Detection of Tissue-specific RNA in the Plasma of Cancer Patients: A Circulating Cell-free Genome Atlas (CCGA) Substudy

2019 Cold Spring Harbor Laboratory Meeting:
The Biology of Genomes
May 7–11, 2019
Cold Spring Harbor, NY

Matthew H. Larson, PhD; Hyunsung J. Kim, PhD; Wenying Pan, PhD; Sarah Stuart, BS; Yiqi Zhou, MS; Archana Shenoy, PhD; Monica Pimentel, MS; Per Knudsgaard, MSc; Vasiliki Demas, PhD; Earl Hubbell, PhD; Alexander M. Aravanis, MD, PhD; Arash Jamshidi, PhD

GRAIL, Inc., Menlo Park, CA

INTRODUCTION

- Cell-free RNA (cfRNA) is a promising analyte for cancer detection, but a comprehensive assessment of cfRNA is lacking.
- To characterize tumor-derived RNA in plasma, we performed an exploratory analysis from a Circulating Cell-free Genome Atlas (CCGA; NCT02889978) substudy to examine cfRNA expression in participants with and without cancer.
- In this analysis, we focused on breast, lung, and colorectal cancers due to their high incidence in the general population and in CCGA.

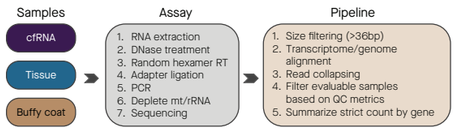
OBJECTIVE

- We conducted a study with the following aims:
  - Characterize cfRNA signal across the whole transcriptome in a non-cancer cohort.
  - Identify biological signals in cfRNA that may be useful for cancer classification.
    - Determine baseline and noise for these signals.
  - Determine concordance of cfRNA signal with RNA signal in tumor tissue.

METHODS

- We selected 210 participants from the previously-described CCGA training set.[1]
  - 98 participants diagnosed with stage III cancer at the time of blood draw (breast [47], lung [32], colorectal [15], and anorectal [4]). Stage III samples were selected to maximize signal in the blood and avoid confounding signal from potential secondary metastases.
  - 112 non-cancer participants frequency-age-matched to the cancer group.
- We extracted cell-free nucleic acids from participant plasma, DNase treated samples to remove cfDNA and genomic DNA, and performed reverse transcription (RT) using random hexamer primers to capture the whole transcriptome for each study participant.
- The resulting cDNA was converted into DNA libraries, amplified, and depleted of abundant sequences arising from ribosomal, mitochondrial, and blood-related transcripts, such as globins.
- The resulting whole-transcriptome RNA-seq libraries were sequenced at a depth of ~750M paired-end reads per sample and analyzed using a custom bioinformatics pipeline that generated unique molecular identifier (UMI)-collapsed counts for each gene on a sample-by-sample basis.
- This same procedure was used to create and analyze RNA-seq libraries from matched buffy coat and tumor tissue RNA when available.
- Due to the presence of residual DNA contamination, all downstream analyses relied on the use of strict RNA reads, defined as read pairs where at least one read overlapped an exon-exon junction. Figure 1 shows a summary of the end-to-end workflow.

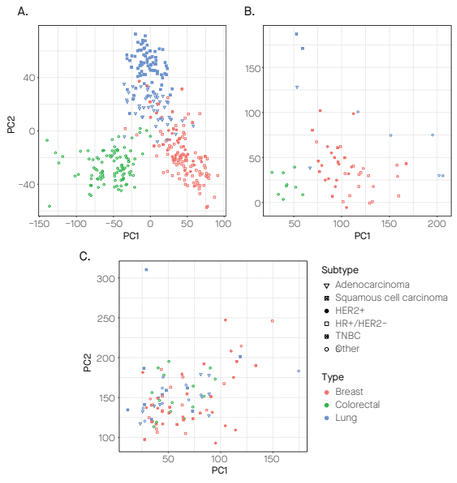
Figure 1. Overview of Assay and Data Processing Workflow



RESULTS

- We set out to determine whether tissue and cell-free RNA samples from different cancer types in our cohort were broadly distinguishable based on their gene expression profiles using principal component analysis (PCA).
- We compared our data to RNA samples from The Cancer Genome Atlas (TCGA) (Figure 2A).
- When we projected CCGA tumor tissue RNA-seq data onto the principal components derived from TCGA tumor tissue RNA-seq data, the CCGA tumor tissue samples were separable by cancer type (Figure 2B).
  - This suggests that the expression profiles of CCGA and TCGA tumors are similar in spite of differences in sample collection/handling/library preparation, and validates the analytical approach.
- A projection of cancer cfRNA samples from the CCGA cohort onto the principal components derived from TCGA tumor tissue RNA-seq data showed no separation of the sample by cancer type (Figure 2C), implying that cancer type was not the dominant source of variance in cfRNA.
- Taken together, these results motivated the development of different feature selection methods to extract tumor-derived signals present in the blood.

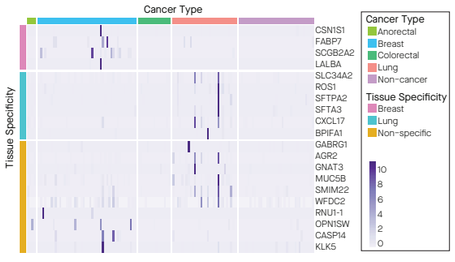
Figure 2. Comparison of TCGA and CCGA RNA-seq Data



A) PCA of stage III TCGA FFPE tissue RNA-seq data. B) CCGA tumor tissue RNA-seq data projected on TCGA PCA axes. C) CCGA cancer cfRNA RNA-seq data projected on TCGA PCA axes. Gene expression levels used for PCA are in reads per million (RPM). PC, principal component.

- The majority of cfRNA in plasma is thought to originate from healthy immune cells. As such, we treated these transcripts as background noise and focused on tumor-derived cfRNA as a source of cancer signal.
- Our analysis identified two classes of genes in cfRNA data: "dark channels" and "dark channel biomarkers" (DCB).
  - Dark channels are genes that were not detected (median gene expression was zero) in the cfRNA of non-cancer participants.
    - Of 57,783 annotated genes, 39,564 (68%) were identified as dark channels.
  - DCB genes met two additional criteria:
    1) Gene was expressed in more than one participant in the cancer cohort, and 2) gene expression was up-regulated in the cancer group.
    - 9 DCB genes were identified for breast cancer: RNU1-1, CSN1S1, FABP7, OPN1SW, SCGB2A2, LALBA, CASP14, KLK5, and WFDC2.
    - 12 DCB genes were identified for lung cancer: SLC34A2, GABRG1, ROS1, AGR2, GNAT3, SFTPA2, MUC5B, SFTA3, SMIM22, CXCL17, BPIFA1, and WFDC2.
    - No DCB genes were identified for colorectal or anorectal cancers.
- DCB genes exhibited several distinct characteristics, including being enriched for tissue-specific genes (Figure 3).
  - Among the 57,783 annotated genes, 0.3% were lung-specific and 0.2% were breast-specific.
  - In comparison, 50% of the lung DCB genes were lung-specific, and 44% of the breast DCB genes were breast-specific (as defined by the protein atlas database).[2]

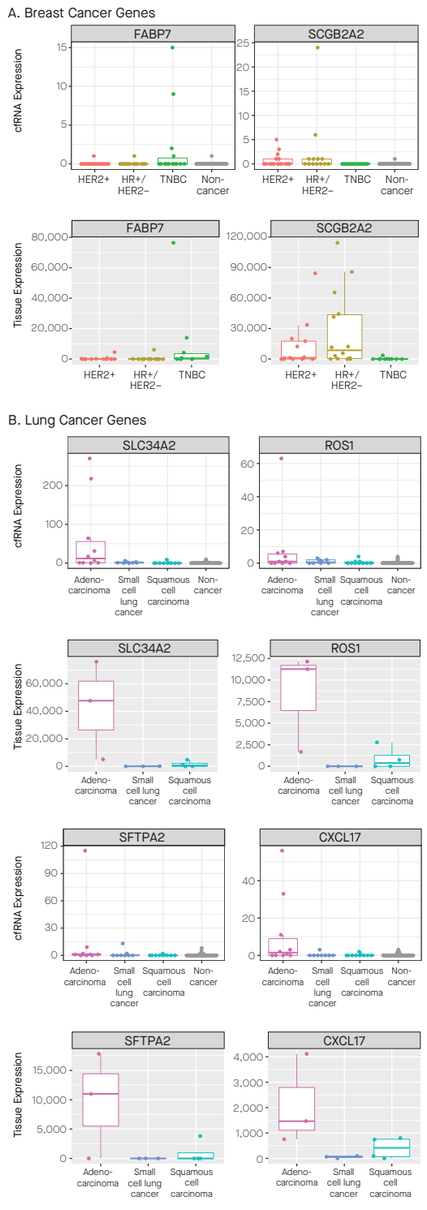
Figure 3. Heatmap of Dark Channel Biomarker Genes



Each column depicts one cfRNA sample, and each row depicts one gene. The color of the rows encodes tissue-specificity. The color of the columns encodes the sample groups. Color scale represents read counts (centered and scaled for each gene).

- In addition, some DCB genes were subtype-specific biomarkers that were only detected in certain cancer subtypes (Figure 4).
  - FABP7 was only detected in triple negative breast cancer (TNBC) samples (Figure 4A).
  - Conversely, SCGB2A2 was not detected in TNBC, but was detected in HER2+ and HR+/HER- breast cancer samples (Figure 4A).
  - SLC34A2, ROS1, SFTPA2 and CXCL17 genes were detected in cfRNA of lung adenocarcinoma patient samples but not in squamous cell carcinoma patient samples (Figure 4B).
  - These genes also exhibited the same subtype-specific expression pattern in matched tumor tissue (Figure 4A-B).

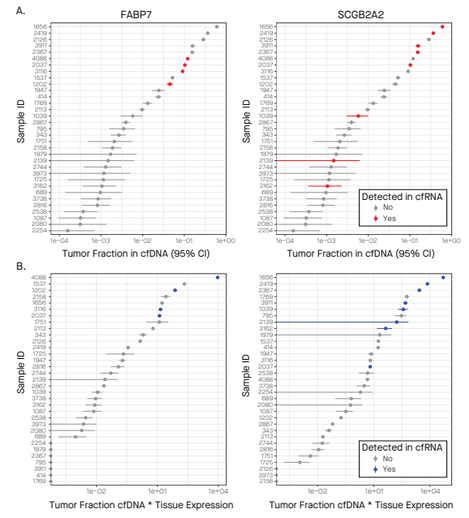
Figure 4. Expression of Subtype-specific DCB Genes in cfRNA and Tissue



A) Breast DCB genes and B) lung DCB genes. The cfRNA and tissue expression levels plotted on the y-axis are normalized read counts.

- In order to determine the source of tumor-associated transcripts in the blood, we assessed concordance between cfRNA and tumor tissue RNA for DCB genes.
- We observed high concordance between cfRNA and tumor tissue expression (Figure 5A). Genes not detected in the tumor tissue were unlikely to be detected in the matched cfRNA sample, and genes detected in the tumor tissue were more likely to be detected in the matched cfRNA sample.
- Additionally, tumor content, defined as the product of cfDNA tumor fraction for a given patient and the gene expression in matched tumor tissue, was a strong predictor of the detectability of a DCB gene in the cfRNA of breast cancer patients (Figure 5B).

Figure 5. Detectability of Two Breast DCB Genes for Breast Cancer Samples with Matched Tumor Tissue



*Sample IDs plotted as a function of tumor content (tumor fraction * tumor tissue expression). Tumor fraction in cfDNA was measured from SNV allele fractions from the cfDNA enrichment assay. Tissue expression was measured from RNA-seq data of matched tumor tissue.*

CONCLUSIONS

- The majority of annotated transcripts are not found in cfRNA from non-cancer subjects. These dark channel biomarkers (DCBs) represent genes that have the potential for high signal-to-noise in cancer patients.
- DCB signal is correlated with tumor content (defined as the product of tumor fraction in the blood and RNA expression in the tissue).
- cfRNA DCBs were identified in cancer participants in a tissue- and subtype-specific manner.
- We observed cases where high tumor tissue expression led to DCB signal amplification and enabled detection of cancer in patients with low cfDNA tumor fraction.
- Taken together, these data suggest that tissue-specific transcripts have potential for use in blood-based multi-cancer detection.

References

1. Klein et al. Development of a comprehensive cell-free DNA (cfDNA) assay for early detection of multiple tumor types: The Circulating Cell-free Genome Atlas (CCGA) study. ASCO (2018).
2. Uhlén et al. Tissue-based map of the human proteome (www.proteinatlas.org). *Science* doi:10.1126/science.1260419 (2015).

Disclosures: Study funded by GRAIL, Inc. All authors are employees of GRAIL, Inc. with equity in the company. AJ is a shareholder of Illumina. MHL is a shareholder of Illumina, Johnson & Johnson, Procter & Gamble. HJK is a current shareholder of Illumina, Intellia Therapeutics, and Bristol-Myers Squibb, and is a former shareholder of Editas, Pacific Biosciences, and NantHealth. SS is a shareholder of Illumina. VD is a shareholder of T2Biosystems and Alphabet.