C139

Robustness of a Targeted Methylation-Based Multi-Cancer Early Detection (MCED) Test to Population Differences in Self-Reported Ethnicity

AACR Conference on The Science of Cancer Health Disparities in Racial/Ethnic Minorities and the Medically Underserved
September 29-October 2, 2023

Oliver Venn, PhD,[1]* Joerg Bredno, PhD,[1]* Alexis Thornton, PhD,[1] Christopher Chang, PhD,[1] Earl Hubbell, PhD,[1] Kathryn Kurtzman, MD,[1] John Beausang, PhD,[1] Charles Swanton, MBPhD, FRS[2,3]
[1]GRAIL, LLC, Menlo Park, California, USA; [2]Francis Crick Institute, London, UK; [3]UCL Cancer Institute, CRUK Lung Cancer Centre of Excellence, London, UK
*Contributed equally

INTRODUCTION

- DNA methylation is influenced by environmental and genetic factors and is known to vary by race and ethnicity[1,2]
- A cell-free DNA (cfDNA) methylation-based multi-cancer early detection (MCED) test has been available as a validated laboratory-developed test (LDT) since June 2021[3–5]
- The targeted methylation-based MCED LDT was designed to identify rare methylation patterns in fragments of circulating tumor DNA that are highly specific to cancer and uncommon in a large and diverse study population without diagnosed active cancer[6]
- Such a test must detect cancer-specific methylation in the presence of population diversity to be useful as a screening tool[6]
- Continued investigation is merited because clinical study cohorts may not adequately represent the entire intended screening population,[7–9] such as underserved populations who may benefit most from improved cancer screening
  - Innovative strategies have been employed to address inequity in health research studies, such as the novel data-enabled invitation algorithm used for recruitment into the NHS-Galleri clinical trial[10]
- Clinical studies often collect participants' self-reported ethnicity (SRE), which reflects both genetic and non-genetic information, including behavioral, cultural, and societal norms, as well as skin color and other qualitative influences[11]
- Genetically-inferred ancestry (GIA) quantifies the heritage contributed to an individual by ancestors from distinct demographics and/or diverse geographic regions around the globe. Compared to SRE, GIA may better capture the presence or absence of ancestry-specific methylation changes interfering with the analysis of cancer-specific methylation changes[12]

OBJECTIVES

With awareness that neither SRE nor GIA completely represents a population's diversity, we aimed to:

- Evaluate the agreement between SRE and GIA when assessing participant diversity in clinical studies that use genomic data
- Evaluate the robustness of a cfDNA methylation-based MCED LDT to population diversity based on participant SRE in the training (N = 4487) and validation (N = 5309) datasets from the 15,254-participant Circulating Cell-free Genome Atlas (CCGA; NCT02889978) study

KEY RESULTS: CANCER DETECTION BY THE MCED LDT DOES NOT VARY STRONGLY BY SELF-REPORTED ETHNICITY

Good agreement of SRE and GIA

- In the MCED LDT validation population, there was good agreement between most SRE categories and the associated top GIA categories
  - **99%** (274/278) of participants with Black, non-Hispanic SRE had the top match of African GIA
  - **99%** (3291/3312) of participants with White, non-Hispanic SRE had the top match of European GIA
  - **84%** (61/73) of participants with Asian, Native Hawaiian, or Pacific Islander SRE had the top match of East Asian or South Asian GIA
  - **70%** (207/295) of participants with Hispanic SRE had the top match of Admixed American GIA
- Using the top predicted GIA category to classify the continuous nature of genetic ancestry has limitations in clinical applications[13]
  - 2.5% (102/4077) of participants reported more than one SRE category. A more detailed investigation beyond the top scoring GIA category may be required to disambiguate the ancestry of this group
  - 15 participants responded as American Indian or Alaska Native SRE, which limits the comparisons that can be made in this subset. Two of the 15 had a top match of Admixed American GIA, suggesting a more detailed investigation may also be required for this SRE category[14]

SRE was not associated with cancer signal scores under multvariable regression

- Eight different linear models were created to assess the impact of sex, age, smoking status, and SRE on classifier cancer signal scores in the non-cancer participants of the MCED LDT validation population
- The model that only considered age had the lowest Akaike Information Criterion (AIC; see Methods) and was, therefore, identified as the best model to explain cancer signal score variation using the fewest possible independent variables (**Table 1**)
- In each of the competing models, neither sex, smoking status, nor SRE were significantly correlated with MCED classifier cancer signal scores (**Table 1**)
- Age was weakly associated with MCED classifier cancer signal score ($r^2 < 0.01$)

Experimental classifiers with intentionally skewed SRE demonstrated robust performance

- The multiple SRE (M) and single SRE (S) experimental classifiers had nearly identical sensitivity (M: **57%** [95% CI: 52-61], 288/507 vs S: **57%** [53-62], 290/507) in a separate held-out population of 507 cancer participants with SRE other than White (**Figure 1**)
- A McNemar test identified no difference between M and S classifier sensitivity ($p > 0.5$; **Table 2**) in the cancer participant population with SRE other than White, despite the relatively large difference in SRE training composition
- M and S classifiers had nearly identical cancer signal origin (CSO) prediction accuracy in the cancer participant population with SRE other than White (M: **91%** [88-94] 263/288 vs S: **90%** [86-93], 262/290; **Figure 2**)

Table 1. Covariates for MCED LDT classifier cancer signal scores.

Variables in linear model	Covariate or confounder	p-value	ΔAIC ≤ 2
Age	Age	**< 0.001**	√
Age, SRE	Age	**< 0.01**	√
	SRE	0.24	
Age, Sex	Age	**< 0.01**	√
	Sex	0.58	
Age, Smoking status	Age	**< 0.001**	√
	Smoking status	0.60	
Age, Sex, Smoking status, SRE	Age	**< 0.01**	**X**
	Sex	0.50	
	Smoking status	0.58	
	SRE	0.23	
SRE	SRE	0.11	**X**
SRE, Sex	Sex	0.23	**X**
	SRE	0.10	
Sex	Sex	0.27	**X**

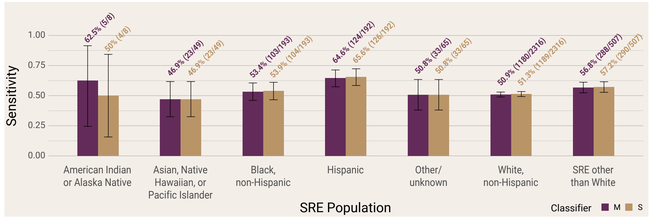
Models are displayed in order of increasing AIC. Models with a ΔAIC within 2 units (√) of the best model (age only) were similar, whereas models with a ΔAIC > 2 units (**X**) from the best model were considered significantly worse at explaining variation with few independent variables. p-values indicate the significance of the covariates within each logistic regression model.
AIC, Akaike Information Criterion; LDT, laboratory-developed test; MCED, multi-cancer early detection; SRE, self-reported ethnicity.

Table 2. McNemar sensitivity comparison for the cancer participant population with SRE other than White (p = 0.617).

	Detected by S classifier	Not detected by S classifier
Detected by M classifier	287	1
Not detected by M classifier	3	216

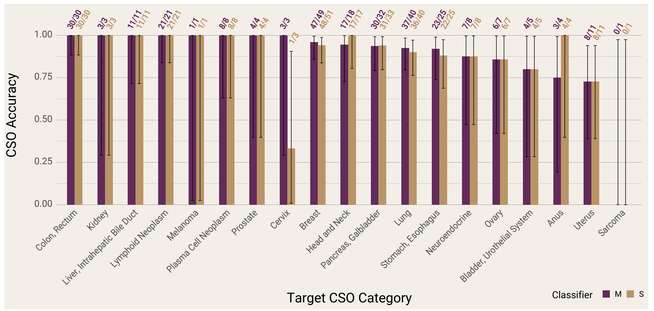
M, multiple SRE; S, single SRE; SRE, self-reported ethnicity.

Figure 1. Sensitivity was largely consistent between M and S experimental classifiers across SRE populations.



M, multiple SRE; S, single SRE; SRE, self-reported ethnicity.

Figure 2. CSO prediction accuracy was consistent between experimental classifiers in the cancer participant population with SRE other than White.



Seven individuals with multiple primaries, cancer of unknown primary, or cancers with no matching CSO target class are not shown.
CSO, cancer signal origin; M, multiple SRE; S, single SRE.

CONCLUSIONS

- SRE analysis supports the CCGA study as a diverse and representative dataset with 19% (765/4077) SRE other than White in the MCED validation subset reported here
- Although DNA methylation patterns at sites across the genome vary by SRE, MCED LDT classifier cancer signal scores were consistent across non-cancer participants in different SRE populations
- In two separate analyses of a diverse population of cancer participants with SRE categories of Black, non-Hispanic; Hispanic; American Indian or Alaska Native; and Asian, Native Hawaiian or Pacific Islander, SRE-associated DNA methylation patterns did not significantly contribute to MCED LDT classifier performance
- Cancer signal detection in a population with SRE categories of Black, non-Hispanic; Hispanic; American Indian or Alaska Native; and Asian, Native Hawaiian or Pacific Islander was consistent regardless of classifier training exposure, including skewed training populations of only White, non-Hispanic participants
- These findings indicate that detection of invasive cancer by the MCED LDT is not strongly influenced by SRE
- Ongoing interventional studies of the MCED test (PATHFINDER2 [NCT05155605] and NHS-Galleri [NCT05611632]) are targeting enrollment of population-representative SREs, which will allow for further validation and comparison of GIA to SRE

METHODS

Collection of Self-Reported Ethnicity (SRE)

- Participants enrolled in the CCGA (NCT02889978) and STRIVE (NCT03085888) studies completed a participant questionnaire that requested the selection of a race and an ethnicity category following the US census race groups at the time of CCGA study initiation
- The questionnaire allowed participants to identify race as "American Indian or Alaska Native", "Asian", "Black or African American", "Native Hawaiian", "Other Pacific Islander", or "White". Participants could choose to not report a category, choose multiple categories, or select "Unknown"
- Participants identified ethnicity as either "Hispanic or Latino" or "Not Hispanic or Latino"
- The final SRE categories—"American Indian or Alaska Native", "Asian, Native Hawaiian, or Pacific Islander", "Black, non-Hispanic", "Hispanic", "White, non-Hispanic", and "Other/unknown"—were derived following the specifications from the Breast Cancer Surveillance Consortium[15]

Genetically-Inferred Ancestry (GIA) and comparison with SRE

- GIA was estimated by isolating N = 6859 single nucleotide polymorphisms (SNPs) with population minor allele frequency (MAF) >5% that overlap the targeted methylation panel used in the MCED assay. Because of bisulfite conversion of cytosine to thymine, SNPs with C>T alleles on the forward strand and A>G alleles on the reverse strand were ignored
- Principal component analysis was performed using the MAF of these SNPs in the 1000 Genomes Project[16] and was used to classify the 5 super populations, ie, African, Admixed American, East Asian, European, and South Asian populations
- The GIA classifier used multinomial logistic regression and was developed using participants (N = 192) from the second CCGA substudy,[3] none of which were used in the experimental classifier analysis described below. The top scoring GIA was assigned to each sample
- GIA for all participants in the MCED LDT validation population[4] (N = 4077 clinically evaluable participants with evaluable assay results) was obtained from their cfDNA samples, and the co-occurrence with their SRE was recorded

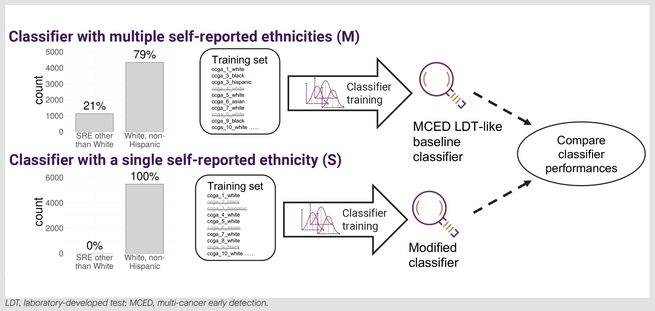
Identification of covariates for MCED classifier cancer signal scores

- Classifier cancer scores from 1254 non-cancer participants from the MCED LDT validation population[4] were regressed against the potential covariates or confounders of sex, age, smoking status, and SRE
- Two SRE groups were considered for the analysis: White, non-Hispanic (996 participants) and all SRE other than White (258 participants)
- Linear models for sex, age, smoking status, SRE, and selected subsets of these covariates or confounders were created. The model that best explains the greatest amount of variation of classifier cancer scores using the fewest possible independent variables was identified using the Akaike Information Criterion (AIC). The AIC supports model selection by assessing each model's trade-off between the likelihood of observing the classifier cancer scores given a fitted model and the number of parameters[17]
- In each of these models, candidate covariates or confounders were analyzed separately by p-value and coefficient of determination r^2 to assess if they significantly affected classifier cancer signal scores

Experimental classifiers with intentionally skewed SRE training populations

- To evaluate classifier generalization across SREs, 2 experimental MCED test classifiers (multiple SRE, M; single SRE, S) were trained similarly to the MCED LDT[3] with intentionally skewed SRE compositions (**Figure S1**)
- M was trained with all available participants with SRE other than White (21%), and S was trained with 100% White, non-Hispanic participants
- Training sets for both classifiers were matched on demographics (age, sex, smoking status), number of participants, number and type of training samples (plasma and tissue samples), and number of cancer types and clinical stages. In addition to participants from CCGA, training sets for both classifiers also included participants from STRIVE (NCT03085888)
- Sensitivity (at 99.4% specificity) and accuracy of cancer signal origin (CSO) prediction for the 2 classifiers were assessed in a separate held-out population of 507 cancer participants with SRE other than White from the MCED LDT validation cohort[4]

Figure S1. Experimental MCED test classifiers were trained with either multiple SRE or single SRE participants.



LDT, laboratory-developed test; MCED, multi-cancer early detection.

References

1. Zhang FF, et al. *Epigenetics*. 2011;6(5):623-629.
2. Adkins RM, et al. *Birt Defects Res A Clin Mol Teratol*. 2011;91(8):728-736.
3. Liu MC, et al. *Ann Oncol*. 2020;31(6):745-759.
4. Klein E, et al. *Ann Oncol*. 2021;32(9):1167-1177.
5. GRAIL. GRAIL Presents Interventional PATHFINDER Study Data at 2021 ASCO Annual Meeting and Introduces Galleri, a Groundbreaking Multi-Cancer Early Detection Blood Test. Published online June 4, 2021. Accessed August 28, 2023. https://grail.com/press-releases/grail-presents-interventional-pathfinder-study-data-at-2021-asco-annual-meeting-and-introduces-galleri-a-groundbreaking-multi-cancer-early-detection-blood-test/
6. Tang WHW, et al. *Prev Med*. 2023;167:107384.
7. Buttenstein I, et al. *Sci Rep*. 2023;13(1):42.
8. Improving Representation in Clinical Trials and Research: Building Research Equity for Women and Underrepresented Groups. Washington (DC): National Academies Press (US); 2022. Available from: https://www.ncbi.nlm.nih.gov/books/NBK584396/.
9. Febbo P. Improving cancer outcomes through equitable access to cfDNA tests. Plenary Session presented at: AACR; April 17, 2023.
10. Neal RD, et al. *Cancers*. 2022;14(19):4818.
11. Mersha TB, et al. *Hum Genomics*. 2015;9(1):1.
12. Newman LA, et al. *JAMA Surg*. 2018;153(4):299-300.
13. Ding, Y, et al. *Nature*. 2023;618(7966):774-781.
14. Graves S, et al. *PLoS Genet*. 2013;9(12):e1004023.
15. Braithwaite D, et al. *JAMA Intern Med*. 2018;178(4):494-501.
16. The 1000 Genomes Project Consortium. *Nature*. 2015;526(7571):68-74..
17. Portet, S. *Infect Dis Model*. 2020;5:111-128.

Acknowledgements

Funded by GRAIL, LLC. Writing, editorial, and graphic assistance provided by Randall Janairo (GRAIL, LLC) and Alexandra Thomas (Prescott Medical Communications Group).

Disclosures

Poster