Cancer detection: Seeking signals in blood:

Combining gene mutations and protein biomarkers for earlier detection and localization

Mark Kalinich[1,2] and **Daniel A. Haber**[1,2]

[1]Massachusetts General Hospital Cancer Center Harvard Medical School, Boston, MA 02467, USA.

[2]Howard Hughes Medical Institute, Bethesda, MD 20815, USA.

Most cancers are detected when thy cause symptoms that lead to medical evaluation. Unfortunately, in too many cases this results in diagnosis of cancers that are locally invasive or already metastatic and hence no longer curable with surgical resection or radiation treatment. Medical therapies, which might be curative in the setting of minimal tumor burden, typically provide more limited benefit in more advanced cancers, given the emergence of drug resistance (1). On page 926 of this issue, Cohen et al. (2) describe a strategy for early cancer detection, CancerSEEK, aimed at screening for multiple different cancers within the general population. This study challenges current assumptions in the field of blood-based biomarkers and sets the stage for the next generation of cancer screening initiatives.

Given the potential curative advantage of earlier diagnosis and treatment, why have so many cancer screening approaches failed? In the past, efforts at screening healthy populations for cancer have relied on tests that were insufficiently specific. For example, most men with rising serum prostate-specific antigen (PSA) do not have prostate cancer but instead have benign prostatic enlargement. However, where accurate tests exist, there have been dramatic improvements in cancer outcomes (3). For example, advanced cervical cancer has virtually disappeared in countries where Pap screening is the standard of care; although less reliable, mammography and screening colonoscopy are recommended for early detection of breast and colon cancers in individuals above ages 40 to 45 and 50, respectively, and screening heavy smokers by use of low-dose chest computed tomography (CT) scans reduces deaths from lung cancer (4). However, these tests are imperfect, and cost-effectiveness for broad deployment remains a challenge, particularly because a multitude of false-positive test results may lead to extensive diagnostic evaluations and unnecessary medical interventions. Unfortunately, for the majority of cancers no effective early screening tests are available.

It is in this setting that emerging molecular analyses of blood specimens, so-called "liquid biopsies," are poised to revolutionize cancer screening (5). Circulating cell- free DNA (cfDNA) in the blood consists of small fragments of DNA that are approximately 150 nucleotides in length. cfDNA is primarily derived from normal tissues, but a small fraction may be derived from tumor cells in individuals who have cancer. This circulating tumor DNA (ctDNA) may be identified by the presence of characteristic mutations in cancer genes or by variations in chromosome copy numbers (6). Recent studies have established the reliability of ctDNA genotyping for monitoring treatment response and identifying drug

resistance mechanisms in patients with advanced cancer (7, 8). However, the much lower amount of ctDNA in the plasma of patients who have a localized tumor poses a challenge for early cancer screening, as does the absence of knowledge about which mutation to look for. Furthermore, some background mutations detectable in the blood may arise from nonmalignant proliferation of blood cells in older individuals, a phenomenon called clonal hematopoiesis of indeterminate potential (CHIP) (9). Importantly, cancer gene mutations alone are insufficient to identify the tissue of origin for a given cancer signal in the blood because similar mutations are present in multiple different cancers. Thus, a tissue-agnostic blood-based screening test has limited clinical utility, unless accompanied by insight into which organ should be investigated for follow-up.

Cohen *et al.* sought to combine ctDNA sequencing of cancer genes with quantitation of tumor-associated serum protein markers, deriving a probabilistic algorithm for the presence of cancer and for the tissue of origin. After estimating the minimal number of recurrent cancer gene mutations required for a robust signal in eight different cancer types, Cohen *et al.* assigned an Omega score to condense the entirety of the ctDNA sequencing data into a single number, based on the most predictive mutation identified. Added to the Omega score are levels for eight cancer-associated serum proteins, which are combined by the CancerSEEK algorithm into a single probability of the sample having come from an individual with cancer (see the figure). Of the 1005 patients studied with operable cancers, the test sensitivity ranged from 33 to 98%, depending on the cancer type, with a test specificity in healthy blood donors greater than 99%. In patients correctly identified as having some type of cancer, a further algorithm that incorporates the Omega score and the initial eight protein panel, as well as measurements of an additional 31 proteins and the patient's gender, correctly localizes the tumor to one of two top predicted anatomic sites in 83% of patients.

Among the key discoveries in this study is that a relatively small panel of cancer genes sequenced repeatedly to extreme depth to find rare alleles, with the pool of templates divided into multiple fractions in order to enhance signal detection, is sufficient to provide information for many different types of cancer. Compared with other efforts that use large-scale ctDNA sequencing (10), this approach will have greatly reduced cost. Also, by combining multiple protein biomarkers with ctDNA genotyping, the devised algorithm can implicate candidate tissues of origin, information unavailable from mutational data alone.

There are a number of important caveats. The predictive value of any diagnostic test relies on the prevalence of the disease within the tested population. For instance, in testing apparently healthy individuals within the general population, the prevalence of all eight cancers can be conservatively estimated as 1% of people over age 64 (11). Hence, in this setting even a test that is 99% sensitive and 99% specific will yield a positive predictive value (PPV) of only 50% (half of all test positives will be a false-positive result). Similarly, a positive CancerSEEK test result would be predicted to have a PPV of 40 to 45% for a person having any of the eight different cancers (2). Although the model was not designed to screen for individual cancer types, breaking down the aggregate PPV into its individual component cancers would result in further reduction in PPV, particularly for rare cancers. Because PPVs improve with higher disease prevalence, application of any cancer screening test to

subpopulations with increased genetic or environmental risk factors (for example, carriers of familial breast cancer susceptibility mutations, heavy smokers at risk for lung cancer, or patients with liver cirrhosis predisposed to hepatocellular carcinoma) would of course increase the likelihood of true-positive results.

A well-documented challenge in early cancer detection studies is that patient populations at increased risk for cancer may also have precancerous or inflammatory conditions resulting in baseline elevation of serum protein biomarkers, a confounding factor that is not well recapitulated in the healthy control population used to build the CancerSEEK test. Although the relative contributions of ctDNA genotyping versus serum protein markers varies among the individual cancers analyzed by Cohen *et al.,* the integration of these potentially orthogonal markers into the CancerSEEK algorithm is likely to strengthen its accuracy when trained on clinically relevant populations. Extending from this study, future research may combine multiple blood-based analytes, including massive ctDNA sequencing for mutations (10), high-throughput screening for chromosome copy number variation (12), scoring for tissue-specific DNA meth- ylation patterns (13), serum-based multi-protein mass spectrometric quantitation (14), and digital quantitation of lineage- specific RNA from circulating tumor cells (15). Each of these blood-based assays may provide optimal capabilities for the detection of specific cancer types within at-risk patient populations, and as elegantly demonstrated by Cohen *et al.*, combinations of tests may be optimal to enable both high- sensitivity detection and identification of the tissue of origin.

The ultimate goal of cancer screening is to diagnose invasive cancers early, while they are still curable. All the patients studied by Cohen *et al.* had been diagnosed as part of standard clinical evaluation and were candidates for surgical resection of their tumors, but many already had local invasion, and their cure rate is unknown. As the authors acknowledge, diagnosing cancers before clinical symptoms trigger an initial diagnostic procedure will require detection of even lower levels of signal, and prospective studies of patients whose cancer is first detected through blood-based screening will be required to determine real-world performance and whether such early screening can lead to improved cure rates. In addition, as the authors suggest, by coupling an initial blood-based screening test with secondary high-specificity confirmatory tests, it may be possible to achieve PPVs that would enable large-scale clinical implementation.

Undoubtedly, effective screening for early invasive cancers represents the best hope for reducing cancer mortality and morbidity. The conceptual advances and the practical feasibility of the CancerSEEK assay constitute an important milestone toward the application of early cancer detection. Most importantly, the ongoing development of cost-effective and accurate blood-based cancer screening strategies is poised to revolutionize clinical cancer care, bringing with it new emphasis on genetic and environmental risk stratification so as to tailor application of screening tests; minimally invasive imaging, biopsy, and molecular characterization of early tumors that are discovered and might be either indolent or invasive; and deployment of increasingly effective therapeutic options to stages of cancer for which they have curative potential. The vision of effective earlier cancer detection and intervention warrants validation in appropriate populations through large-scale clinical trials that are likely to radically change the way we diagnose and treat cancer.

REFERENCES

1. Haber DA, Gray NS, Baselga J, Cell 145, 19 (2011). [PubMed: 21458664]

2. Cohen JD, Science 359, 926 (2018). [PubMed: 29348365]

3. Smith RA et al., CA Cancer J. Clin 67, 100 (2017). [PubMed: 28170086]

4. National Lung Screening Trial Research Team et al., N. Engl. J. Med365, 395 (2011). [PubMed: 21714641]

5. Bardelli A, Pantel K, Cancer Cell 31, 172 (2017). [PubMed: 28196593]

6. Wan JCM et al., Nat. Rev. Cancer 17, 223 (2017). [PubMed: 28233803]

7. Dawson SJ et al., N. Engl. J. Med 368, 1199 (2013). [PubMed: 23484797]

8. Goyal L et al., Cancer Discov 7, 252 (2017). [PubMed: 28034880]

9. Jaiswal S et al., N. Engl. J. Med 371, 2488 (2014). [PubMed: 25426837]

10. Aravanis AM, Lee M, Klausner RD, Cell 168,571 (2017). [PubMed: 28187279]

11. Surveillance, Epidemiology, and End Results (SEER) Program; https://seer.cancer.gov.

12. Adalsteinsson VA et al., Nat Commun. 8, 1324 (2017). [PubMed: 29109393]

13. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J, Cell 164, 57 (2016). [PubMed: 26771485]

14. Geyer PE,Holdt LM, Teupser D, Mann M, Mol. Sys. Bio 13, 942 (2017)

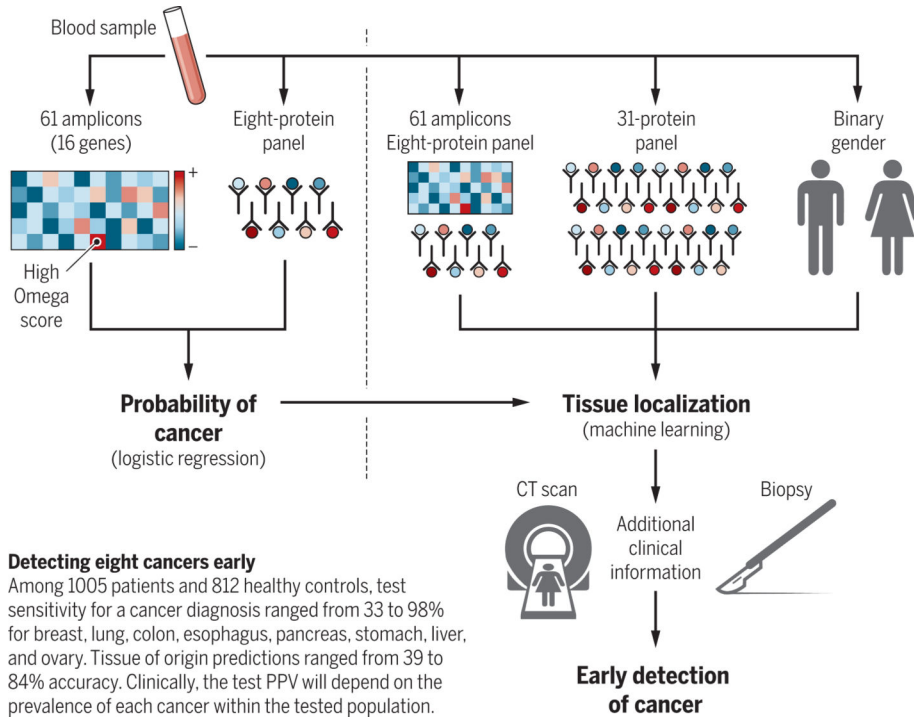15. Kalinich M et al., Proc. Natl. Acad. Sci. U.S.A 114, 1123 (2017). [PubMed: 28096363]

"...effective screening for early- invasive cancers represents the best hope for reducing cancer mortality and morbidity."

How the CancerSEEK algorithm works

16 cancer genes generates an Omega score that is combined with eight cancer-associated serum proteins to derive a probability for having any of eight different types of cancer. A machine learning algorithm then integrates these data with 31 additional serum proteins and patient gender to predict the tissue of origin.