3135

Validation of a bioinformatic model for classifying non-tumor variants in a cell-free DNA liquid biopsy assay

Jennifer Yen[1], Pegah Safabakhsh[1], Jeffrey L. Werbin[1], Denis Tolkunov[1], Yu Fu[1], Katie Quinn[1], Robert Foley[1], Minh Tran[1], Xianxian Liu[1], Carin Espenschied[1], Scott Higdon[1], Brett Kennedy[1], Shile Zhang[1], Han-Yu Chuang[1]

[1] Guardant Health, Palo Alto, CA

GUARDANTINFINITY™

Introduction

Liquid biopsy is a powerful, non-invasive tool for profiling tumors and identifying clinically relevant variants.

The presence of clonal hematopoiesis (CH) variants, and biological noise, due to aging and therapy has potential to confound biomarker interpretation.

Currently, comprehensive methods to filter out non-tumor variants require genotyping the white blood cell (WBC) fraction of the paired plasma sample, which is a costly, complicated workflow.

A plasma-only, bioinformatics solution to identify non-tumor variants is needed for accurate biomarker assessments in the cell-free DNA (cfDNA).

Methods

Variant calls were obtained from >250,000 plasma samples comprising healthy donor, early and late-stage cancer patients sequenced on the Guardant360™, GuardantREVEAL™, GuardantOMNI™ and GuardantInfinity™ liquid biopsy panels as well as public tissue datasets.

The model was trained on paired plasma and WBC datasets and optimized with 10-fold cross-validation to produce a non-tumor and tumor variant classifier.

To validate these calls, an independent cohort of 72 paired plasma and WBC advanced cancer samples were genotyped on the GuardantInfinity™ assay. A cohort of 76 healthy donor samples, genotyped on the GuardantOMNI assay was also assessed.

Results

Model design: Leveraging internal database of >250K clinical patients

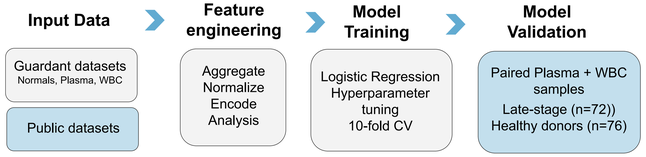
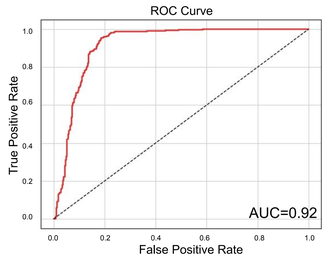


Figure 1. Model design. Features were engineered from Guardant internal and external public datasets and trained using 10-fold cross validation using multiple models. Only results from the Logistic Regression model are shown. Model validation was performed on an independent cohort of paired plasma and WBC late-stage samples sequenced on the GuardantInfinity™ panel, and healthy donors sequenced on the GuardantOMNI™ panel.

Model performance: High ROC AUC and accuracy for predicted calls

A. Validation: Late-stage cohort of 72 paired plasma and WBC samples



WBC \ Predicted	Non-Tumor	Tumor	Total
Non-Tumor	**318**	70	388
Tumor	15	**310**	325
Total	333	380	713

Sensitivity: 0.934
Specificity: 0.815
Precision: 0.819
Accuracy: 0.881

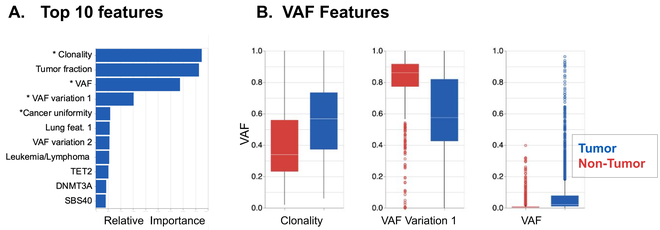
B. Validation: Healthy donor cohort of 76 paired plasma and WBC samples

All somatic SNV/Indels in healthy donors are considered non-tumor

Confirmation Method >	WBC Sequencing	Model Prediction
≥0.6% VAF	92.8% (64/69)	100% (69/69)
<0.6% VAF	42.5% (74/174)	100% (174/174)

Figure 2. Model performance. Predictions for tumor and non-tumor status were compared to WBC confirmation in A) 713 somatic SNV/Indels from 72 paired plasma and WBC GuardantInfinity™ samples and B) 243 somatic SNV/Indels from 76 paired plasma and healthy donors on GuardantOMNI™. Lower confirmation rate in WBC sequencing observed for low VAF variants (<0.6%) likely attributed to the limit of detection in for WBC variant calling and/or possible non-WBC lineage origin.

Feature importance: Assay-specific engineered features among most important



C. Cancer uniformity feature

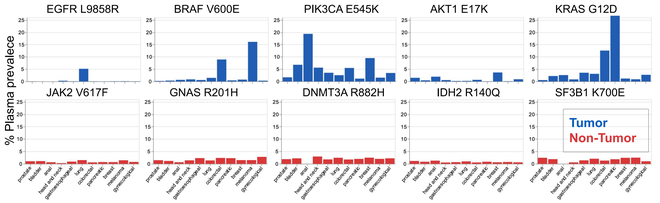


Figure 3. Feature importance and examples. A) Top 10 features ranked by relative importance on validation dataset. Individual gene names were included with one-hot encoding. B) Highly ranked engineered features include clonality, defined as VAF / tumor fraction as measured by methylation or max somatic VAF (left), VAF variation across timepoints (middle), mean percentage (right) and C) uniformity in variant prevalence across solid tumor cancer types in the Guardant plasma database.

Concordance in non-tumor predictions: gene prevalence and correlation with age

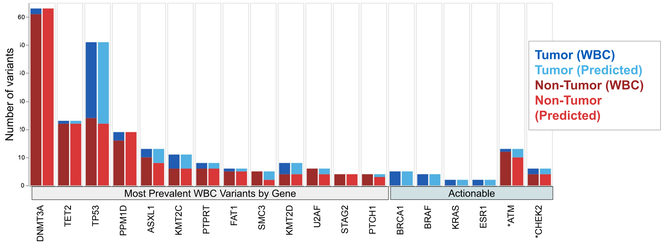
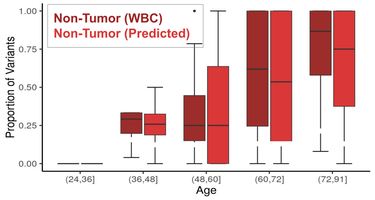


Figure 4. Number of variants within each gene predicted as non-tumor or tumor-derived as confirmed by WBC or model prediction in the late-stage validation cohort from Figure 2. Variant counts are shown for the genes with cfDNA variants most commonly confirmed in WBC samples, along with counts in clinically actionable gene (*BRCA1, BRAF, KRAS,* ESR1, *ATM, CHEK2*). Most frequent WBC-confirmed genes are consistent with previous reports, including high prevalence of clonal hematopoiesis in *ATM* and *CHEK2* (*)(Coombs et al. 2018).

Figure 5. Proportion of variants detected in WBC or predicted as non-tumor in late-stage validation cohort by age ranges. As expected from the literature (Jaiswal et al. 2014), variants predicted or confirmed as non-tumor are highly correlated with age.



References:
Coombs et al. Identification of Clonal Hematopoiesis Mutations in Solid Tumor Patients Undergoing Unpaired Next-Generation Sequencing Assays. Clin Cancer Res (2018).
Jaiswal et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med (2014).

Conclusions

Our bioinformatic model exhibits high sensitivity and specificity with WBC for discriminating tumor and non-tumor using only cfDNA.

Our bioinformatic model has improved sensitivity for identifying non-tumor variants over WBC sequencing at low VAFs (<0.6%).

In a paired plasma and WBC late stage cancer cohort, the majority of non-tumor variants were in known clonal hematopoiesis genes and variants of uncertain significance. No clinically actionable variants, except in *ATM* and *CHEK2*, were confirmed or annotated as non-tumor.