Racial disparities in tumor profiling testing inferred from continental genetic ancestry determination of 100,000 cancer patients

Francisco M. De La Vega, Yannick Pouliot, Brooke Rhead, and Justin Guinney
Tempus Labs, Inc., Chicago, IL

INTRODUCTION

Racial and ethnic disparities in the incidence of cancer and treatment outcomes are well documented. This is credited to a variety of factors which include social, structural, and access to care inequities, as well as biological differences that may correlate with race/ethnicity. We aimed to measure racial differences in testing for cancer therapy decision support using real-world data (RWD) from 100,000 de-identified patients who underwent tumor genomic profiling with the Tempus xT next-generation sequencing assay (targeting 648 genes).

METHODS

We inferred genetic ancestry from approximately 100,000 de-identified records from cancer patients of diverse histology who underwent tumor genomic profiling with the 648-gene Tempus xT next-generation sequencing (NGS) assay. We used 654 ancestry informative markers selected to overlap the target regions of the assay to infer global ancestry proportions at the continental level: Africa (AFR), Americas (AMR), Europe (EUR), East Asia (EAS), and South Asia (SAS).

Using a heuristic that combines continental ancestry proportions, we imputed four race/ethnicity labels: Asian, Non-Hispanic (NH) Black, Hispanic/Latino/Native American (NA), and NH White, showing a classification error of <2% evaluated in patients for which race/ethnicity was provided. We were unable to classify 3% of patients due to complex ancestry admixture and these were excluded.

Inclusion Criteria:
- Cancer types with at least 1,000 patients
- Cancer types with recorded incidence at the SEER database

Statistical Analysis:
- The distribution of race/ethnicity by cancer type was compared to the expected distribution under the null hypothesis of no relationship by the $\chi 2$ test of independence.
- Significantly over/under-represented race/ethnicities in each cancer were compared to ranks of race/ethnicities in SEER at the national level.

SUMMARY

- Our results show that genetic ancestry inference on genomic data from tumor profiling can partially compensate for the lack of race/ethnicity information in RWD.
- We observe disparities in the representation of patients of different races and ethnicities across cancer types, which often reflects racial differences of cancer incidence rates, but there are notable exceptions.
- Future work aims to compare RWD racial distributions *vs* localized cancer incidence rates at the county level to better understand the differences presented here.

RESULTS

Distribution of global continental ancestry in pan-cancer cohort

Figure 1. We estimated global continental ancestry proportions from data of Tempus xT NGS assay using ancestry informative markers. Data was obtained from normal tissue when available and tumor tissue otherwise. Each vertical line represents the continental ancestry inference for a patient according to the color codes indicated in top right legend. Patients were stratified by major cancer types present in our cohort. "Other" represents minor cancer subtypes and cancers of unknown origin. Most patients were of European descent (72%, not shown), however, continental genetic ancestry inference identified 4.7 and 3.8-fold more patients with substantial (>50%) AFR and AMR ancestry, correspondingly, compared with TCGA (not shown).

Racial/ethic disparities in the distribution of patients sequenced per cancer type with respect to cohort-level distributions and compared to SEER incidence trends

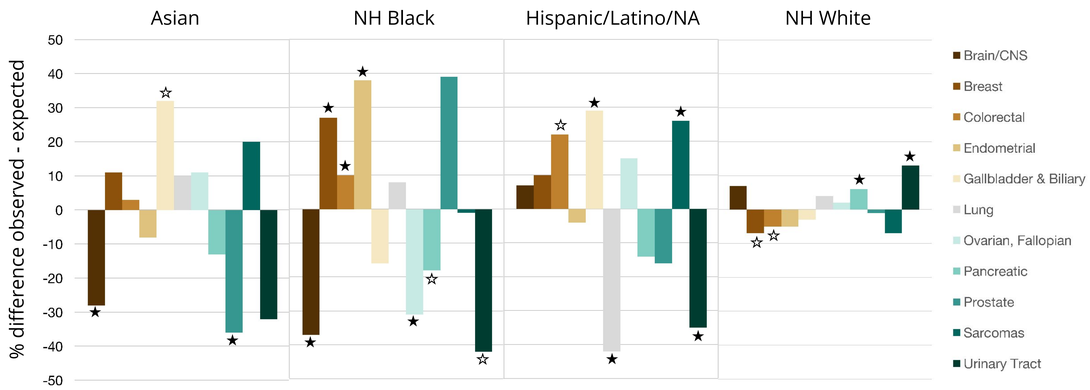

Figure 3. We observed racial/ethnic disparities in the distribution of patients of different imputed race/ethnicity as compared to expectations from the the null hypothesis of no association between race/ethnicity and cancer type. The y-axis shows the percentage difference in counts as (observed – expected)/expected x 100. A star indicates statistically higher or lower than expected number of patients (χ^2 test $p<0.05$). An open star indicates this trend is discrepant to reported incidence of that cancer type in SEER. Sample sizes: Asian = 2,657; NH Black = 6,327; Hispanic/Latino/NA = 4,799; and NH White = 38,299.

Imputing race/ethnicity using genetic ancestry unlocks additional diversity in RWD

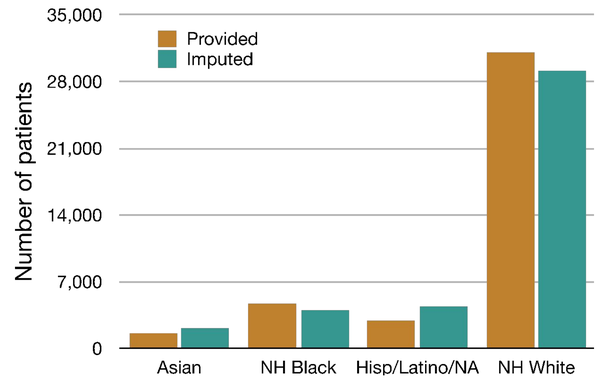

Figure 2. Race/ethnicity imputation from genetic ancestry allowed us to remediate the prevalent missingness of race/ethnicity in RWD to respectively identify 60% and 121% more patients as likely Black and Hispanic/Latino/NA for further analysis. Notably, among those missing race/ethnicity, Asian and Hispanic/Latino/NA groups are overrepresented.

Our data illustrates disparities that are in partial agreement with differences in US-level incidence rates by group (SEER). Differences are of a smaller magnitude in NH Whites despite a larger representation in the cohort.

These differences are the result of a complex interplay of factors which include stage of the disease, availability of curative therapies, comorbidities/cofactors correlated with incidence (e.g. smoking), access to care, health insurance status, socioeconomic status, and others.

Limitations: Our cohort does not represent a uniform sampling of cancer patients across the US.

ACKNOWLEDGEMENTS

We thank Vanessa Nepomuceno, Ph.D., from Tempus Scientific Communications for visualization and poster review.
Correspondence: Francisco.delavega@tempus.com.