Simultaneous Multi-cancer Detection and Tissue of Origin (TOO) Localization Using Targeted Bisulfite Sequencing of Plasma Cell-free DNA (cfDNA)

Geoffrey R. Onnord, MD; Eric A. Klein, MD; Michael V. Seiden, MD; Earl Hubbard, PhD; Oliver Venn, DPh; Arash Jamshidi, PhD; Nan Zhang, PhD; John F. Beausang, PhD; Samuel Gross, PhD; Kathryn N. Kurtzman, MD; Eric T. Fung, MD, PhD; Jessica Yecies, PhD; Rita Shakhnovich, MD, PhD; Alexander P. Felsid, PhD; Mikhail A. Sekeres, MD; Donald Richards, MD, PhD; Peter P. Yu, MD; Alexander M. Aravanis, MD, PhD; Anne-Renee Hartman, MD; Minetta C. Liu, MD*

**INTRODUCTION**

- A non-invasive cell-free DNA blood test detecting multiple cancers at earlier stages (stages I-III) could decrease cancer mortality.
- For a multi-cancer test to be effective at population scale, it should:
  - Detect clinically significant cancer with a low false positive rate (ie, very high specificity [>99%] to limit overdiagnosis)
  - Identify specific tissue origin to direct appropriate diagnostic work-up for detected cancers.1
- In earlier discovery work, whole-genome bisulfite sequencing outperformed whole genome and targeted sequencing approaches for multi-cancer detection across cancer stages at high specificity,2,3 targeted methylation was selected for further assay development, including training and internal cross-validation.
- Presented here are data from a second pre-specified subability of Circulating Cell-free DNA Atlas (CCDA; n=354,884), in which a multi-cancer detection and tissue of origin (TOO) localization approach using bisulfite sequencing of plasma cfDNA to identify methylation signatures was validated in preparation for returning results in a clinical setting.

**METHODS**

- The primary analysis population used for this validation was comprised of 1,265 participants derived from the CCGA and STRIVE study populations (Figure 1). CCGA is a multi-center, case-control, observational study with longitudinal follow-up (n=354 participants enrolled: 150% cancer, 46% non-cancer) and STRIVE is a multi-center, prospective, cohort study enrolling women undergoing screening mammography (n=252 participants enrolled).
- The trained classifier targeting specificity of >99% (see Methods) achieved specificity of 99.8% in the cross-validated training set and 99.3% in the independent validation set (P=0.095). The trained classifier targeting specificity of >99% (see Methods) achieved specificity of 99.8% in the cross-validated training set and 99.3% in the independent validation set (P=0.095).
- Among pre-specified high-signal cancer types, the stage-specific consistency was >90% (69/76) in stage II and >90% (71/79) in stage I (Tables 1). Therefore, assay performance reflected a consistent false positive rate of <1%.
- Across stages, multiple deadly cancer types that currently have no screening paradigm were detected, and high specificity and identify conspecific tissue types for 96% (344/359) of cases; of these (and consistent with training set analyses), the TOO classification agreed perfectly with tissue type for 90.7% (318/352) of cases. The trained classifier targeting specificity of >99% (see Methods) achieved specificity of 99.8% in the cross-validated training set and 99.3% in the independent validation set (P=0.095).

**RESULTS**

- The trained classifier targeting specificity of >99% (see Methods) achieved specificity of 99.8% in the cross-validated training set and 99.3% in the independent validation set (P=0.095).
- Therefore, assay performance reflected a consistent false positive rate of <1%
- Importantly, the assay specificity and sensitivity were consistent between the cross-validated training set and independent validation set across stages (Figure 3), confirming that training data were not overfitted. This was also consistent for all (120) cancer types.

**CONCLUSIONS**

- Across stages, multiple deadly cancer types that currently have no screening paradigm were detected, and simultaneously accurately localized to a TOO, using methylation signatures in plasma cfDNA.
- This was achieved with trained thresholds that resulted in a single fixed, low false positive rate (<1%) in an independent validation set.
- Importantly, results in the independent validation set were indistinguishable from the training set, demonstrating robustness of machine learning classifier training, including evidence of overtraining.
- This validation supports the feasibility of a single blood-based test that can simultaneously detect multiple cancers and supports further clinical development for the preparation of return results.