Specific and Differential Detection of Breast and Lung Cancer Via Cell-free RNA

Biology of Genomes Meeting
May 10 – 14, 2022
Virtual

David H. Burkhardt; Ruth E. Mauntz; Sarah M. Stuart; Monica P. Pimentel; Yiqi Zhou; Kelly McClintock; Jason Pallas; Ashton Teng; Matthew H. Larson
GRAIL, LLC, a subsidiary of Illumina, Inc., Menlo Park, CA, USA*.

INTRODUCTION

- Cancer is a leading cause of death in the United States as many cancers are detected too late.[1]
- Early cancer diagnosis can improve cancer survival.[2]
- Recently, a validated blood-based test has demonstrated detection of cancer signals through analysis of circulating nucleic acids.[3]
- Cell-free RNA (cfRNA) is a promising analyte for early cancer detection.
  - Evidence suggests cancer cells release cfRNA[4-8] into the blood by mechanisms besides cell death.[9]
  - cfRNA offers the potential to detect cancer in people with low tumor release of circulating tumor DNA.
- We previously performed an exploratory whole-transcriptome study to examine cfRNA expression in a subset of participants with cancer (breast or lung) and without cancer from the Circulating Cancer Genome Atlas (CCGA, NCT02889978).[10]
- Larson et al. (2021) identified 23 Dark Channel Biomarkers (DCBs), that are breast and lung cancer specific.[10]
  - DCBs are a set of transcripts that are: (1) not detected in cfRNA from participants without cancer (non-cancer samples) (2) recurrently expressed in cancer tissue, and (3) detectable in plasma of participants with cancer (cancer samples).
  - These DCBs exhibited cancer subtype specificity and were upregulated in tissue and plasma prepared from participants with triple-negative breast cancer (TNBC), hormone receptor positive breast cancer (HR+), and lung adenocarcinoma.

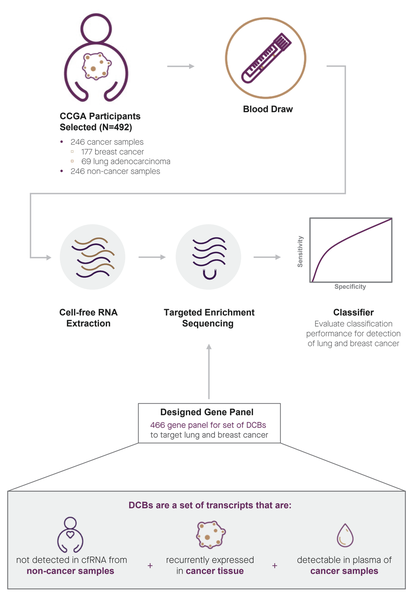
OBJECTIVES

We conducted a study with the following aims:

- Validate a targeted set of DCBs selected based on high tumor tissue expression.
- Determine classification performance for detection of lung and breast cancer status from plasma samples using these targeted DCBs.

STUDY DESIGN

Figure 1. Study Design Overview

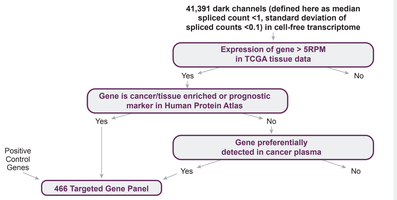


METHODS

Gene Panel Design

- We designed a 466-gene panel to target putative breast cancer (HR+ and TNBC) and lung cancer (adenocarcinoma) DCBs in a cohort of 492 participants selected from CCGA based on cancer type and cell-free DNA (cfDNA) tumor fraction (TF) levels (Figure 2).
  - Panel design integrated information from The Cancer Genome Atlas (TCGA) tumor tissue samples and GRAIL cfRNA data.
  - Genes were selected from the panel based on evidence of low expression in plasma collected from participants without cancer, and high expression in plasma from participants with cancer and in tumor tissue.
- To evaluate these cfRNA DCBs as potential biomarkers for cancer detection, we quantified DCB levels in samples from participants with cancer and without cancer using targeted cfRNA sequencing.

Figure 2. Targeted Panel Design Utilized TCGA Expression to Identify Dark-Channel Genes That May Be Preferentially Expressed in Plasma of Participants With Cancer



Sample Selection

- We selected samples from 492 participants in the CCGA cohort.
  - 246 samples were from participants diagnosed with cancer at the time of blood draw (177 breast [139 HR+; 38 TNBC]; 69 lung adenocarcinoma) (Table 1).
  - 246 samples were from frequency-age-matched participants without cancer (non-cancer samples).
- "High tumor fraction" (high TF) samples had >0.3% tumor fraction and "low tumor fraction" (low TF) samples had <0.3% tumor fraction, as estimated by methyl variants in cfDNA.[11]

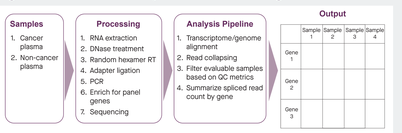
Table 1. Sample Characteristics

		Lung Cancer (n=69)	Breast Cancer (n=177)	Non-Cancer (n=246)
Tumor Fraction	High TF	14	23	—
	Low TF	55	154	—
Clinical Stage	Stage I	38	91	—
	Stage II	3	71	—
	Stage III	11	8	—
	Stage IV	17	7	—
Sex	Male/Female	38/31	1/176	83/163
Age Group	< 50	3	49	79
	50-65	28	75	110
	> 65	38	53	57

Sample and Data Processing

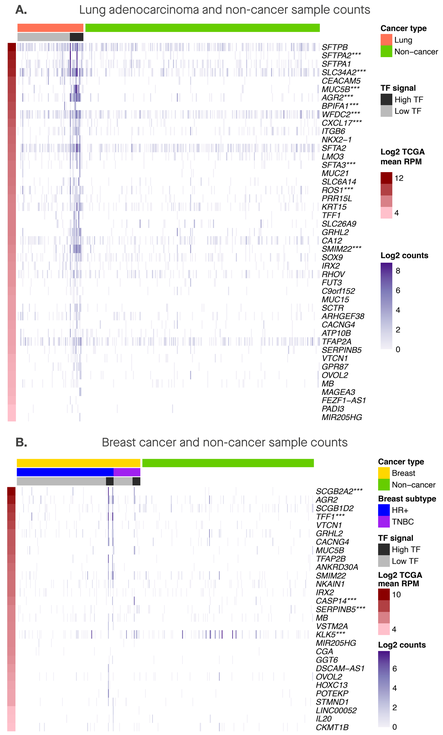
- RNA-seq libraries were prepared from isolated plasma as previously described[10]. Libraries were enriched with a hybridization capture approach. Enriched libraries were sequenced at a depth of ~100M paired-end reads per sample and analyzed using a custom bioinformatics pipeline that generated unique molecular identifier-collapsed counts for each gene on a sample-by-sample basis.
- All downstream analyses relied on the use of RNA reads with evidence of splicing, defined as read pairs where at least 1 read overlapped an exon-exon junction. These reads are unique to RNA and help to filter DNA-derived reads.
- Classification was evaluated on breast cancer samples against non-cancer samples and lung cancer samples against non-cancer samples with a leave-one-out method.
  - For each classification fold, features were selected by the cancer vs. non-cancer Mann-Whitney *p*-value.
  - For each fold, the probability of cancer of the left-out sample was estimated by standardizing evidence of expression for each selected panel gene by expression observed in non-cancer plasma samples in the inner fold, and aggregating evidence for expression observed on all selected panel genes.

Figure 3. Overview of Assay and Data Processing Workflow



KEY RESULTS: cfRNA ENABLES SPECIFIC AND DIFFERENTIAL DETECTION OF BREAST AND LUNG CANCER

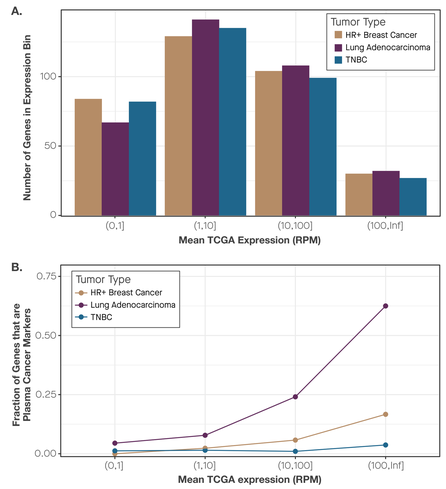
Figure 4. High-Specificity Biomarkers Were Identified in (A) Lung and (B) Breast Cancer



*Each column depicts one cfRNA sample, and each row depicts one gene. For each gene, the relationship between cfRNA expression and cancer status was evaluated by Mann-Whitney test within each cancer type, considering only high tumor fraction samples and non-cancer samples. The most significantly expressed DCBs are shown, for lung cancer (Bonferroni-corrected p-value<0.0005) and breast cancer (Bonferroni-corrected p-value<0.05). Log2-transformed spliced counts for each gene in each sample are shown in purple. Genes previously identified as DCBs that pass significance threshold of <0.05 are included and are marked with '***'.*

- We identified DCB transcripts that enabled detection of both breast and lung cancers (Figure 4).
  - In total, 10 of the 12 lung cancer specific DCBs previously identified[10] were validated as significantly lung cancer specific features in this cohort and shown above (Bonferroni-corrected *p*-value<0.05; *AGR2, BPIFA1, CXCL17, MUC5B, ROS1, SFTA3, SFTPA2, SLC34A2, SMIM22, WFDC2*).
  - 5 of the 12 breast cancer specific DCBs previously identified[10] were significantly expressed in breast cancer samples (Bonferroni-corrected *p*-value<0.05; *CASP14, KLK5, TFF1, SCGB2A2, SERPINB5*).

Figure 5. Panel Genes With Greater Cancer Detection Performance Were More Highly Expressed in Tumor Tissue for Lung Adenocarcinoma and HR+ Breast Cancer

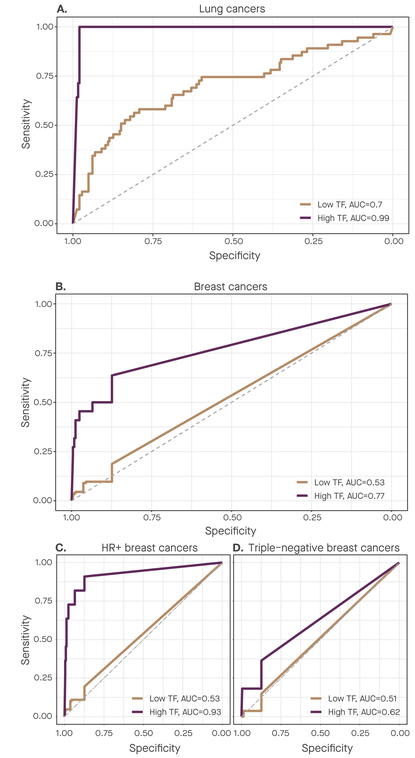


(A) For DCB panel genes (n=396), we determined the mean expression in TCGA primary tumor tissue samples for HR+ breast cancer, TNBC, and lung adenocarcinomas. Panel genes are binned by TCGA expression for each tumor type.

(B) For each DCB panel gene, the relationship between cfRNA expression and cancer status was evaluated by Mann-Whitney test within each cancer type, considering only high TF samples and non-cancer samples. All genes with Bonferroni-corrected p-values < 0.05 were considered "plasma cancer markers". The fraction of panel genes that are plasma cancer markers is plotted for each TCGA expression bin.

- The distribution of TCGA expression on panel genes is similar for each cancer type (Figure 5A).
- The number of panel DCBs that are plasma cancer markers differs substantially by cancer type (Figure 5B).
  - For lung adenocarcinoma, 60% (20/32) of genes with mean expression in TCGA tumor tissue > 100 RPM are plasma cancer markers. By contrast, only 5% (3/67) of genes with mean TCGA tissue tumor expression < 1 RPM are plasma cancer markers.
  - For HR+ breast cancer 17% (5/30) of genes with mean expression in TCGA tumor tissue > 100 RPM are plasma cancer markers, in contrast to 0% (0/84) of genes with mean TCGA tumor tissue expression < 1 RPM.
  - For TNBC, few cfRNA plasma cancer markers exist, with only 3% (1/28) of of genes with mean expression in TCGA tumor tissue > 100 RPM plasma cancer markers, and 1% (1/82) of genes with mean TCGA tumor tissue expression < 1 RPM.
- This analysis across tumor types indicates that:
  - TCGA tumor tissue expression is valuable for identifying probable plasma cancer markers for some cancer types (lung adenocarcinoma, HR+ breast cancer)
  - For TNBC, even the most highly expressed RNAs are not detected in plasma, suggesting that some tumor types may not release detectable RNA into blood.

Figure 6. cfRNA Enables Specific Detection of Lung and Breast Cancers



Receiver operating characteristic (ROC) curves for the (A) lung vs. non-cancer and (B) breast vs. non-cancer classifiers are shown.

- For lung cancer, we observed 100% (95% CI: 77–100%) and 25% (95% CI: 15–39%) sensitivity at 95% specificity for high and low TF samples, respectively (Figure 6A).
- For breast cancers, we observed 43% (95% CI: 24–68%) and 9% (95% CI: 5–15%) sensitivity at 95% specificity for high and low TF samples, respectively (Figure 6B).
- Breast cancer performance was subtype specific, with higher detection in HR+ breast cancers (78% [95% CI: 44–100%] and 11% [95% CI: 4–17%] at 95% specificity for high and low TF samples, respectively) compared to TNBC (22% [95% CI: 0–10%] and 3% [95% CI: 0–56%] at 95% specificity for high and low TF samples, respectively) (Figure 6C and 6D).

CONCLUSIONS

cfRNA detected breast and lung cancers with differential performance at 95% specificity.

- DCBs that were more highly expressed in TCGA tumor tissue samples were more informative for cancer classification than DCBs with low expression, supportive of their origination from cancer tissue.
- 10/12 lung cancer markers and 5/12 breast cancer markers identified in our prior investigation were preferentially detected in cancer participants in this cohort.
  - While the discovery cohort in our prior investigation consisted of participants with stage III breast and lung cancers, the majority of cancer samples in this cohort were from participants with stage I and II cancer to determine sensitivity of cfRNA in early cancer detection. This and other factors beyond staging may affect DCB abundance in the plasma.
- cfRNA detection differed substantially between lung adenocarcinoma and breast cancer subtypes, even in samples with high cfDNA tumor fraction.
  - The difference is consistent with differential shedding of cfRNA relative to cfDNA for different tumor types.

Further studies will be needed to understand the utility of cfRNA in multi-cancer early detection.

References

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. *CA Cancer J Clin.* 2022;72(1):7-33.
2. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence - SEER 18 Regs Research Data, Nov 2018 Sub. Includes persons aged 50-79 diagnosed 2006-2015. http://seer.cancer.gov/csr/1975_2015/, based on November 2017 SEER data submission, posted to the SEER website April 2018.
3. Klein E, Richards D, Cohn A, et al. *Ann Oncol.* 2021;32(9):1167-1177.
4. Lo K-W, Lo YM, Leung SF, et al. *Carcinoma Clin Chem.* 1999;45:1292–1294.
5. Kopreski MS, Benko FA, Kwak LW, Gocke CD. *Clin Cancer Res.* 199;5:1961–1965.
6. Kopreski MS, Benko FA, Gocke CD. *Ann NY Acad Sci.* 2001;945:172–178.
7. Castellanos-Rizaldos E, Zhang X, Tadigotla VR, et al. *Oncotarget.* 2019;10:2911–2920.
8. Perhavec A, Cerkovnik P, Novakovic S, Zgajnar J. *Neoplasma.* 2008;55:549–555.
9. Skog J, Würdinger T, van Rijn S, et al. *Nat Cell Biol.* 2008;10:1470–1476.
10. Larson MH, Pan W, Kim HJ, et al. *Nature Communications.* 2021;12:2357.
11. Melton C, Singh P, Venn O, et al., Tumor methylation patterns accurately measure tumor fraction in cell-free DNA. Poster presented at: *ASCO*; May 29-31, 2020.

Disclosures

Study funded by GRAIL, LLC., a subsidary of Illumina Inc.* All authors are employees of GRAIL, LLC. with equity in the company*.

*GRAIL, LLC., a subsidiary of Illumina Inc., currently held separate from Illumina Inc. under the terms of the Interim Measures Order of the European Commission dated 29 October 2021.

Acknowledgements

Funded by GRAIL, LLC, a subsidiary of Illumina, Inc*. This work incorporates data generated by the TCGA Research Network: https://www.cancer.gov/tcga.