Genetic ancestry associations with actionable somatic mutations from tumor profiling data of 100,000 cancer patients

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

INTRODUCTION

The incidence and mortality of cancer vary widely across race and ethnicity. This is attributed to an interplay of socioeconomic factors, environmental exposures, and genetic background. Cancer genomic studies have underrepresented individuals of non-European descent, thus limiting a comprehensive understanding of disparities in the diagnosis, prognosis, and treatment of cancer among these populations. Furthermore, the social constructs of race and ethnicity are far from precise categories to understand the biological underpinnings of such differences. In this study, we use a large real-world data (RWD) patient cohort to examine associations of genetic ancestry with actionable somatic alterations in known cancer driver genes.

METHODS

We inferred genetic ancestry from approximately 94,687 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).

Inclusion Criteria:

- Cancer types with at least 1,000 patients
- Genes for which at least 1% of patients harbored an actionable somatic mutation (defined as OncoKB, Levels 1 &2, R1).

Statistical Analysis:

- Ancestry proportions were transformed into isometric logratio pivot coordinates, then used as predictors in logistic regression models to discover associations between genetic ancestry proportions and the presence of somatic mutations. Each ancestry proportion association is thus adjusted for every other ancestry.
- *P*-values were adjusted for multiple testing by the Benjamini-Hochberg method to control the false discovery rate at 5%.

SUMMARY

- Genetic ancestry inference by ancestry informative markers in tumor profiling data permits to directly study the association **of ancestry with somatic mutations** and overcomes lack and ambiguity of race/ethnicity labels
- We identify several **associations** between **continental ancestry** and presence of **actionable somatic mutations** in cancer genes, some previously known (e.g. *EFGR* in lung cancer), but also several not previously described.

RESULTS

Figure 1. Distribution of global continental ancestry in pan-cancer cohort

Fig. 1: We estimated global continental ancestry proportions from data of Tempus xT NGS assay using the 654 ancestry informative markers. Data was from normal tissue when available and from tumor tissue otherwise. Patients are stratified by imputed race/ethnicity categories (NH: Non-Hispanic; Hisp/Lat: Hispanic/Latino/NA).

Figure 2. Imputing race from genetic ancestry unlocks additional diversity

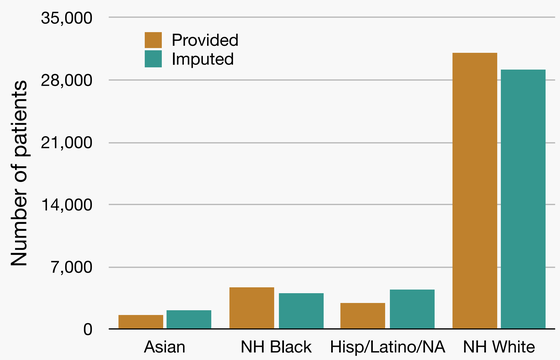


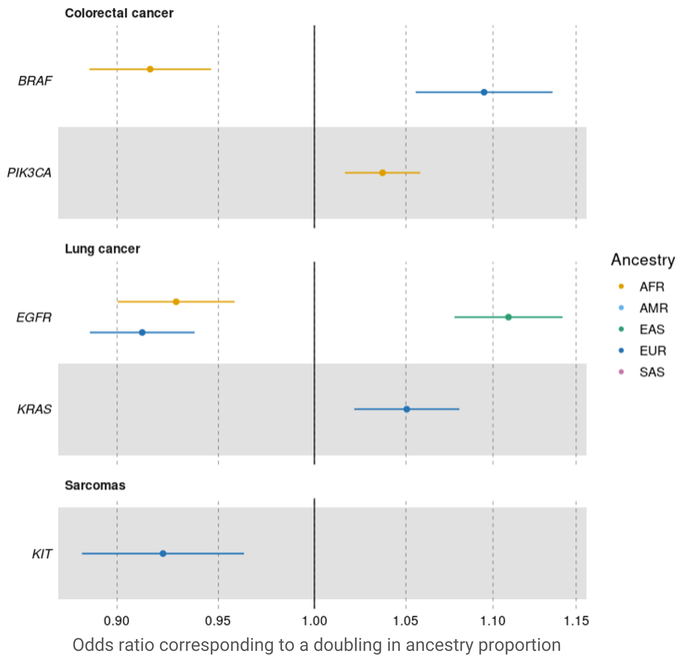
Fig. 2: Most patients were of European descent (72%), however, continental genetic ancestry inference identified 4.7 and 3.8-fold more patients with substantial (>50%) AFR and AMR ancestry, correspondingly, compared with TCGA (not shown). Using imputation based on genetic ancestry thresholds, we identified 60% and 121% more patients as likely Non-Hispanic Black and Hispanic/Latino/Native American, respectively, compared to provided race/ethnicity categories.

Fig. 3: We performed association tests between proportion of each continental ancestry and counts of actionable somatic mutations for the cancer genes and cancer types that met our criteria (see Methods). The forest plot shows the odds ratios for the presence of an actionable somatic mutation in a particular gene per a doubling of ancestry proportion.

We identify a positive association between the presence of actionable OncoKB somatic mutations in *EGFR* with EAS ancestry in lung cancer (OR=1.11), which has been previously reported using race categories. In contrast, we observe a negative association with AFR (OR=0.93) and EUR (OR=0.91) ancestries with actionable mutations in this gene.

Furthermore, we also identify 5 novel significant associations. EUR ancestry with mutations in *KRAS* in lung cancer (OR=1.05); EUR ancestry with *KIT* in sarcomas (OR=0.92); AFR in *PIK3CA* (OR= 1.04) , AFR (OR=0.92) and EUR (OR=1.09) both in *BRAF*, all in colorectal cancer.

Figure 3. Ancestry associations with actionable somatic mutations in cancer genes



This analysis is biased to variants currently in labels of FDA approved antineoplastic drugs and known resistance mutations (and hence present in OncoKB) and does not address the question of whether there are cancer driver mutations present in non-European ancestries of importance that have not been used in drug development. This is a question that warrants further exploration.

Acknowledgements: We thank Vanessa Nepomuceno, Ph.D., from Tempus Scientific Communications for visualization and poster review.

Correspondence: Francisco.delavega@tempus.com