Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Mahal BA, Alshalalfa M, Kensler KH, et al. Racial differences in genomic profiling of prostate cancer. N Engl J Med 2020;383:1083-5. DOI: 10.1056/NEJMc2000069

SUPPLEMENTAL APPENDIX

Table of Contents

1. Methods/Acknowledgments………………………………………………...Pg. 2-5
2. Supplemental Figure …………………………..……………………………Pg. 6
3. References…………………………………………………………………..Pg. 7

METHODS:

Patient Cohort

Next generation sequencing (NGS) tumor genomic profiles (somatic non-synonymous genomic alteration data) of patients treated for primary or metastatic prostate cancer were identified from the American Association for Cancer Research (AACR) Project Genomics Evidence Neoplasia Information Exchange (GENIE) 7.0-registry (released January, 2020).[1,2] The GENIE v7.0-registry contains genomic data that is obtained during the course of routine practice at multiple participating institutions. The AACR Project GENIE is a multiphase, multi-year, data-sharing project that captures genomic data for cancer patients across a multiple international institutions. As described in the AACR Project GENIE Data Guide (https://www.aacr.org/Research/Research/Documents/20190713_GENIE_Data_Guide_6.1-public_final.pdf), "the database currently contains CLIA-/ISO-certified genomic data obtained during the course of routine practice at multiple international institutions (Table 1), and will continue to grow as more patients are treated at additional participating centers."

The majority of prostate cancer cases (86%) in the GENIE v7.0-registry are captured by Memorial Sloan Kettering Cancer Center (MSK)-IMPACT and Dana-Farber Cancer Institute (DFCI)-Oncopanel NGS initiatives. MSK and DFCI were the only institutions with race demographic data that also captured >3% of prostate cancer patients in the registry. Therefore, MSK and DFCI samples were used for cohort selection. Data were downloaded from cBioPortal (v3.2.1) GENIE cohort v7.0.

Genomic Sequencing Statistical Methods

MSK and DFCI each contributed NGS data from 3 different panels (with increasing coverage) to the GENIE 7.0-registry (MSK 341, MSK 410, MSK 468, DFCI 1, DFCI 2, DFCI 3); specifics about the DFCI and MSK NGS assays are described below.[1,2] In total, these NGS panels analyzed 474 unique genes, including all actionable targets (i.e., genes that are targeted by therapies and that provide information about the disease); **Supplemental Table 1** describes the genes included in each NGS panel (MSK 341, MSK 410, MSK 468, DFCI 1, DFCI 2, DFCI 3). Mutational profiles were examined by race (White, Black, Asian) in primary and metastatic disease; **Supplemental Table 2** describes the distribution of patient samples by race, stage, and panel. Mutational frequencies were calculated and compared across race. To account for panel differences, the denominator used to calculate mutational frequency for each gene included the number of patients who received sequencing for each specific gene, rather than the total number of patients sequenced; the AR gene was captured across all panels. The Benjamini-Hochberg method was used to control for false discovery rate (FDR), with q-value < 0.05 considered significant.[3] Difference in mutation proportions (PropDiff) across race with associated 95% confidence intervals were reported. Only P- that were significant after controlling for FDR were reported. The Dana-Farber/Harvard Cancer Center institutional review board granted a waiver of informed consent for this study.

Next Generation Sequencing Assays

Specifics about genomic profiling at each center are provided below, as copied from the AACR Project GENIE Data Guide:

(https://www.aacr.org/Research/Research/Documents/20190713_GENIE_Data_Guide_6.1-public_final.pdf).

Dana-Farber Cancer Institute (DFCI)

“DFCI uses a custom, hybridization-based capture panel (OncoPanel) to detect single nucleotide variants, small indels, copy number alterations, and structural variants from tumor-only sequencing data. Three (3) versions of the panel have been submitted to GENIE: version 1 containing 275 genes, version 2 containing 300 genes, version 3 containing 447 genes. Specimens are reviewed by a pathologist to ensure tumor cellularity of at least 20%. Tumors are sequenced to an average unique depth of coverage of approximately 200x for version 1 and 350x for version 2. Reads are aligned using BWA, flagged for duplicate read pairs using Picard Tools, and locally realigned using GATK. Sequence mutations are called using MuTect for SNVs and GATK SomaticIndelDetector for small indels. Putative germline variants are filtered out using a panel of historical normals or if present in ESP at a frequency >= .1%, unless the variant is also present in COSMIC. Copy number alterations are called using a custom pipeline and reported for fold-change >1. Structural rearrangements are called using BreaKmer. Testing is performed for all patients across all solid tumor types. Version 3 includes the exonic regions of 447 genes and 191 intronic regions across 60 genes targeted for rearrangement detection. 52 genes present in previous versions were retired in the v3 test.”

Memorial Sloan Kettering Cancer Center (MSK)

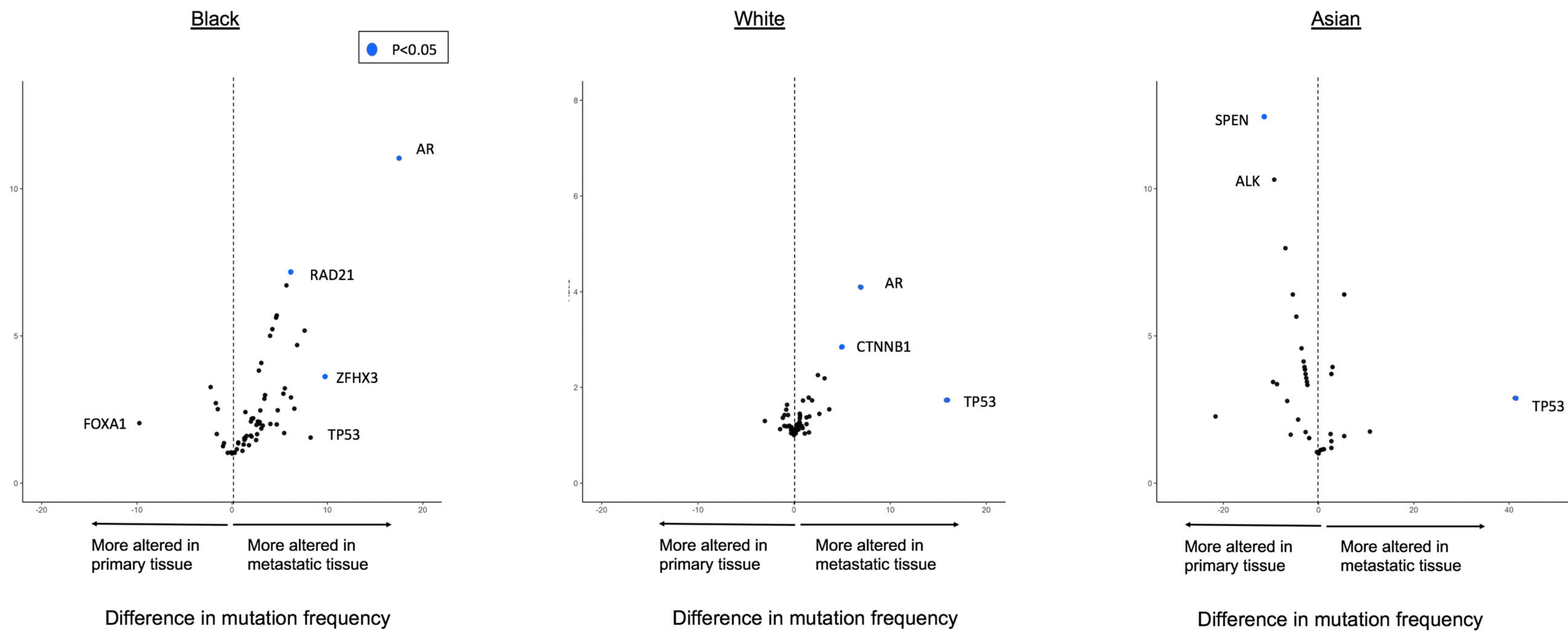
"MSK uses a custom, hybridization-based capture panel (MSK-IMPACT) to detect single nucleotide variants, small indels, copy number alterations, and structural variants from matched tumor-normal sequence data. Three (3) versions of the panel have been submitted to GENIE: version 1 containing 341 genes, version 2 containing 410 genes, version 3 containing 468 genes. Specimens are reviewed by a pathologist to ensure tumor cellularity of at least 10%. Tumors are sequenced to an average unique depth of coverage of approximately 750X. Reads are aligned using BWA, flagged for duplicate read pairs using GATK, and locally realigned using ABRA.

Sequence mutations are called using MuTect, VarDict, and Somatic indel detector, and reported for >5% allele frequency (novel variants) or >2% allele frequency(recurrent hotspots). Copy number alterations are called using a custom pipeline and reported for fold-change >2. Structural rearrangements are called using Delly. All somatic mutations are reported without regard to biological function. Testing is performed for patients with advanced metastatic cancer across all solid tumor types."

ACKNOWLEDGEMENTS:

The authors would like to acknowledge the American Association for Cancer Research (AACR) and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of study authors.

Supplemental Figure. Relative difference (x-axis) and ratio (y-axis) of mutation frequency in metastatic and primary disease for each race in 2393 patients (2109 White, 204 Black, 80 Asian).
Blue circles represent statistically significant differences (P<0.05).



REFERENCES:

1. Zehir A, Benayed R, Shah RH, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med.* 2017;23(6):703-713.
2. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov.* 2017;7(8):818-831.
3. Betensky RA, Newberger NG. New Guidelines for Statistical Reporting. *N Engl J Med.* 2019;381(16):1597.