1</sup>

**GENERALISABILITY OF TCGA**

Much of the focus of TCGA, rightfully so, has been on genomic alterations across a tumour's genetic code. The associated linked variables, such as age, race, gender and stage of diagnosis are often a dismissed co-variable when performing genomic analyses. Yet, it is clear that patients of different age or race have variable prognosis and tumour biology. Classical examples of this are the extreme difference in prognosis of thyroid or endometrial cancer by age, or the difference in prevalence of *EGFR* mutations between Asians and Whites.<sup>2</sup> Therefore, we must now question how well the TCGA represents patients with cancer within the USA, let alone the global community. If we pursue medical advancements based on genomic sequencing efforts that are not representative of the population, then we have the potential to widen many of the already pervasive disparities in oncology.

To investigate this, multiple studies have been performed to assess how representative the TCGA dataset is compared to the USA cancer population, as inferred from the Surveillance, Epidemiology, and End Results (SEER) database. Two such studies were specifically dedicated to disparities in race and age, respectively.<sup>3,4</sup> The first study, from our group and published in

2016, shed light on the low absolute number of racial minorities within TCGA.<sup>3</sup> Of the 10 tumour types studied comprising 5729 samples, 77% ( $n = 4389$ ) were white, 12% ( $n = 660$ ) were black, 3% ( $n = 173$ ) were Asian, 3% ( $n = 149$ ) were Hispanic, and less than 0.5% combined were from patients of Native Hawaiian, Pacific Islander, Alaskan Native or American Indian descent. Furthermore, even if the relative distribution by race was similar to the actual population, usually sequencing fewer than 50 Black patients per tumour type has powerful implications to the generalisability of results from the TCGA. Spratt et al.<sup>3</sup> performed a series of power calculations to demonstrate that there is currently insufficient power to detect even a 10% mutational frequency in Black patients across many tumour types within TCGA. This means that there could easily be a relatively frequent mutation present in Black patients that has yet to be detected, given the limited sample size.

The second study, also by our group, investigated disparities within the elderly population and their participation within TCGA.<sup>4</sup> This study showed that patients aged 80 to 99 years old were underrepresented in all 9 cancer types assessed, with a median TCGA underrepresentation of 167% compared to SEER. Similar to the racial disparities study, no cancer type had enough elderly patients to detect even a 10% mutational frequency. This is not likely to be the fault of TCGA, as elderly patients, especially over 80 years old, rarely undergo radical surgery and thus there are fewer tissues available for sequencing. However, given the close correlation between age and oncological outcomes in a multitude of tumour types, we are probably missing valuable biological insights by the omission of these patients from sequencing efforts. The size of the elderly population is rapidly growing, and the relevance of understanding the biology of their cancer is of increasing importance. Within this study by Wahl et al.<sup>4</sup> the investigators specifically looked at gene expression in young versus older men with prostate cancer, and identified multiple differentially expressed genes involved with androgen signalling and DNA repair, further supporting the influence age has on cancer biology.<sup>4</sup>

In the current issue, Wang et al.<sup>1</sup> report a well-performed analysis of all cancer types within TCGA. They have completed the most comprehensive comparison of patient characteristics within TCGA and SEER, not only for race and age, but also for gender, stage, and survival outcomes. They confirmed the previous findings, and expand upon them to demonstrate that 20 cancer types within TCGA contain significantly younger patients than SEER, 13 had disparities in relative racial distributions, and 25 cancer types had different stage distributions. Furthermore, the mean 1-year survival estimate for 27 of the 33 tumour types within TCGA were longer than SEER estimates. These studies clearly demonstrate important differences between patients contained within TCGA and the general cancer population.

Received: 12 July 2018 Revised: 18 July 2018 Accepted: 18 July 2018  
Published online: 21 August 2018

## BEYOND TCGA

The overwhelming genomic data gathered to date has been almost exclusively on individuals of European ancestry, with an estimated 10:1 ratio of genome-wide association studies (GWAS) from European ancestry vs all other groups combined. For example, one study by Need et al highlighted this point by showing that over 1.5 million patients of European ancestry have been studied in GWAS cohorts, compared with only ~1000 Hispanic patients (0.07%).<sup>5</sup> This has been due in part to Europe and the United States providing a large majority of genomic samples and data, as well as a low percent of minorities taking part in genomic sequencing.

This is important, as there has been an increasing trend towards somatic tumour profiling without combined germline testing. Garofalo et al. nicely demonstrated that in studies without matched germline testing, there was a higher percentage for non-white patients to have false positive results.<sup>6</sup> The higher false positive variants in non-white patients highlights the limitations of utilising genomic reference databases, as these minorities are often not well represented. Thus, the clinical impact of disparities in genomic sequencing can impact real-world clinical results from tumour profiling.

## IMPLICATIONS

The implications and downstream impact of these findings is unclear. It is plausible that, in an effort to further develop precision medicine through advanced technology and big data, the scientific community has inadvertently widened an already existent disparity for racial minorities and the elderly. Similar to the low enrolment of minorities and the elderly within clinical trials, dedicated efforts have improved clinical trial participation.<sup>7</sup> This is critical, because just this year at the Annual Society of Clinical Oncology 2018 meeting, there were two studies that compared the results of Black versus White men with advanced prostate cancer enrolled on various clinical trials, and showed equivalent, if not potentially superior outcomes for the Black men.<sup>8,9</sup> It is possible that with further investigation of the populations under-represented within TCGA that additional clinically meaningful findings could be generated.

Going forward, dedicated initiatives will be required to increase the enrolment of genomic sequencing of racial minorities, especially

when there are known disparities in clinical outcomes by race, and the elderly. Given the ability to now perform next-generation sequencing on limited amounts of biopsy tissue, this should aid in a broader ability to capture these under-represented populations. New discoveries often lead to more questions and challenges, and we must unite to ensure these benefits reach everyone.

## ADDITIONAL INFORMATION

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

**Note:** This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Daniel E. Spratt<sup>1</sup>

<sup>1</sup>*Department of Radiation Oncology, University of Michigan, Ann Arbor, MI, USA*

*Correspondence: Daniel E. Spratt (sprattda@med.umich.edu)*

## REFERENCES

1. Wang, X. et al. Characteristics of The Cancer Genome Atlas cases relative to U.S. general population cancer cases. *Br. J. Cancer* (2018).
2. Shi, Y. et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). *J. Thorac. Oncol.* **9**, 154–162 (2014).
3. Spratt, D. E. et al. Racial/ethnic disparities in genomic sequencing. *JAMA Oncol.* **2**, 1070–1074 (2016).
4. Wahl, D. R. et al. Pan-cancer analysis of genomic sequencing among the elderly. *Int. J. Radiat. Oncol. Biol. Phys.* **98**, 726–732 (2017).
5. Need, A. C. & Goldstein, D. B. Next generation disparities in human genomics: concerns and remedies. *Trends Genet.* **25**, 489–494 (2009).
6. Garofalo, A. et al. The impact of tumor profiling approaches and genomic data strategies for cancer precision medicine. *Genome Med.* **8**, 79 (2016).
7. Spratt, D. E. & Osborne, J. R. Disparities in Castration-resistant prostate cancer trials. *J. Clin. Oncol.* **33**, 1101–1103 (2015). JCO. 2014.2058. 1751.
8. George, D. J. et al. Abi Race: a prospective, multicenter study of black (B) and white (W) patients (pts) with metastatic castrate resistant prostate cancer (mCRPC) treated with abiraterone acetate and prednisone (AAP). *ASCO 2018 Abstract LBA5009* (2018).
9. Halabi, S. et al. Overall survival between African-American (AA) and Caucasian (C) men with metastatic castration-resistant prostate cancer (mCRPC). *ASCO 2018 Abstract LBA5005* (2018).