Genetic ancestry correlates of the cancer somatic mutational landscape from tumor profiling data of 50,000 cancer patients

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

TEMPUS

Poster #10573

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 somatic alterations in known cancer driver genes.

METHODS

We inferred genetic ancestry from approximately 50,000 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).

Inclusion Criteria:
- cancer types with at least 1,000 patients
- genes for which at least 1% of patients harbored a somatic mutation

Statistical Analysis:
- Logistic regression was utilized to directly test for associations between continental ancestry proportions and presence of somatic mutations in cancer genes, controlling for assay version, gender and age.
- *P*-values were adjusted for multiple testing by the Benjamini-Hochberg method to control the false discovery rate at 5%.

For all significant associations, additional confounders such as smoking status, tumor grade, tumor tissue site, cancer stage, cancer primary site, cancer primary histology, TMB count, and MSI status were tested.

SUMMARY

- Genetic ancestry inference by ancestry informative markers in tumor profiling data permits to directly study the **influence of ancestry on somatic mutation patterns** and overcomes lack and ambiguity of race/ethnicity labels
- We identify several **associations** between **continental ancestry** and presence of **somatic mutations** in cancer genes, replicating previous observations and adding new findings that warrant further study

RESULTS

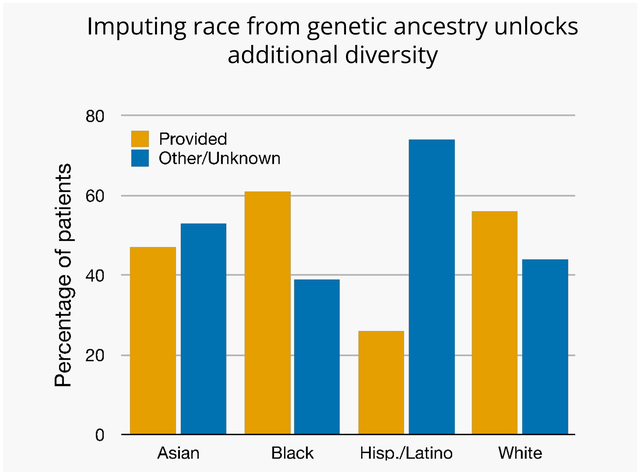

Figure 1. Most patients were of European descent (72%, not shown), however, continental genetic ancestry inference identified 4.7 and 3.8-fold more patients with substantial (>50%) AFR and AMR ancestry, correspondingly, compared with TCGA (not shown). Using imputation, we identified 60% and 121% more patients as likely Black and Hispanic/Latino, respectively, compared to provided race/ethnicity categories.

Racial disparities in cohort composition

	Asian		Black		Hisp./Latino		White	
Cancer	N	Diff	N	Diff	N	Diff	N	Diff
Brain/CNS	59	-	102	-2%	156	-	1236	-
Breast	207	-	612	3%	417	-	2731	-1%
Colorectal	248	-	669	2%	578	2%	3516	-1%
Heme	55	-	144	-	127	-	1063	-
Lung	386	-	883	-	401	-7%	5383	-
Ovarian	109	-	124	-2%	202	-	1360	-
Pancreatic	133	-	304	-1%	245	-2%	2533	-
Prostate	72	-2%	421	3%	211	-	1795	-

Table 1. We observed several racial disparities in the distribution of patients of different imputed race/ethnicity as compared to expectations from the overall cohort-level distributions *vs.* SEER incidence. Orange shading – higher than expected number of patients; blue shading – lower than expected (chi-squared test p<0.05).

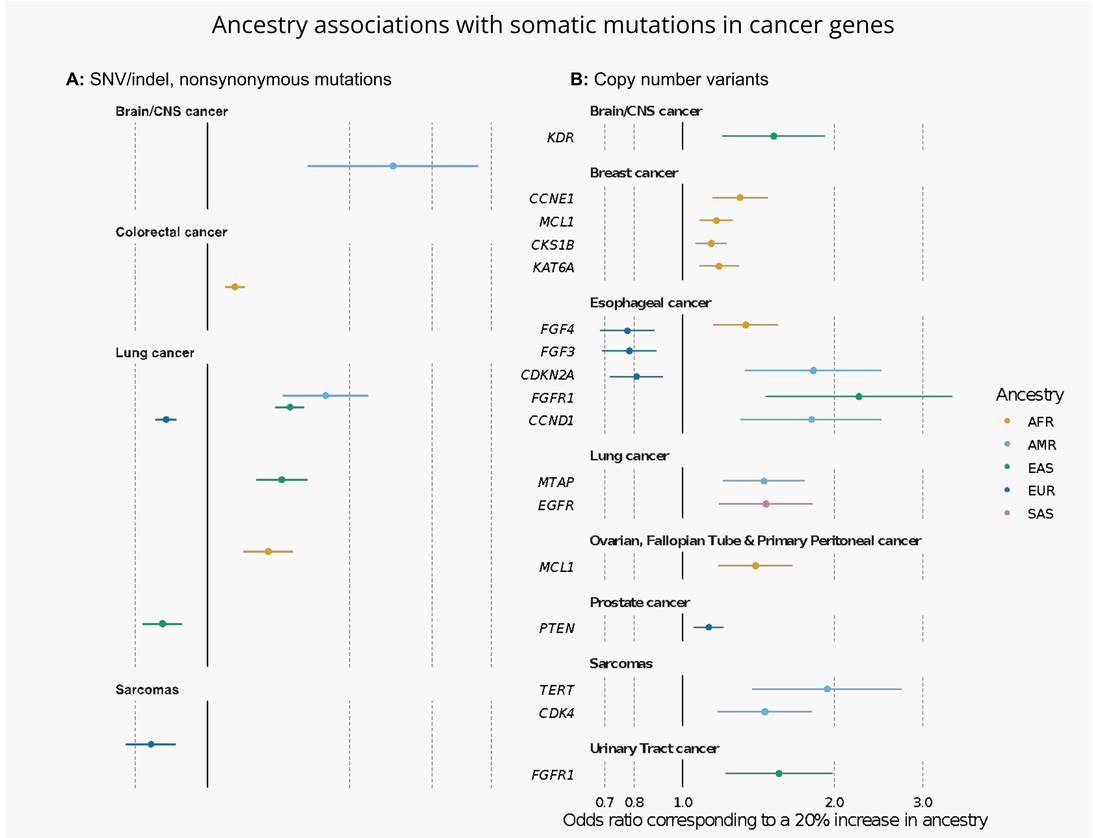

Figure 2. We performed association tests between proportion of each continental ancestry and counts of somatic mutations for the cancer genes and cancer types that met our criteria (see Methods). Panel A shows the results for statistically significant associations for nonsynonymous small variants (SNVs and small indels) and Panel B shows the results for copy number variants (CNVs). The forest plots show the increased/decreased odds of having a variant in a gene for every 20% increase in ancestry proportion. Some of our analyses replicate previous findings (e.g. increased somatic mutations in *EGFR* associated with Native American ancestry in lung cancer), but not others. Furthermore, we find several new associations with both small variants and CNVs across several cancer types that are not fully explained by the confounders tested.

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

Correspondence: Francisco.delavega@tempus.com