2225

Limitations on Mutation Detection for Early Detection of Cancer

Imran S. Haque,[1] Gabriel Otte,[1] Olivier Elemento[2]

[1]Freenome Inc, South San Francisco, CA; [2]Weill Cornell Medical College, New York, NY

INTRODUCTION

- The presence of tumor-derived genomic alterations in cell-free DNA (cfDNA) circulating in the blood of patients with cancer[1] has led to development of blood-based assays for tumor genomic profiling
- Circulating tumor (ctDNA) is present in many different malignancies, making it potentially useful for the early noninvasive detection of cancer (**Figure 1**)
- While current methods relying on deep sequencing of a few genes can detect ctDNA in some patients with early-stage cancers,[2,3] limited sensitivity has so far precluded their use for screening
- For detection of tumor derived mutations, it is critical to estimate the tumor fraction (ratio of ctDNA to cfDNA) and the tumor variant allele frequency (VAF; fraction of cfDNA molecules bearing a mutation known to be present in a tumor). Both tumor fraction and VAF vary widely by tumor type and stage[4]
- Aravanis and colleagues[5] recently proposed that the advances required to enable a ctDNA-based early detection test are around 100× more sequencing coverage, improvements in variant interpretation, and sensitivity to VAF of 0.01% or lower

OBJECTIVES

- Reanalyze published data on the expected ctDNA allele fraction in early-stage cancer and assess the feasibility of ctDNA mutational assays for early cancer detection based on their physiologic and economic requirements
- Review alternative biologic signals of early cancer and the potential of machine learning to integrate these signals into reliable diagnostics

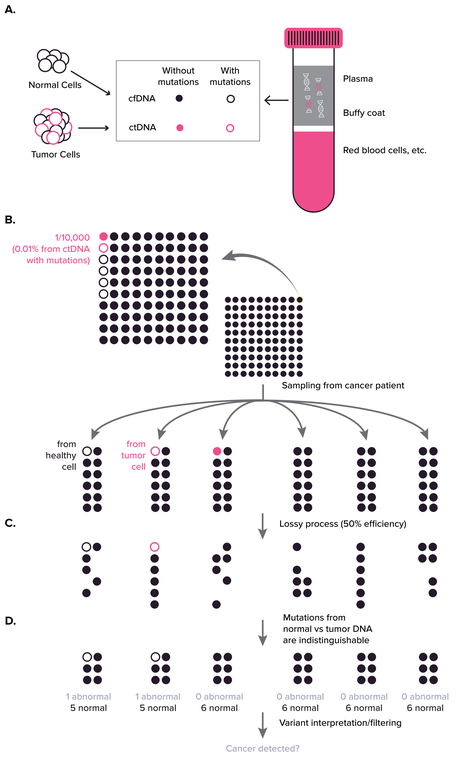
METHODS

- A binomial model was used to assess depth and input requirements, with parameters derived from published data on cfDNA sequencing[2]
- Model parameters:
  - No more than 5% of samples may fail because of insufficient cfDNA quantity
  - 95% sensitivity to detect one read from any cancer-derived allele, assuming that one is present in the sample
  - 50% process efficiency: Half of the cfDNA molecules in the input blood sample are represented in the sequencer output
  - 5× oversampling in sequencing for error correction
  - 100% on-target rate in target enrichment
  - "$1000 genome" sequencing costs: US $1000/(30 × 3 Gbp) of sequencing bandwidth
  - Only sequencing costs computed; all other costs (e.g., labor, equipment, facilities, depreciation) accounted at $0
  - Panel expansion neither reduces input requirements nor increases sequencing requirements

ctDNA Mutation Detection Is Statistically Limited by Input Volume

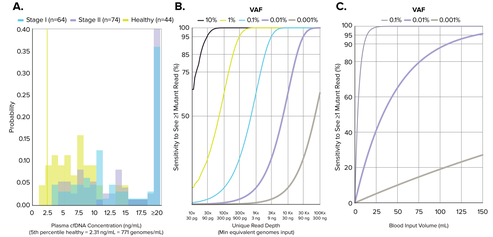
- The binomial model, which analyzed a range of VAFs and sequencing depths, suggested that 3000× unique coverage is required for 95% sensitivity at 0.1% VAF, and 30,000× is required for 95% sensitivity at 0.01% VAF (**Figure 2**)
- The model was validated using real-world VAF and unique coverage data reported by Phallen et al[2] in 190 patients (**Figure 3**)
  - No patients were observed below the modeled boundary (95% sensitivity), and patients with VAF below 0.1% had higher depth of unique coverage (most >5000)
  - Depth (the number of molecules assayed) is a significant factor when detecting low-frequency mutations

FIGURE 1. Challenges of ctDNA Mutation Analysis. cfDNA Comprises Short DNA Fragments Present in the Blood, and May Include DNA Derived From Tumor Cells (ctDNA) and Normal Cells[6]



(A) Illustration of the sources of cfDNA and that somatic mutations (open circles) can arise from both tumor and nontumor cells. (B) Binomial sampling: a finite-size sample is drawn from the entire population of cfDNA, implying that some samples will contain no mutations. Shows 0.01% VAF to scale. (C) Process efficiency: Because no lab process is 100% efficient, some molecules present in the input tube may not make it to the sequencer (and therefore the data readout), creating another opportunity for dropout of rare variants. (D) Variant interpretation: Although 2 mutations were read out at the end of the process, the mutation tissue of origin (healthy or tumor) is indistinguishable in the actual sample, making it difficult to interpret whether the sample actually contained a tumor-derived mutation.

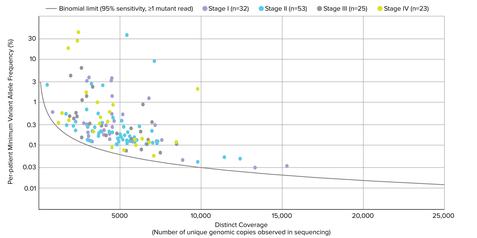
FIGURE 2. Binomial Model for ctDNA Sequencing. cfDNA Is Less Frequent in Healthy Individuals Than Those With Early-Stage Cancer



(A) Distribution of cfDNA concentrations observed in healthy individuals and stage I/II cancer patients.[2] The green line = 5th percentile of the distribution of healthy individuals, cfDNA concentration. (B) Upper bound on sensitivity to detect one mutant molecule as a function of sequencing depth and VAF; note logarithmic x-axis. Also shown is the minimum amount of unique DNA input required for sequencing, assuming 3 pg haploid genome mass and 100% process efficiency. (C) Sensitivity as a function of blood input volume, assuming 2.3 ng cfDNA/mL plasma, plasma volume 55% of blood volume, and 50% process efficiency.

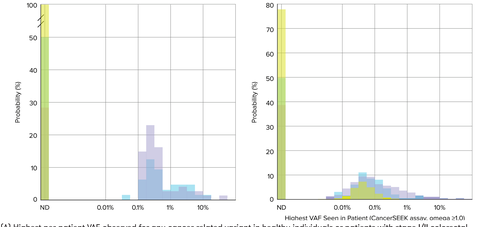
- Although it is possible that the patients without detected ctDNA would have had mutant alleles detected at greater depth, it is not possible to arbitrarily increase sequencing depth because unique coverage is limited by the number of input molecules
  - 30,000× unique coverage requires at least 30,000 distinct copies of the tested region in the input sample
  - Process efficiency (the probability a molecule in an input sample is represented in the output) is below 1% for typical sequencing assays,[7] which means some molecules present in the original sample will be lost during processing
  - Even with 50-fold improvement in efficiency (50% efficiency), 60,000 input molecules would be required to achieve 30,000× unique depth
- The model suggests that even with 50-fold improvement in efficiency, 150 mL of blood would be required to achieve 95% detection of 0.01% VAF mutations (**Figure 2C**), which is impractical for a general-population screening test

FIGURE 3. Validation of Binomial Model for ctDNA Sequencing



Comparison of observed VAF in data reported by Phallen et al[2] to the binomial model. Dots correspond to the VAF of the lowest-frequency cancer-derived variant detected and the unique depth of coverage for that patient. Gray curve shows the VAF expected to be detected with 95% confidence by the binomial model.

FIGURE 4. VAF Is Below 0.01% in a Substantial Fraction of Healthy Individuals and Stage I/II Cancer Patients



(A) Highest per-patient VAF observed for any cancer-related variant in healthy individuals or patients with stage I/II colorectal, breast, ovarian, or lung cancer, as measured by the targeted error correction sequencing (TEC-Seq) protocol.[2] Samples listed as "ND" had no cancer-derived alleles observed. In patients with multiple cancer alleles detected in plasma, the highest VAF is shown. Healthy samples had zero cancer-related variants detected. (B) Highest per-patient VAF observed for any cancer-related variant in healthy individuals or patients with stage I/II colorectal, breast, esophageal, liver, ovarian, lung, stomach, or pancreatic cancer based on sequencing data from Cohen and colleagues.[3] Samples listed as "ND" had no cancer-derived alleles observed (50% of stage I/II cancer patients and almost 80% of healthy patients). Variants with an omega score below 1.0 were excluded based on the analysis described in Cohen et al.[3]
ND = not detected.

ctDNA Mutation Detection Is Biologically Limited by Somatic Heterogeneity

- As large-scale cancer genomics projects have revealed that most tumors contain multiple somatic variants,[8] it may be possible to reduce input and sequencing depth requirements by detecting any of a large number of tumor-specific mutations rather than focusing on one specific mutation (the assumption of the binomial model)
  - For example, detecting any of 10 independent VAF=0.01% mutations would have the same sampling probability as detecting a single 0.1% variant
- However, low levels of cancer-associated mutations are observed in healthy individuals[9] and may increase with age,[10,11] which could complicate interpretation when mutations are detected
- VAFs observed using sequencing data from Cohen and colleagues[3] confirmed that even using a stringent filtering threshold such that nearly half of early-stage cancers had no detected ctDNA, over 20% of healthy individuals had a low-level "cancer-related" mutation (**Figure 4**)
- To mitigate false positives, many of the detected alleles may need to be filtered out and more than one mutation detection event would likely be required
- Even successful filtering would imply that the binomial limits derived above are highly optimistic

TABLE 1. Assay Requirements and Cost for Tumor Liquid Biopsy and Mutation-Based Early Cancer Detection

	VAF 95% Sensitivity	Corrected Depth	Raw Depth	Input Volume (blood)	Sequencing Cost by Panel Size: *TEC-Seq*[2] 58 genes 81 kb	*Razavi et al*[11] 508 genes 2000 Kb	*WES* ≈20k genes 50,000 Kb
Tumor liquid biopsy	0.1%	3000×	15,000×	15 mL	$14	$340	$8300
Early cancer detection	0.01%	30,000×	150,000×	150 mL	$140	$3400	$83,000

WES, whole exome sequencing.

ctDNA Mutation Detection Is Economically Limited by Sequencing Cost

- Depth requirements from the binomial model enabled estimation of the costs of a mutation detection assay
- The input volume required and corresponding sequencing cost of a mutation-calling ctDNA-based early detection assay under highly conservative assumptions is estimated in **Table 1**
- The model suggests that early detection may be infeasible:
  - Small panels (e.g., the 81-kb TEC-Seq panel) have achievable sequencing costs, but have input volumes of >150 mL of blood, which are likely prohibitive
  - Larger panels (e.g., the 2-Mb panel reported by Razavi et al[12]) have significantly higher sequencing costs ($3400)

Alternatives to ctDNA Mutation Detection

- There are a number of blood-based analytes that may be useful for early cancer screening (**Table 2**)
- Computational integration of these multi-analyte signals may provide improved power for phenotype classification[3]
- Repeat screening also offers a unique opportunity to improve accuracy with longitudinal data on individuals

TABLE 2. Biologic Components Other Than ctDNA With Potential for Cancer Screening

Category	Analyte	Explanation
Tumor-derived material	Tumor-derived proteins and RNA[3]	Protein and RNA are present at higher copy number than DNA, potentially enabling detection via nonzero count even at low concentration
	Exosomes/microvesicles/ circulating tumor cells or cell clusters[13,14]	Tumor-derived bodies may contain macromolecular markers in sufficiently high concentration to detect efficiently
	Platelets[15]	Platelets contain proteins and RNA that function in immune signaling pathways, whose composition may vary in the presence of cancer
Immuno-surveillance	Cytokines, antibodies, and other immune signaling molecules[16,17]	Differential cytokine and autoantibody production has been observed in cancer patients as a component of the immune response to a tumor
	Immune cell subpopulations[18]	Differential composition of immune cells may indicate the presence of cancer and inform prognosis
Tumor microenvironment and host response	cfDNA[6,19]	Patterns in cfDNA beyond sequence variation, including epigenetic modifications and fragmentation, may serve as a marker for host gene expression

CONCLUSIONS

- This analysis demonstrated that for 3 reasons, tests using detection of tumor-derived mutations in cfDNA alone are unlikely to achieve the clinical and operational performance characteristics, including sensitivity and specificity, required for population screening due to:
  - Statistical limitations driven by the physiology of cfDNA
  - Biologic limitations driven by recently discovered somatic heterogeneity in healthy tissue
  - Economic limitations related to the costs and reimbursement for such an assay
- Further research using multi-analyte and/or longitudinal analysis methods holds promise for the development of clinically useful and economically viable tests for the early detection of cancer

ACKNOWLEDGMENTS

We thank Christina Curtis, Adam Drake, Kate Niehaus, Girish Putcha, and David Weinberg for their suggestions and feedback on the analysis. Medical writing and editorial support, which was in accordance with Good Publication Practice (GPP3) guidelines, was provided by Amy Agbonbhase, PhD of The Lockwood Group (Stamford, CT), and funded by Freenome Inc.

REFERENCES

1. Sorenson GD, et al. *Cancer Epidemiol Biomarkers Prev.* 1994;3:67-71.
2. Phallen J, et al. *Sci Transl Med.* 2017;9(403).
3. Cohen JD, et al. *Science.* 2018;359(6378):926-930.
4. Newman AM, et al. *Nat Med.* 2014;20(5):548-554.
5. Aravanis AM, et al. *Cell.* 2017;168:571-574.
6. Snyder MW, et al. *Cell.* 2016;164:57-68.
7. Fu GK, et al. *Proc Natl Acad Sci USA.* 2014;111(5):1891-1896.
8. Ciriello G, et al. *Nat Genet.* 2013;45(10):1127-1133.
9. Martincorena I, et al. *Science.* 2015;348:880-886.
10. Jaiswal S, et al. *New Engl J Med.* 2014;371:2488-2498.
11. Razavi P, et al. *J Clin Oncol.* 2017;35(suppl):abstr 11526.
12. Razavi P, et al. *J Clin Oncol.* 2017;35(suppl):abstr LBA11516.
13. Hannafon BN, et al. *Breast Cancer Res.* 2016;18(1):90.
14. Xu YF, et al. *Oncotarget.* 2017;8(14):77028-77040.
15. Best MG, et al. *Cancer Cell.* 2015;28(5):666-676.
16. Hanash SM, et al. *Nature.* 2008;452:571-579.
17. Chen H, et al. *Oncotarget.* 2016;7(13):16420-16432.
18. Gentles AJ, et al. *Nat Med.* 2015;21(8):938-945.
19. Li W, et al. *Cell Res.* 2017;27:1243-1257.

Presented at the American Association for Cancer Research (AACR) Annual Meeting, April 14-18, 2018; Chicago, Illinois, USA.