



Published in final edited form as:

*Science*. 2018 February 23; 359(6378): 926–930. doi:10.1126/science.aar3247.

## Detection and localization of surgically resectable cancers with a multi-analyte blood test

Joshua D. Cohen<sup>1,2,3,4,5</sup>, Lu Li<sup>6</sup>, Yuxuan Wang<sup>1,2,3,4</sup>, Christopher Thoburn<sup>3</sup>, Bahman Afsari<sup>7</sup>, Ludmila Danilova<sup>7</sup>, Christopher Douville<sup>1,2,3,4</sup>, Ammar A. Javed<sup>8</sup>, Fay Wong<sup>1,3,4</sup>, Austin Mattox<sup>1,2,3,4</sup>, Ralph. H. Hruban<sup>3,4,9</sup>, Christopher L. Wolfgang<sup>8</sup>, Michael G. Goggins<sup>3,4,9,10,11</sup>, Marco Dal Molin<sup>4</sup>, Tian-Li Wang<sup>3,9</sup>, Richard Roden<sup>3,9</sup>, Alison P. Klein<sup>3,4,12</sup>, Janine Ptak<sup>1,2,3,4</sup>, Lisa Dobbyn<sup>1,3,4</sup>, Joy Schaefer<sup>1,3,4</sup>, Natalie Silliman<sup>1,2,3,4</sup>, Maria Popoli<sup>1,3,4</sup>, Joshua T. Vogelstein<sup>13</sup>, James D. Browne<sup>14</sup>, Robert E. Schoen<sup>15,16</sup>, Randall E. Brand<sup>15</sup>, Jeanne Tie<sup>17,18,19,20</sup>, Peter Gibbs<sup>17,18,19,20</sup>, Hui-Li Wong<sup>17</sup>, Aaron S. Mansfield<sup>21</sup>, Jin Jen<sup>22</sup>, Samir M. Hanash<sup>23</sup>, Massimo Falconi<sup>24</sup>, Peter J. Allen<sup>25</sup>, Shibin Zhou<sup>1,3,4</sup>, Chetan Bettegowda<sup>1,3,4</sup>, Luis A. Diaz Jr.<sup>1,3,4,\*</sup>, Cristian Tomasetti<sup>3,6,7,†</sup>, Kenneth W. Kinzler<sup>1,3,4,†</sup>, Bert Vogelstein<sup>1,2,3,4,†</sup>, Anne Marie Lennon<sup>3,4,8,10,11,†</sup>, and Nickolas Papadopoulos<sup>1,3,4,†</sup>

<sup>1</sup>Ludwig Center for Cancer Genetics and Therapeutics, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>2</sup>Howard Hughes Medical Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>3</sup>Sidney Kimmel Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>4</sup>Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>5</sup>Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA

<sup>6</sup>Department of Biostatistics, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205, USA

<sup>7</sup>Division of Biostatistics and Bioinformatics, Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA

<sup>8</sup>Department of Surgery, Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA

<sup>9</sup>Department of Pathology, Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA

<sup>†</sup>Corresponding author. ctomasetti@jhu.edu (C.T.); amlennon@jhmi.edu (A.M.L.); kinzlike@jhmi.edu (K.W.K.); bertvog@gmail.com (B.V.); npapado1@jhmi.edu (N.P.).

<sup>\*</sup>Present address: Division of Solid Tumor Oncology, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA.

### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/359/6378/926/suppl/DC1](http://www.sciencemag.org/content/359/6378/926/suppl/DC1)

Material and Methods

Figs. S1 to S4

Tables S1 to S11

References (31–37)

- <sup>10</sup>Department of Medicine, Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA
- <sup>11</sup>Department of Oncology, Johns Hopkins Medical Institutions, Baltimore, MD 21287, USA
- <sup>12</sup>Department of Epidemiology, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD 21205, USA
- <sup>13</sup>Institute for Computational Medicine, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
- <sup>14</sup>Department of Computer Science, Johns Hopkins University Whiting School of Engineering, Baltimore, MD 21218, USA
- <sup>15</sup>Department of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA
- <sup>16</sup>Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA 15260, USA
- <sup>17</sup>Division of Systems Biology and Personalized Medicine, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3021, Australia
- <sup>18</sup>Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne, VIC 3010, Australia
- <sup>19</sup>Department of Medical Oncology, Western Health, Melbourne, VIC 3021, Australia
- <sup>20</sup>Department of Medical Oncology, Peter MacCallum Cancer Center, Melbourne, VIC 3000, Australia
- <sup>21</sup>Division of Medical Oncology, Department of Oncology, Mayo Clinic, Rochester, MN 55902, USA
- <sup>22</sup>Division of Experimental Pathology, Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, MN 55902, USA
- <sup>23</sup>Sheikh Ahmed Center for Pancreatic Cancer Research, University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA
- <sup>24</sup>Division of Pancreatic Surgery, Department of Surgery, San Raffaele Scientific Institute Research Hospital, 20132 Milan, Italy
- <sup>25</sup>Department of Surgery, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA

## Abstract

Earlier detection is key to reducing cancer deaths. Here, we describe a blood test that can detect eight common cancer types through assessment of the levels of circulating proteins and mutations in cell-free DNA. We applied this test, called CancerSEEK, to 1005 patients with nonmetastatic, clinically detected cancers of the ovary, liver, stomach, pancreas, esophagus, colorectum, lung, or breast. CancerSEEK tests were positive in a median of 70% of the eight cancer types. The sensitivities ranged from 69 to 98% for the detection of five cancer types (ovary, liver, stomach, pancreas, and esophagus) for which there are no screening tests available for average-risk individuals. The specificity of CancerSEEK was greater than 99%: only 7 of 812 healthy controls scored positive. In addition, CancerSEEK localized the cancer to a small number of anatomic sites in a median of 83% of the patients.

The majority of localized cancers can be cured by surgery alone, without any systemic therapy (1). Once distant metastasis has occurred, however, surgical excision is rarely curative. One major goal in cancer research is therefore the detection of cancers before they metastasize to distant sites. For many adult cancers, it takes 20 to 30 years for incipient neoplastic lesions to progress to late-stage disease (2–4). Only in the past few years of this long process do neoplastic cells appear to successfully seed and give rise to metastatic lesions (2–5). Thus, there is a wide window of opportunity to detect cancers before the onset of metastasis. Even when metastasis has initiated but is not yet evident radiologically, cancers can be cured in up to 50% of cases with systemic therapies, such as cytotoxic drugs and immunotherapy (6–9). Once large, metastatic tumors are formed, however, current therapies are rarely effective (6–9).

The only widely used blood test for earlier cancer detection is based on measurement of prostate-specific antigen, and the proper use of this test is still being debated (10). The approved tests for cancer detection are not blood-based and include colonoscopy, mammography, and cervical cytology. New blood tests for cancer must have very high specificity; otherwise, too many healthy individuals will receive positive test results, leading to unnecessary follow-up procedures and anxiety. Blood tests that detect somatic mutations (“liquid biopsies”) offer the promise of exquisite specificity because they are based on driver gene mutations that are expected to be found only in abnormal clonal proliferations of cells, such as cancers (11–18). To date, the vast majority of cancer patients evaluated with mutation-based liquid biopsies have advanced-stage disease. In addition, no studies have examined a large number of healthy control individuals, which is essential for evaluation of the specificity of such tests (19). Diagnostic sensitivity is also an issue for liquid biopsies. Available evidence indicates that patients with early-stage cancers can harbor less than one mutant template molecule per milliliter of plasma (11, 20), which is often beyond the limit of detection of previously reported technologies that assess multiple mutations simultaneously (19, 21). Yet another issue with liquid biopsies is the identification of the underlying tissue of origin. Because the same gene mutations drive multiple tumor types, liquid biopsies based on genomic analysis alone generally cannot identify the anatomical location of the primary tumor.

We describe here a new blood test, called CancerSEEK, that addresses the issues described above. The test uses combined assays for genetic alterations and protein biomarkers and has the capacity not only to identify the presence of relatively early cancers but also to localize the organ of origin of these cancers.

Initial studies demonstrated that the maximum sensitivity of plasma DNA-based tests—liquid biopsies—was limited for localized cancers (11). A subsequent study suggested that the combination of four protein biomarkers with one genetic marker (*KRAS*) could enhance sensitivity for the detection of pancreatic cancers (20). We sought to generalize this approach by evaluating a panel of protein and gene markers that might be used to detect many solid tumors at a stage before the emergence of distant metastases. We began by designing a polymerase chain reaction (PCR)-based assay that could simultaneously assess multiple regions of driver genes that are commonly mutated in a variety of cancer types. In designing this test, we were confronted by four competing challenges. First, the test must

query a sufficient number of bases to allow detection of a large number of cancers. Second, each base queried in the test must be sequenced thousands of times to detect low-prevalence mutations (11, 19, 21, 22). Third, there must be a limit on the number of bases queried in the test because the more bases queried, the more likely that artifactual mutations would be identified, reducing the signal-to-noise ratio. And fourth, for implementation in a screening setting, the test must be cost effective and amenable to high throughput, factors that limit the amount of sequencing that can be performed. To overcome these challenges, we searched for the minimum number of short amplicons that would allow us to detect at least one driver gene mutation in each of the eight tumor types evaluated. Using publicly available sequencing data, we found that there was a fractional power law relationship between the number of amplicons required and the sensitivity of detection, with a plateau at ~60 amplicons (Fig. 1). Once this plateau was reached, raising the number of amplicons would not detect substantially more cancers but would increase the probability of false-positive results. This decreasing marginal utility defined the optimal number of amplicons.

On the basis of these data, we designed a 61-amplicon panel, with each amplicon querying an average of 33 base pairs (bp) within one of 16 genes (table S1). As shown in Fig. 1, this panel would theoretically detect 41% (liver) to 95% (pancreas) of the cancers in the Catalog of Somatic Mutations in Cancer (COSMIC) data set (23). In practice, the panel performed considerably better, detecting at least one mutation in 82%, two mutations in 47%, and more than two mutations in 8% of the 805 cancers evaluated in our study (Fig. 1, colored dots; fig. S1; and table S2). We were able to detect a larger fraction of tumors than predicted by the COSMIC data set because the PCR-based sequencing assay we used was more sensitive for detecting mutations than conventional genome-wide sequencing. On the basis of this analysis of the DNA from primary tumors, the predicted maximum detection capability of circulating tumor DNA (ctDNA) in our study varied by tumor type, ranging from 60% for liver cancers to 100% for ovarian cancers (Fig. 1).

Armed with this small but robust panel of amplicons, we developed two approaches that enabled the detection of the rare mutations expected to be present in plasma ctDNA. First, we used multiplex-PCR to directly and uniquely label each original template molecule with a DNA barcode. This design minimizes the errors inherent to massively parallel sequencing (24) and makes efficient use of the small amount of cell-free DNA present in plasma. Additionally, we divided the total amount of DNA recovered from plasma into multiple aliquots and performed independent assays on each replicate. In effect, this decreases the number of DNA molecules per well; however, it increases the fraction of each mutant molecule per well, making the mutants easier to detect. Because the sensitivity of detection is often limited by the fraction of mutant alleles in each replicate, this partitioning strategy allowed us to increase the signal-to-noise ratio and identify mutations present at lower prevalence than possible if all of the plasma DNA was evaluated at once.

The second component of CancerSEEK is based on protein biomarkers. Previous studies have demonstrated that a major fraction of early-stage tumors do not release detectable amounts of ctDNA, even when extremely sensitive techniques are used to identify them (11, 20). Many proteins potentially useful for early detection and diagnosis of cancer have been described in the literature (25–27). We searched this literature to find proteins that had

previously been shown to detect at least one of the eight cancer types described above with sensitivities >10% and specificities >99%. We identified 41 potential protein biomarkers (table S3) and evaluated them in preliminary studies on plasma samples from normal individuals as well as from cancer patients. We found that 39 of these proteins could be reproducibly evaluated through a single immunoassay platform, and we then used this platform to assay all plasma samples (table S3). Eight of the 39 proteins proved to be particularly useful for discriminating cancer patients from healthy controls (table S3).

We then used CancerSEEK to study 1005 patients who had been diagnosed with stage I to III cancers of the ovary, liver, stomach, pancreas, esophagus, colorectum, lung, or breast. No patient received neo-adjuvant chemotherapy before blood sample collection, and none had evident distant metastasis at the time of study entry. The median age at diagnosis was 64 (range 22 to 93). The eight cancer types were chosen because they are common in western populations and because no blood-based tests for their earlier detection are in common clinical use. The histopathological and clinical characteristics of the patients are summarized in table S4. The most common stage at presentation was American Joint Commission on Cancer (AJCC) stage II, accounting for 49% of patients, with the remaining patients harboring stage I (20%) or stage III (31%) disease. The number of samples per stage for each of the eight tumor types is summarized in table S11. The healthy control cohort consisted of 812 individuals of median age 55 (range 17 to 88) with no known history of cancer, high-grade dysplasia, auto-immune disease, or chronic kidney disease.

CancerSEEK evaluates levels of eight proteins and the presence of mutations in 1933 distinct genomic positions; each genomic position could be mutated in several ways (single base substitutions, insertions, or deletions). The presence of a mutation in an assayed gene or an elevation in the level of any of these proteins would classify a patient as positive. It was therefore imperative to use rigorous statistical methods to ensure the accuracy of the test. We used log ratios to evaluate mutations and incorporated them into a logistic regression algorithm that took into account both mutation data and protein biomarker levels to score CancerSEEK test results (supplementary materials). The mean sensitivities and specificities were determined by 10 iterations of 10-fold cross-validations. The receiver operating characteristic (ROC) curves for the entire cohort of cancer patients and controls in one representative iteration is shown in Fig. 2A.

The median sensitivity of CancerSEEK among the eight cancer types evaluated was 70% ( $P < 10^{-96}$  one-sided binomial test) and ranged from 98% in ovarian cancers to 33% in breast cancers (Fig. 2C). At this sensitivity, the specificity was >99%; only 7 of the 812 individuals without known cancers scored positive. We could not be certain that the few false positive-testing individuals identified among the healthy cohort did not actually have an as-yet undetected cancer, but classifying them as false positives provided the most conservative approach to classification and interpretation of the data.

The features of the test that were most important to the algorithm were the presence of a ctDNA mutation followed by elevations of cancer antigen 125 (CA-125), carcinoembryonic antigen (CEA), cancer antigen 19-9 (CA19-9), pro-lactin (PRL), hepatocyte growth factor (HGF), osteopontin (OPN), myeloperoxidase (MPO), and tissue inhibitor of

metalloproteinases 1 (TIMP-1) protein levels (table S9). Waterfall plots for each of the ctDNA and protein features used in CancerSEEK illustrate their distribution among individuals with and without cancer (fig. S2). The importance ranking of the ctDNA and protein features used in CancerSEEK are provided in table S9, and a principal component analysis displaying the clustering of individuals with and without cancer is shown in fig. S3. The complete data set, including the levels of all proteins studied and the mutations identified in the plasma samples, are provided in tables S5 and S6. The probabilistic rather than deterministic nature of the approach used here to call a sample positive is evident from fig. S4; each panel represents the sensitivity of CancerSEEK when one specific feature was excluded from the analysis.

One of the most important attributes of a screening test is the ability to detect cancers at relatively early stages. The median sensitivity of CancerSEEK was 73% for the most common stage evaluated (stage II), similar (78%) for stage III cancers, and lower (43%) for stage I cancers (Fig. 2B). The sensitivity for the earliest-stage cancers (stage I) was highest for liver cancer (100%) and lowest for esophageal cancer (20%).

The basis of liquid biopsy is that mutant DNA templates in plasma are derived from dying cancer cells and thus serve as exquisitely specific markers for neoplasia. To investigate whether CancerSEEK meets this expectation, we evaluated tumor tissue from 153 patients in whom ctDNA could be detected at statistically significant levels (supplementary materials) and for whom primary tumors were available. We found that the mutation in the plasma was identical to a mutation found in the primary tumor of the same individual in 138 (90%) of these 153 cases (table S7). This concordance between plasma and primary tumor was evident in all eight cancer types, and ranged from 100% in ovarian and pancreatic cancers to 82% in stomach cancers.

One limitation of liquid biopsies is their inability to determine the cancer type in patients who test positive, which poses challenges for clinical follow-up. To examine whether the CancerSEEK test can help identify a cancer's tissue of origin, we used supervised machine learning to predict the underlying cancer type in patients with positive CancerSEEK tests. The input algorithm took into account the ctDNA and protein biomarker levels as well as the gender of the patient (supplementary materials). One of the main purposes of such predictions is to determine the most appropriate follow-up test for cancer diagnosis or monitoring after a positive CancerSEEK test. We therefore grouped together patients with esophageal and gastric cancers because endoscopy would be the optimal follow-up in both instances. We then used this algorithm (supplementary materials) to study the 626 cancer patients with positive CancerSEEK tests. Without any clinical information about the patients, we were able to localize the source of the cancer to two anatomic sites in a median of 83% of these patients ( $P < 10^{-77}$  one-sided binomial test) (Fig. 3 and table S8). Furthermore, we were able to localize the source of the positive test to a single organ in a median of 63% of these patients ( $P < 10^{-47}$  one-sided binomial test) (Fig. 3 and table S8). Given that driver gene mutations are usually not tissue-specific, the vast majority of the localization information was derived from protein markers. The accuracy of prediction varied with tumor type; it was highest for colorectal cancers and lowest for lung cancers (Fig. 3 and table S10).

We have designed a multi-analyte blood test that can detect the presence of eight common solid tumor types. The advantage of combining completely different agents, with distinct mechanisms of action, is widely recognized in therapeutics (28–30) but has not been routinely applied to diagnostics. We combined protein biomarkers with genetic biomarkers to increase sensitivity without substantially decreasing specificity. Other cancer biomarkers—such as metabolites, mRNA transcripts, miRNAs, or methylated DNA sequences—could be similarly combined to increase sensitivity and localization of cancer site. Such multi-analyte tests are not meant to replace other non-blood-based screening tests, such as those for breast or colorectal cancers, but to provide additional information that could help identify those patients most likely to harbor a malignancy.

Several limitations of our study should be acknowledged. First, the patient cohort in our study was composed of individuals with known cancers, most diagnosed on the basis of symptoms of disease. Although none of our patients had clinically evident metastatic disease at the time of study entry, most individuals in a true screening setting would have less advanced disease, and the sensitivity of detection is likely to be less than reported here. Second, our controls were limited to healthy individuals, whereas in a true cancer screening setting, some individuals might have inflammatory or other diseases, which could result in a greater proportion of false-positive results than observed in our study. Third, although multiple-fold cross-validation is a powerful and widely used technique for demonstrating robust sensitivity and specificity on a cohort of this study's scale, we were not able to use a completely independent set of cases for testing, which would have been optimal. Last, the proportion of cancers of each type in our cohort was purposefully not representative of those in the United States as a whole because we wanted to evaluate at least 50 examples of each cancer type with the resources available to us. When weighted for actual incidence in the United States, we estimate the sensitivity of CancerSEEK to be 55% among all eight cancer types. This weighting would not affect the high sensitivities of CancerSEEK (69 to 98%) to detect five cancer types (ovary, liver, stomach, pancreas, and esophagus) for which there are no screening tests available for average-risk individuals.

Our study lays the conceptual and practical foundation for a single, multi-analyte blood test for cancers of many types. We estimate the cost of the test to be less than \$500, which is comparable or lower than other screening tests for single cancers, such as colonoscopy. The eight cancer types studied here account for 360,000 (60%) of the estimated cancer deaths in the United States in 2017, and their earlier detection could conceivably reduce deaths from these diseases. To actually establish the clinical utility of CancerSEEK and to demonstrate that it can save lives, prospective studies of all incident cancer types in a large population will be required.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank our patients for their courage and generosity. We are grateful to C. Blair and K. Judge for expert technical and administrative assistance. We thank H. Ren, J. Olson, M. Hathcock, H. Zeh, A. Singhi, S. Crippa, M. Ryan, and

L. Ryan for their assistance with this study. This work was supported by the Lustgarten Foundation for Pancreatic Cancer Research; The Virginia and D. K. Ludwig Fund for Cancer Research; The Conrad N. Hilton Foundation; The Commonwealth Fund; The John Templeton Foundation; the Clinomics Program; Mayo Clinic Center for Individualized Medicine; the Mayo Clinic Biobank; The Sol Goldman Center for Pancreatic Cancer Research; The Michael Rolfe Pancreatic Cancer Research Foundation; The Benjamin Baker Scholarship; The Gray Foundation; S. Wojcicki and D. Troper; The Marcus Foundation; The Honorable Tina Brozman Foundation; The Burroughs Wellcome Career Award for Medical Scientists; The Doris Duke Charitable Foundation (2014107); and National Institutes of Health grants P50-CA62924, P50-CA102701, CA06973, CA152753, GM-07309, U01CA200469, and U01CA152753. N.P., S.Z., K.W.K., L.D., and B.V. are founders of Personal Genome Diagnostics and PapGene. B.V. and K.W.K. are on the Scientific Advisory Board of Sysmex-Inostics. B.V. is also on the Scientific Advisory Boards of Exelixis GP. R.H.H. is on the Board of Directors of MiDiagnostics. These companies and others have licensed technologies from Johns Hopkins, and N.P., K.W.K., L.D., B.V. and R.H.H. receive equity or royalties from these licenses. The terms of these arrangements are being managed by Johns Hopkins University in accordance with its conflict of interest policies. L.D. is on the Board of Directors of Jounce Therapeutics and is a Scientific Advisor for Genocea, Cell Design Labs, and Merck. A.S.M. is a consultant for Abbvie, Genentech, Bristol-Myers Squibb, and Trovogene. B.V., N.P., and K.W.K. are inventors on a patent (U.S. 20140227705 A1) held by Johns Hopkins University that covers basic aspects of the SafeSeqS technology used in this paper. B.V., K.W.K., N.P., J.D.C., C.T., and A.M.L. are inventors on a patent application to be submitted by Johns Hopkins University that covers other aspects of SafeSeqS as well as the multi-analyte approach described in this paper. All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplementary materials. Contact C.T. for questions about the algorithms; A.M.L. for questions about clinically related issues; K.W.K. about the sequencing analyses; B.V. about experimental procedures; and N.P. about the overall design of the study.

## REFERENCES AND NOTES

1. Siegel RL, Miller KD, Jemal A. *CA Cancer J Clin.* 2017; 67:7–30. [PubMed: 28055103]
2. Vogelstein B, et al. *Science.* 2013; 339:1546–1558. [PubMed: 23539594]
3. Jones S, et al. *Proc Natl Acad Sci USA.* 2008; 105:4283–4288. [PubMed: 18337506]
4. Yachida S, et al. *Clin Cancer Res.* 2012; 18:6339–6347. [PubMed: 22991414]
5. Vogelstein B, Kinzler KW. *N Engl J Med.* 2015; 373:1895–1898. [PubMed: 26559569]
6. Bozic I, et al. *eLife.* 2013; 2:e00747. [PubMed: 23805382]
7. Semrad TJ, Fahrni AR, Gong IY, Khatri VP. *Ann Surg Oncol.* 2015; 22(suppl 3):S855–S862. [PubMed: 26100816]
8. Moertel CG, et al. *Ann Intern Med.* 1995; 122:321–326. [PubMed: 7847642]
9. Huang AC, et al. *Nature.* 2017; 545:60–65. [PubMed: 28397821]
10. Pinsky PF, Prorok PC, Kramer BS. *N Engl J Med.* 2017; 376:1285–1289. [PubMed: 28355509]
11. Bettegowda C, et al. *Sci Transl Med.* 2014; 6:224ra24.
12. Haber DA, Velculescu VE. *Cancer Discov.* 2014; 4:650–661. [PubMed: 24801577]
13. Dawson SJ, et al. *N Engl J Med.* 2013; 368:1199–1209. [PubMed: 23484797]
14. Wang Y, et al. *Sci Transl Med.* 2015; 7:293ra104.
15. Forshew T, et al. *Sci Transl.* 2012; 4:136ra68. *Med.*
16. Abbosh C, et al. *Nature.* 2017; 545:446–451. [PubMed: 28445469]
17. Beddowes E, Sammut SJ, Gao M, Caldas C. *Breast.* 2017; 34(suppl 1):S31–S35. [PubMed: 28694015]
18. Phallen J, et al. *Sci Transl Med.* 2017; 9:eaan2415. [PubMed: 28814544]
19. Cree IA, et al. *BMC Cancer.* 2017; 17:697. [PubMed: 29061138]
20. Cohen JD, et al. *Proc Natl Acad Sci USA.* 2017; 114:10202–10207. [PubMed: 28874546]
21. Bardelli A, Pantel K. *Cancer Cell.* 2017; 31:172–179. [PubMed: 28196593]
22. Diehl F, et al. *Proc Natl Acad Sci USA.* 2005; 102:16368–16373. [PubMed: 16258065]
23. Forbes SA, et al. *Nucleic Acids Res.* 2017; 45(D1):D777–D783. [PubMed: 27899578]
24. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. *Proc Natl Acad Sci USA.* 2011; 108:9530–9535. [PubMed: 21586637]
25. Liotta LA, Petricoin EF 3rd. *Clin Adv Hematol Oncol.* 2003; 1:460–462. [PubMed: 16258432]
26. Wang H, et al. *Expert Rev Proteomics.* 2016; 13:99–114. [PubMed: 26581546]
27. Patz EF Jr, et al. *J Clin Oncol.* 2007; 25:5578–5583. [PubMed: 18065730]



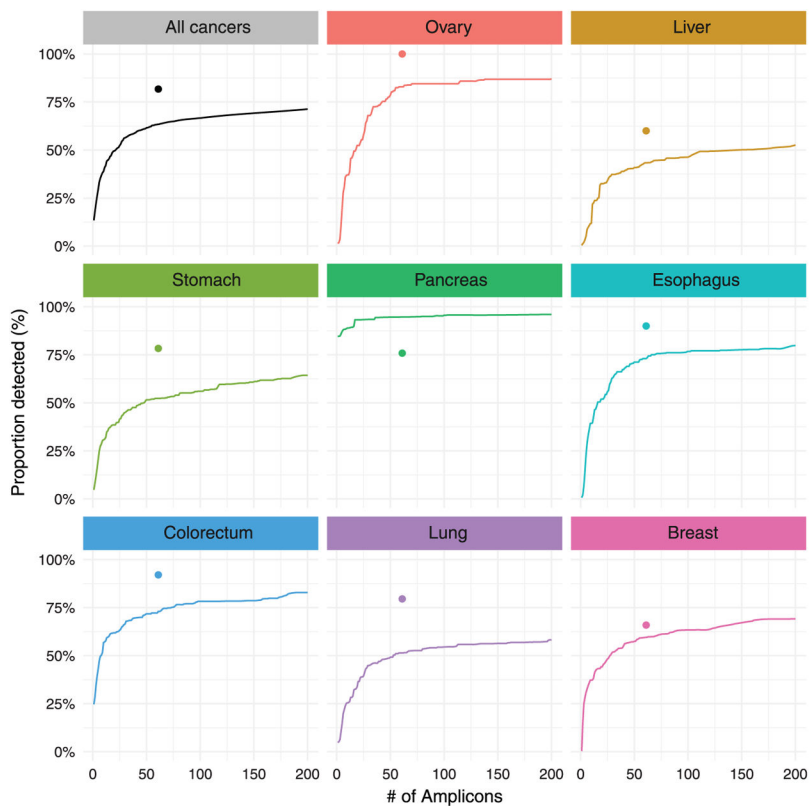
28. World Health Organization. Treatment of Tuberculosis: Guidelines. World Health Organization; 2010.
29. World Health Organization. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach. World Health Organization; 2016.
30. Benson AB 3rd, et al. J Natl Compr Canc Netw. 2017; 15:370–398. [PubMed: 28275037]

Author Manuscript

Author Manuscript

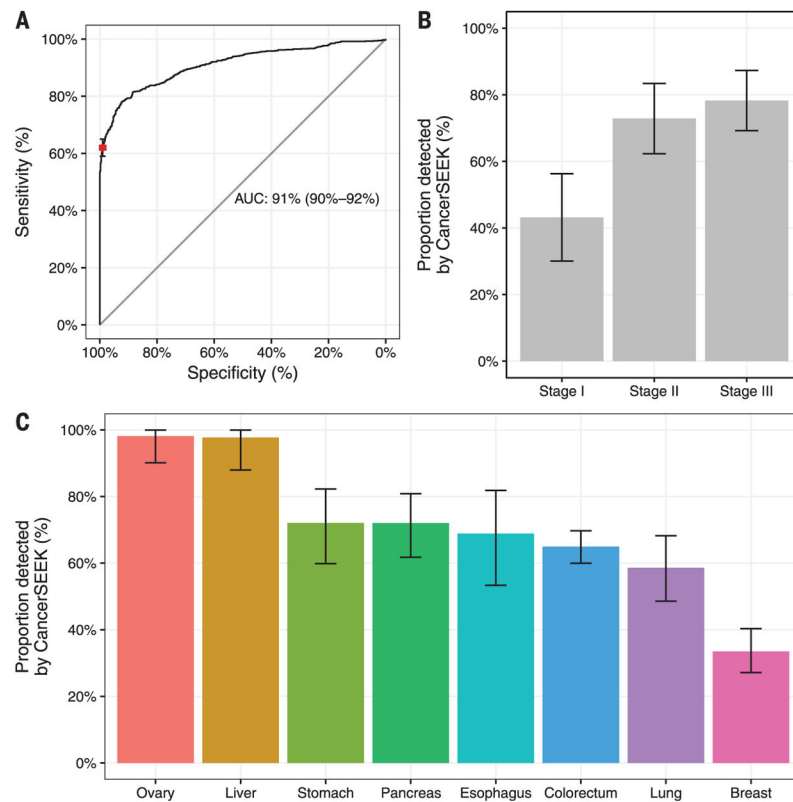
Author Manuscript

Author Manuscript



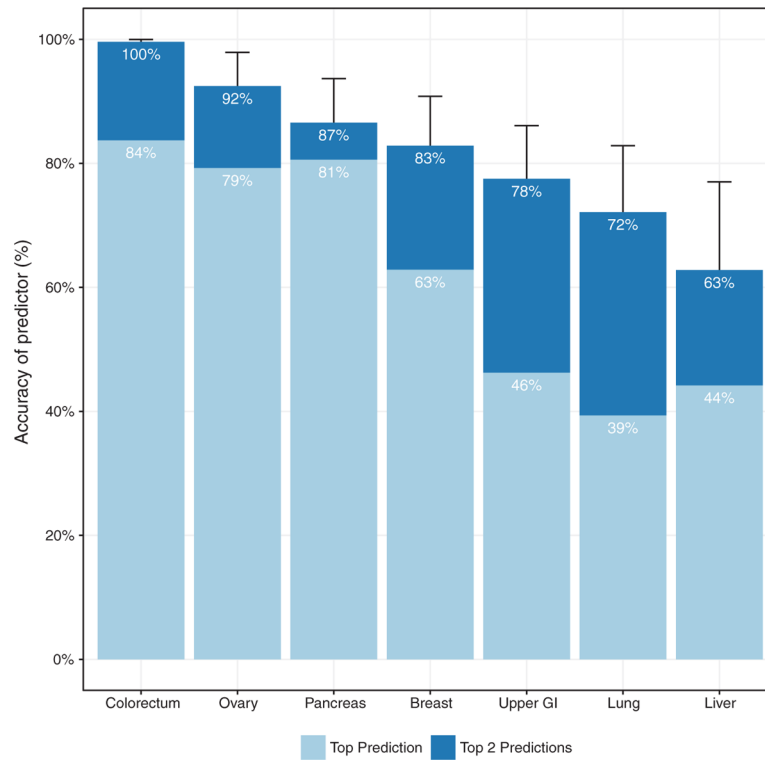
**Fig. 1. Development of a PCR-based assay to identify tumor-specific mutations in plasma samples**

Colored curves indicate the proportion of cancers of the eight types evaluated in this study that can be detected with an increasing number of short (<40 bp) amplicons. The sensitivity of detection increases with the number of amplicons but plateaus at ~60 amplicons. Colored dots indicate the fraction of cancers detected by using the 61-amplicon panel used in 805 cancers evaluated in our study, which averaged 82%. Publicly available sequencing data were obtained from the COSMIC repository.



**Fig. 2. Performance of CancerSEEK**

(A) ROC curve for CancerSEEK. The red point on the curve indicates the test's average performance (62%) at >99% specificity. Error bars represent 95% confidence intervals for sensitivity and specificity at this particular point. The median performance among the eight cancer types assessed was 70%. (B) Sensitivity of CancerSEEK by stage. Bars represent the median sensitivity of the eight cancer types, and error bars represent standard errors of the median. (C) Sensitivity of CancerSEEK by tumor type. Error bars represent 95% confidence intervals.



**Fig. 3. Identification of cancer type by supervised machine learning for patients classified as positive by CancerSEEK**

Percentages correspond to the proportion of patients correctly classified by one of the two most likely types (sum of light and dark blue bars) or the most likely type (light blue bar).

Predictions for all patients for all cancer types are provided in table S8. Error bars represent 95% confidence intervals.