African ancestry-informative markers and their potential in identifying cancer-related SNPs

Boudeau Samantha[1,2], Ramakodi Meganathan P. PhD[2], Kulathinal Robert J., PhD[1] , Ragin Camille, PhD, MPH[2]

[1]Department of Biology, College of Science and Technology, Temple University, Philadelphia, PA

[2] Cancer Prevention and Control Program, Fox Chase Cancer Center- Temple University Health System, Philadelphia, PA

INTRODUCTION

- Certain ethnic populations, namely populations of high African ancestry, are critically underrepresented in biomedical research even as more data is available than ever before
- Most of the available data represents homogeneous populations and miss a significant portion of the global human genetic diversity
- Tools and approaches developed from and applied to such data for the study of diseases fail to translate well across populations
    - Zavala et al.(2021) found that SNPs associated with breast cancer identified by GWAS done in mainly European and Asian populations have poor predictive power in people with high level of African ancestry
- Ancestry informative markers are SNPs that appear at highly different frequencies in different populations
    - There is evidence that ancestral variants might be contributing to disease risk (Campbell & Tishkoff (2008))
- Consequently, it is plausible that population specific AIMs are associated with disease risk and could help elucidate the genetic basis of health disparity in diseases like cancer

OBJECTIVES

- Identify Ancestry Informative Markers (AIMs) for populations of African descent
- Functionally characterize the AIMs and Identify potentially disease-relevant AIMs
- Investigate potential role of AIMs in cancer health disparity

METHODS

- 1000 Genome Project genotype data for the 26 sub-populations
- Infolcalc was used to identify the AIMs (In > = 0.25) – Fig 1.
- A subset of ~ 9000 SNPs were used to genotype ~1400 cancer patients and controls (approximately half of African descent and half of European descent) to validate the AIMs
- SNPs annotation using ANNOVAR, POLYPHEN, SIFT, ML-LR/SVM
- Enrichment analysis using GSEA

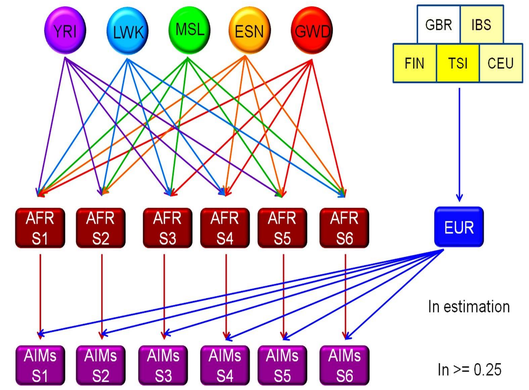

Fig 1. Schematic of methodology used to generate an African ancestry informative markers (AIMs) panel

RESULTS

Population Analysis

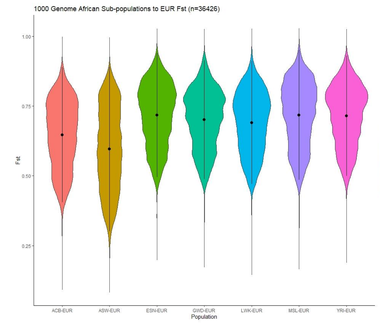

Fig 2. Distribution of the Fst values as estimated between each African subpopulation and the European super-population. -

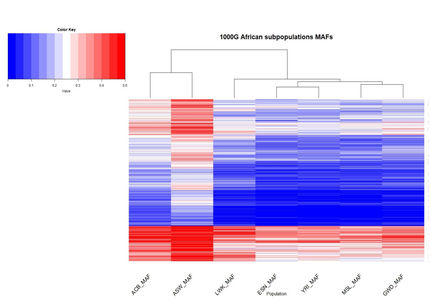

Fig 3. Heatmap showing the MAF of the AIMs in the African subpopulations and the 2 other subpopulations with high African ancestry

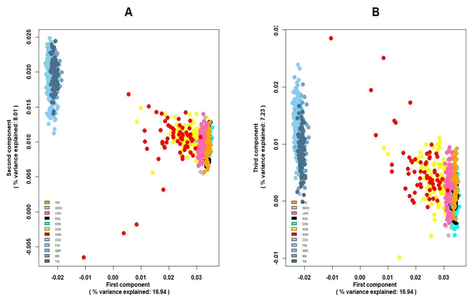

Fig 4. Principal component analysis of the AIMs showing them clearly discriminate between the Europeans subpopulations (blues) and the African subpopulations (multicolor).

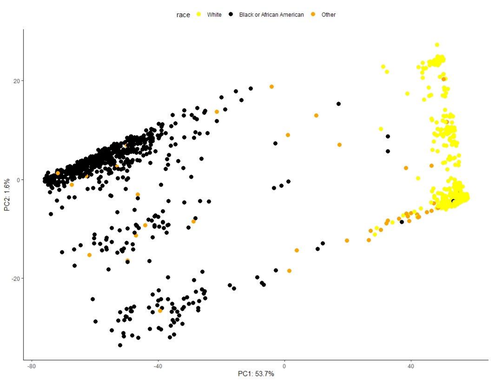

Fig 5. Principal component analysis of 1445 head and neck cancer patients and controls genotyped for 9158 AIMs showing the participants self-identified as Black (Black dots) and White (Yellow dots)

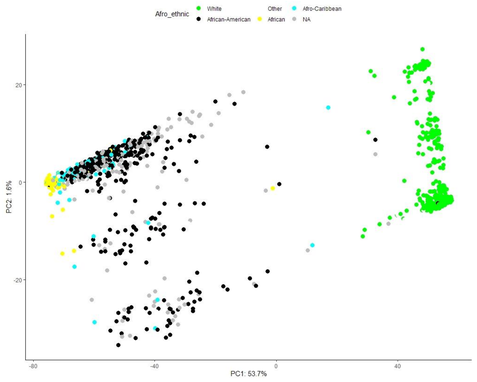

Fig 6. Principal component analysis of 1445 head and neck cancer patients and controls genotyped for 9158 AIMs showing the participants highlighting the clustering of the 3 subgroups of Black or African-Americans (Yellow, Blue, and Black)

Code	Population
YRI	Yoruba in Ibadan, Nigeria
LWK	Luhya in Webuye, Kenya
MSL	Mende in Sierra Leone
ESN	Esan in Nigera
ASW	Americans of African Ancestry in SW USA
ACB	African Carribean in Barbados
GWD	Gambian in Western Division

Table 1: 1000 Genome Project population code and associated population

Functional Analysis

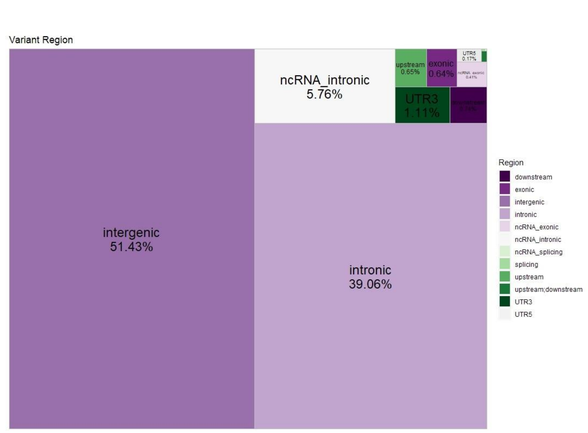

Fig 7. Figure showing the locations of the AIMs in the genome with 0.64% being exonic, 51.43% being intergenic, and 39.06% being intronic

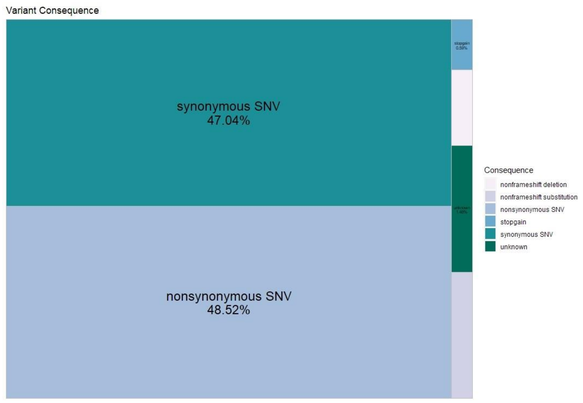

Fig 8. Figure showing the consequence of the exonic variants showing a nearly equal distribution of non-synonymous and synonymous variants

CONCLUSIONS

- 46,737 AIMs were identified from the analysis in Fig. 1 representing SNPs common to panels S1-S6
- Fig 2. provide further evidence that these SNPs are African AIMs as evidenced by the high Fst values between the African and African diaspora subpopulations and the European superpopulation
- The bulk of the AIMs have very low MAFs in the African subpopulations studied and also reflect the changes in frequency likely brought on by admixture in ASW and ACB
- The AIMs characterize the population substructures differences expected between the continental African sub-populations and the admixed Africans of the Americas namely the African-Caribbean of Barbados and the African-Americans (Fig 2., 3., 4.)
- The AIMs effectively differentiate between the African sub-population and European populations(Fig 4.)
- Tested on “real world” data, the AIMs cluster patients as expected based on their self-identified race and also show that they pick up within group differences without people of African descent (fig 5., 6.)
- Pre-ranked GSEA performed on the non-synonymous SNPs show non-statistically significant enrichment in the DODD_NASOPHARYNGEAL_CARCINOMA_UP pathway which represents Genes up-regulated in nasopharyngeal carcinoma (NPC) compared to the normal tissue.
- Of the nonsynonymous SNPs, rs12186491 and rs6601495 have been predicted as deleterious by at least 3 variant effect predictors while rs9830253, rs7645635, rs16891982 and rs10238965 have been identified by at least 2 predictors
- Preliminary results from analysis looking at association between the AIMs and HNSCC disparity show that rs2238151 act as an eQTL affect ALDH2. Low level of ALDH2 is associated with poor cancer prognosis.

References

Campbell, M. C., & Tishkoff, S. A. (2008). African Genetic Diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. *Annual Review of Genomics and Human Genetics*, *9*(1), 403–433. https://doi.org/10.1146/annurev.genom.9.081307.164258

Zavala, V. A., Bracci, P. M., Carethers, J. M., Carvajal-Carmona, L., Coggins, N. B., Cruz-Correa, M. R., Davis, M., de Smith, A. J., Dutil, J., Figueiredo, J. C., Fox, R., Graves, K. D., Gomez, S. L., Llera, A., Neuhausen, S. L., Newman, L., Nguyen, T., Palmer, J. R., Palmer, N. R., … Fejerman, L. (2021). Cancer health disparities in racial/ethnic minorities in the United States. *British Journal of Cancer*, *124*(2), 315–332. https://doi.org/10.1038/s41416-020-01038-6