www.sciencemag.org/cgi/content/full/science.aar3247/DC1

Supplementary Material for

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

Joshua D. Cohen, Lu Li, Yuxuan Wang, Christopher Thoburn, Bahman Afsari, Ludmila Danilova, Christopher Douville, Ammar A. Javed, Fay Wong, Austin Mattox, Ralph. H. Hruban, Christopher L. Wolfgang, Michael G. Goggins, Marco Dal Molin, Tian-Li Wang, Richard Roden, Alison P. Klein, Janine Ptak, Lisa Dobbyn, Joy Schaefer, Natalie Silliman, Maria Popoli, Joshua T. Vogelstein, James D. Browne, Robert E. Schoen, Randall E. Brand, Jeanne Tie, Peter Gibbs, Hui-Li Wong, Aaron S. Mansfield, Jin Jen, Samir M. Hanash, Massimo Falconi, Peter J. Allen, Shibin Zhou, Chetan Bettegowda, Luis A. Diaz Jr., Cristian Tomasetti,* Kenneth W. Kinzler,* Bert Vogelstein,* Anne Marie Lennon,* Nickolas Papadopoulos*

*Corresponding author. Email: ctomasetti@jhu.edu (C.T.); amlennon@jhmi.edu (A.M.L.); kinzlke@jhmi.edu (K.W.K); bertvog@gmail.com (B.V.); npapado1@jhmi.edu

(N.P.) Published 18 January 2018 as *Science* First Release

DOI: 10.1126/science.aar3247

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:
(available at www.sciencemag.org/content/science.aar3247/DC1)

Tables S1 to S11 as a separate Excel file

Materials and Methods

Plasma, white blood cell and tumor DNA samples

The study was approved by the Institutional Review Boards for Human Research at each institution, and complied with Health Insurance Portability and Accountability Act. Informed consent was obtained from all patients. Patients with cancers of the ovary, liver, esophagus, pancreas, stomach, colorectum, lung or breast who were thought to have Stage I to III disease prior to surgery were eligible for inclusion in the study. Peripheral blood was collected after informed consent was obtained and prior to the patients undergoing surgical resection. Patients who were later recognized to have received neo-adjuvant therapy, patients who were found to have stage IV cancer at the time of surgical resection, and patients in whom blood was documented to be collected while anesthesia was administered (*20*), were excluded from the study. General demographics, surgical pathology, and AJCC stage (7th edition) were documented. The 'healthy' cohort consisted of peripheral blood samples obtained from 812 individuals of median age 55 (IQR interquartile range 28 to 65) with no history of cancer. The cancer and healthy control samples were processed in an identical manner. Plasma samples from 46 of the 1,005 cancer patients and 181 of the 812 normal samples had been previously evaluated with a different approach (*20*) (table S4).

DNA was purified from an average of 7.5 mL plasma (table S4) using a QIASymphony circulating DNA kit (cat # 1091063), as specified by the manufacturer. DNA from peripheral WBCs was also purified with the QIAsymphony DP DNA Midi Kit (Cat # 937255) as specified by the manufacturer. Tumor tissues were formalin-fixed and paraffin-embedded (FFPE) according to standard histopathologic procedures and also purified with a QIAsymphony DP DNA Midi Kit (Cat # 937255).

Mutation detection and analysis

For amplification of DNA from plasma, 61 primer pairs were designed to amplify 66 to 80 bp segments containing regions of interest from 16 genes (table S1). The 61 primer pairs were divided into two non-overlapping sets each containing either 28 or 33 primer pairs. Each of these two primer sets were used to amplify DNA in six independent 25 μl reactions as previously described (*31*) except that 15 cycles were used for the initial amplification. We implemented this partitioning approach for two reasons. First, it reduces the complexity of the template facilitating the detection of rare alleles, and second provides duplicate signals for mutations that are detectable in multiple wells. The choice of 6 wells was dictated by technical limitations rather than by optimal design. Experimentally, we found that the optimal elution volume after DNA purification from 7.5 mL of plasma was 75 uL, and this volume allows the use of six wells per amplicon set (one set contains 28 amplicons, the other contains 33 amplicons). Thus, a total of 12 wells (each containing 5 uL of template DNA) can be evaluated with our approach, which represents one row of a standard 8 x 12 well PCR plate, easily handled robotically. The PCR products were purified with AMPure XP beads (Beckman Coulter, PA, USA) and 1% of the purified PCR products were then amplified in a second round of PCR as described in (*31*), but using 21 cycles. PCR products from the second round of amplification were then purified with AMPure and sequenced on an Illumina MiSeq or

HiSeq 4000 instrument. FASTQ files can be obtained from the European Genome-phenome Archive (accession number EGAS00001002764).

The template-specific portion of the reads was matched to reference sequences using custom scripts written in Python, SQL, and C# (Python version of analysis pipeline available for download at https://github.com/InSilicoSolutions/SafeSeqS, In Silico Solutions, Falls Church, VA). Reads from a common template molecule were then grouped based on the unique identifier sequences (UIDs) that were incorporated as molecular barcodes (*24*). Artefactual mutations introduced during the sample preparation or sequencing steps were reduced by requiring a mutation to be present in > 90% of reads in each UID family. Redundant reads arising from optical duplication were eliminated by requiring reads with the same UID and sample index to be at least 5,000 pixels apart when located on the same tile. Mutations that met one of the two following criteria were considered (i) present in the COSMIC database (*23*), or (ii) predicted to be inactivating in tumor suppressor genes (nonsense mutations, out-of-frame insertions or deletions, canonical splice site mutations). Synonymous mutations, except those at exon ends (*32*), and intronic mutations, except for those at splice sites, were excluded. The mutant allele frequency within a positive well was defined as the proportion of UIDs in the positive well that are mutant. Thus, the MAFs reflect the mutant fraction within each well and represents an independent sampling of the mutant allele frequency in the sample of interest. The MAF of a mutation in a sample (rather than the well) was defined as the total number of supermutants present in all six wells divided by the total number of UIDs in all six wells.

Evaluation of plasma proteins

The Bioplex 200 platform (Biorad, Hercules CA) was used to determine the concentration of multiple target proteins in the plasma samples. Luminex bead based immunoassays (Millipore, Bilerica NY) were performed following the manufacturers protocols and concentrations were determined using 5 parameter log curve fits (using Bioplex Manager 6.0) with vendor provided standards and quality controls. The HCCBP1MAG-58K panel was used to detect FGF2, Osteopontin, sFas, IL-8/CXCL8, Prolactin, HE4, HGF, AFP, CA125, IL6, CA15-3, TGFa, CYFRA21-1, CEA, CA19-9 and Leptin. The HANG2MAG-12K panel was used to detect PAR, sPECAM-1, TSP-2, sEGFR, AXL and sHER2/sEGFR2/sErbB2. The HCMBMAG-22K panel was used to detect DKK1, GDF15, Osteoprotegerin (OPG) and Neuron-specific enolase (NSE). The HCCBP4MAG-58K panels was used to detect Kallikrein-6, CD44, Midkine and Mesothelin. The HAGP1MAG-12K panel was used to detect Follistatin, G-CSF, Angiopoietin-2 and Endoglin. The HCCBP3MAG-58K panel was used to detect SHBG, Galectin and Myeloperoxidase. The HTMP1MAG-54K panel was used to detect TIMP-1 and TIMP-2. LRG-1 and Vitronectin were not included in this study since they could not be reproducibly evaluated with a single immunoassay platform.

CancerSEEK algorithm for sample classification and tissue localization

The classification of a sample's ctDNA status was obtained from a statistical test comparing the normalized mutation frequencies of the sample of interest to the

distributions of the normalized mutation frequencies of, respectively, normal and cancer samples in the training set. A step-by-step description of the algorithm is as follows:

1) MAF normalization. All mutations that did not have >1 supermutant in at least one well were excluded from the analysis. The mutant allele frequency (MAF), defined as the ratio between the total number of supermutants in each well from that sample and the total number of UIDs in the same well from that sample, was first normalized based on the observed MAFs for each mutation in a set of normal controls comprising the normal plasmas in the training set plus a set of 256 WBCs from unrelated healthy individuals. All MAFs with <100 UIDs were set to zero. This normalization was performed by first calculating the average MAF (ave_i) for each mutation i=1, … n, found among the normal controls. Using the 25th percentile of the distribution generated by these averages as the reference value (ave_ref), each MAF was normalized multiplying it by the ratio ave_ref / ave_i. For example, if the observed average MAF of a mutation in a set of controls was 10 times higher than ave_ref, then each MAF for that mutation was multiplied by 1/10. If a mutation in a test sample was not observed in any normal control, it was not normalized. Standard normalization, i.e. subtracting the mean and dividing by the standard deviation, did not perform as well in cross-validation.

2) Reference distributions and p-values. Following this mutation-specific normalization, the UID range was split in 10 intervals (<1,000, 1,000 - 2,000, ... , 8,000 - 9,000, > 9,000). Depending on the number of UIDs, the MAF of each mutation in each well was compared to two reference distributions of MAFs built from samples in the corresponding UID range: 1) a distribution built from all the normal control plasmas in the training set plus a set of 256 WBCs from unrelated, healthy individuals; and 2) a distribution built from the plasma samples from cancer patients in the training set. The cancer training set included only those in which the same mutation was present in the plasma and in the corresponding primary tumor, with an MAF > 5% in the tumor. Corresponding p-values, p^N and p^C, were thus obtained. The reference distributions for both the normal and cancer samples were built independently, from the training sets, in each round and each iteration of 10-fold cross-validation, i.e., 90% of the samples in each iteration were used for training and 10% of the samples were used for testing.

3) Log ratios and omega scores. For each mutation, the log ratio of these two p-values, p^C / p^N was then calculated, and the minimum and maximum of these log ratios across the six wells were eliminated so that the results would be less sensitive to outliers. We considered the log ratio of the p-values rather than the standard log-likelihood ratio because the relatively low number of data points available did not allow a robust estimation of the densities of the MAF distributions (particularly for p^C). An "omega" score was then determined according to the following formula:

$$\Omega = \sum_{i=1}^{4} w_i * \ ln \frac{p_i^C}{p_i^N} \ ,$$

where w_i is the number of UIDs in well *i* divided by the total number of UIDs for that mutation in the four wells that were included in the analysis (the two outlying wells

were excluded, as noted above). We weighted the log ratio of p-values so that those wells containing more template molecules would have a greater impact on the final statistic (the omega score). The rationale for this weighting was that the larger the number of template molecules in a well, the more confidence in the result.

To further illustrate how the omega score is obtained, a specific example of its calculation is provided here. Consider the KRAS p.G12S, c.34G>A mutation found in sample INDI 256 PLS 1. The number of supermutants and UIDs in each of the six wells were (161, 3755), (78, 2198), (99, 2966), (84, 2013), (177, 3694), (117,3427), respectively. These pairs yield the six MAFs (0.043, 0.035, 0.033, 0.042, 0.048, 0.034, or (0.0057, 0.0047, 0.0044, 0.0056, 0.0064, 0.0045) after normalization. These normalized MAFs correspond to the six p-values (1.06E-06, 5.70E-06, 1.02E-05, 1.03E-06, 3.09E-07, 8.83E-06) when compared to the reference MAF distribution among controls in the training set, and to the six p-values (0.100, 0.124, 0.128, 0.114, 0.094, 0.112) when compared to the reference MAF distribution among cancers in the training set. The ratio of those two vectors yields the vector p^C / p^N = (94243, 21716, 12510, 110752, 305090, 12680). By eliminating the minimum and maximum values of those ratios, and applying the above formula for omega, we obtain the omega score for that mutation:

$$\Omega = \frac{3755}{11393} ln(94243) + \frac{2198}{11393} ln(21716) + \frac{2013}{11393} ln(110752) + \frac{3427}{11393} ln(12680) = 10.60.$$

When a mutation identified in a plasma sample had $\Omega > 1$, and was not identified in the primary tumor of the patient, we evaluated DNA from white blood cells (WBCs) of the same patient whenever WBCs were available (23% of the cancer patients). WBC DNA was tested with the same 61-amplicon panel to ensure that the plasma mutation was not a result of Clonal Hematopoiesis of Indeterminate Potential (*33*). WBCs from the normal individuals were evaluated identically whenever a mutation with $\Omega > 1$ was found in the plasma. Any mutation that was identified in the WBCs as well as in the plasma was excluded from the analysis. The requirement for exclusion was that the ratio between the max MAF in the plasma and the max MAF in the WBC was less than 100. The mutation with the greatest Ω score in each patient or normal control was then deemed the "top mutation" and is listed in table S5.

4) Protein's normalization and transformation. To account for the variations in the lower and upper limits of detection across different experiments, we set all values smaller than *m*, defined as the maximum among all lower limits of detection for a given protein among all experiments, equal to *m*. By symmetry, we set all values larger than *M*, defined as the minimum among all upper limits of detection for that protein across all experiments, equal to *M*. To be conservative, a further transformation was applied to the proteins levels. Specifically, if a protein's concentration in the sample of interest was lower than the 95th percentile of the concentration found for that same protein among the normal samples in the training set, then the protein's concentration was set equal to zero; otherwise its original concentration value was used. For the Ω score, the same threshold

transformation was used but with a constant threshold equal to 0, because $\Omega > 0$ indicates an MAF that is more likely to originate from a cancer than from normal tissue.

5) Logistic Regression. The omega score was used as a feature in logistic regression (LR). The other 8 features used in LR were the concentrations of the following 8 proteins, selected from the original 39 proteins via a straightforward optimization: CA-125, CA19-9, CEA, HGF, MPO, OPN, PRL, TIMP-1. The optimization first eliminated any protein that, according to a Mann-Whitney-Wilcoxon test, had higher median values in normal than in cancer samples, eliminating 13 proteins and leaving 26 proteins to be evaluated. This was followed by a forward selection based on the importance of each feature, as evaluated by the decrease in accuracy of the same logistic regression when that protein alone was dropped from the remaining 26 protein features. The *R glmnet* package (version 2.10-13) was then used to perform the Logistic Regression, with the lambda parameter set to zero (*34*). Ten rounds of 10-fold cross-validations were performed. The classification calls obtained in an average round of 10-fold cross-validation (CV) are listed for each of the 812 normal individuals and the 1,005 cancer patients in table S4.

5) *Tissue localization*. For the prediction of the cancer type, we used the same 9 features (mutation omega score and levels of eight proteins) plus patient gender and the other 31 proteins evaluated in this study (table S3). Cancer type prediction was performed only on the cancer samples that were correctly classified as cancer by LR. Random Forest, as implemented in the *randomForest* package (version 4.6-12) (*35*) was used for this prediction. Ten rounds of 10-fold CV were performed and, for consistency, in each round and in each fold the same partition used by LR was used by Random Forest. The classification calls obtained in an average round of 10-fold CV (the same round for which cancer status is reported in table S5), are listed in table S8.

For determining the concordance between mutations identified in the plasma with those identified in primary tumors (table S7), we only considered the 153 cases in which a mutation could be identified with high confidence in the plasma (Ω score > 3, table S5) and in which the primary tumor contained *any* mutation that was present at a mutant allele fraction of > 5% (table S2). This approach allowed us to avoid scoring tumors that had low neoplastic contents (*36*).

Sample identification

To confirm that plasma, WBC, and primary tumor DNA samples originated from the same patient, we utilized primers that could be used to amplify ~38,000 unique long interspersed nucleotide elements (LINEs) from throughout the genome (*37*). These ~38,000 LINEs contain 26,220 common polymorphisms which can establish or refute sample identity among plasma, white blood cell and tumor samples. We identified the genotype at each polymorphic location and calculated the percent concordance between the samples of interest. Concordance was defined as the number of matched polymorphic sites that were identical in both samples divided by the total number of genotypes that had adequate coverage in both samples. Two samples were considered a match if concordance was > 0.99 and at least 5,000 amplicons had adequate coverage.

Statistical analysis

Continuous variables were reported as means and standard deviations or medians and range, while categorical variables were reported as whole numbers and percentages. Confidence intervals (CI) for sensitivities were calculated using a binomial distribution. Principle component analysis was performed using the R *stats* package (version 3.4.0). One-sided p-values for the obtained accuracy were calculated assuming a binomial with success probability equal to the no-information rate (NIR), taken to be the largest class percentage in the data, using the R *stats* package (version 3.4.0). The one-sided probability was chosen because the goal was not to test whether the obtained accuracy was significantly higher or lower than the NIR. Rather, we wanted to report the probability that an accuracy as high as the one obtained, or even higher, could be obtained when assuming a binomial with a success probability equal to the NIR.

Tables S1 to S11:

Table S1 (Microsoft Excel Format): Primer sequences for multiplex PCR assays.

Table S2 (Microsoft Excel Format): Mutations identified in primary tumors.

Table S3 (Microsoft Excel Format): Protein biomarkers analyzed and included in CancerSEEK test.

Table S4 (Microsoft Excel Format): Histopathological and clinical characteristics of the cancer patients and healthy controls.

Table S5 (Microsoft Excel Format): Mutations identified in plasma samples from cancer patients and healthy controls.

Table S6 (Microsoft Excel Format): Concentrations of assayed protein biomarker in plasma samples from cancer patients and healthy controls.

Table S7 (Microsoft Excel Format): Concordance between mutations identified in the plasma with those identified in primary tumors.

Table S8 (Microsoft Excel Format): Cancer type localization results for the 617 cancer patients identified by CancerSEEK.

Table S9 (Microsoft Excel format): Logistic regression model coefficients and importance scores.

Table S10 (Microsoft Excel format): Confusion matrix of top predictions from cancer type localization results.

Table S11 (Microsoft Excel format): Cancer patients evaluated in this study by tumor type and stage.

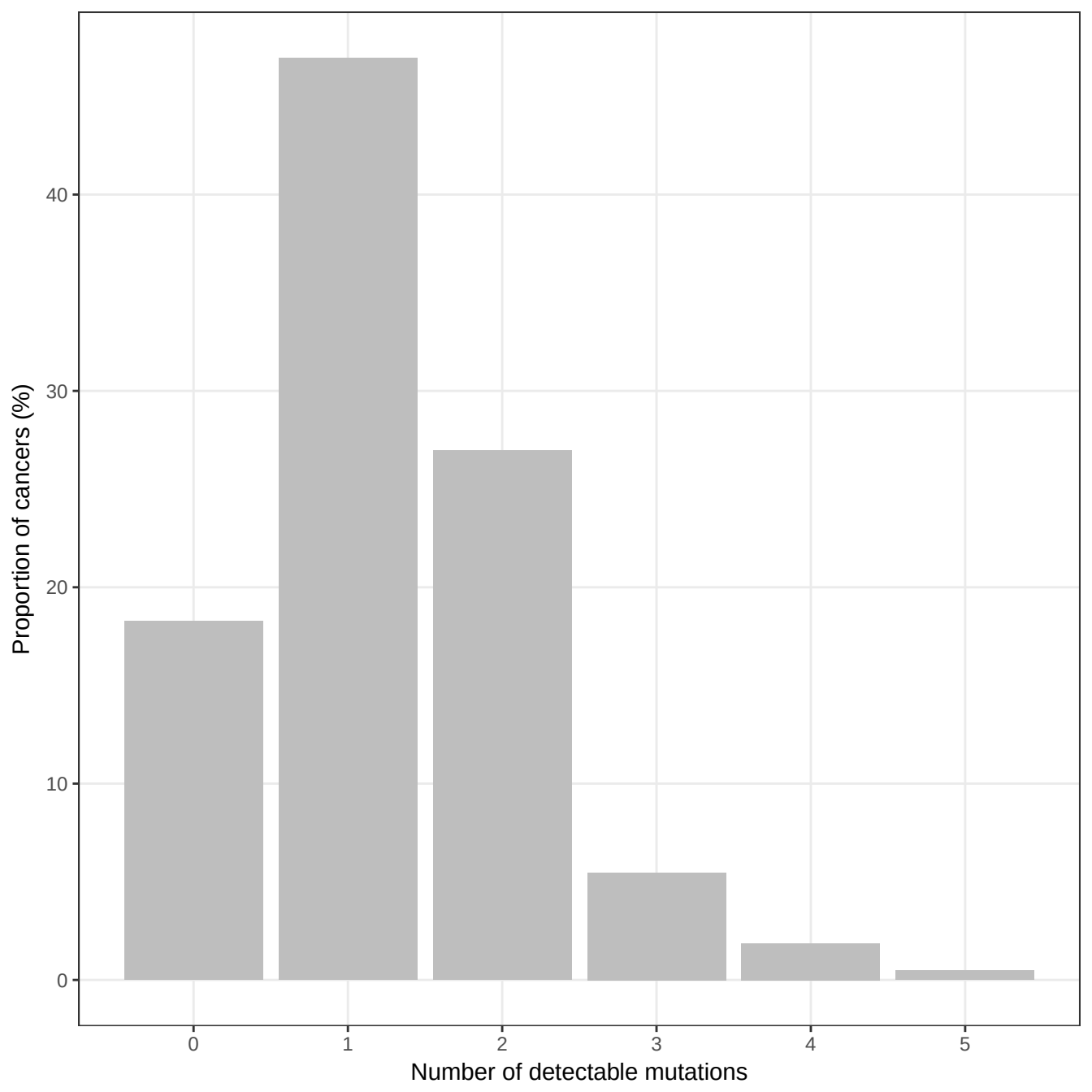
Proportion of cancers (%)
40
30
20
10
0
0
1
2
3
4
5
Number of detectable mutations

Fig. S1. Distribution of the number of detectable mutations within the 805 primary tumors evaluated.

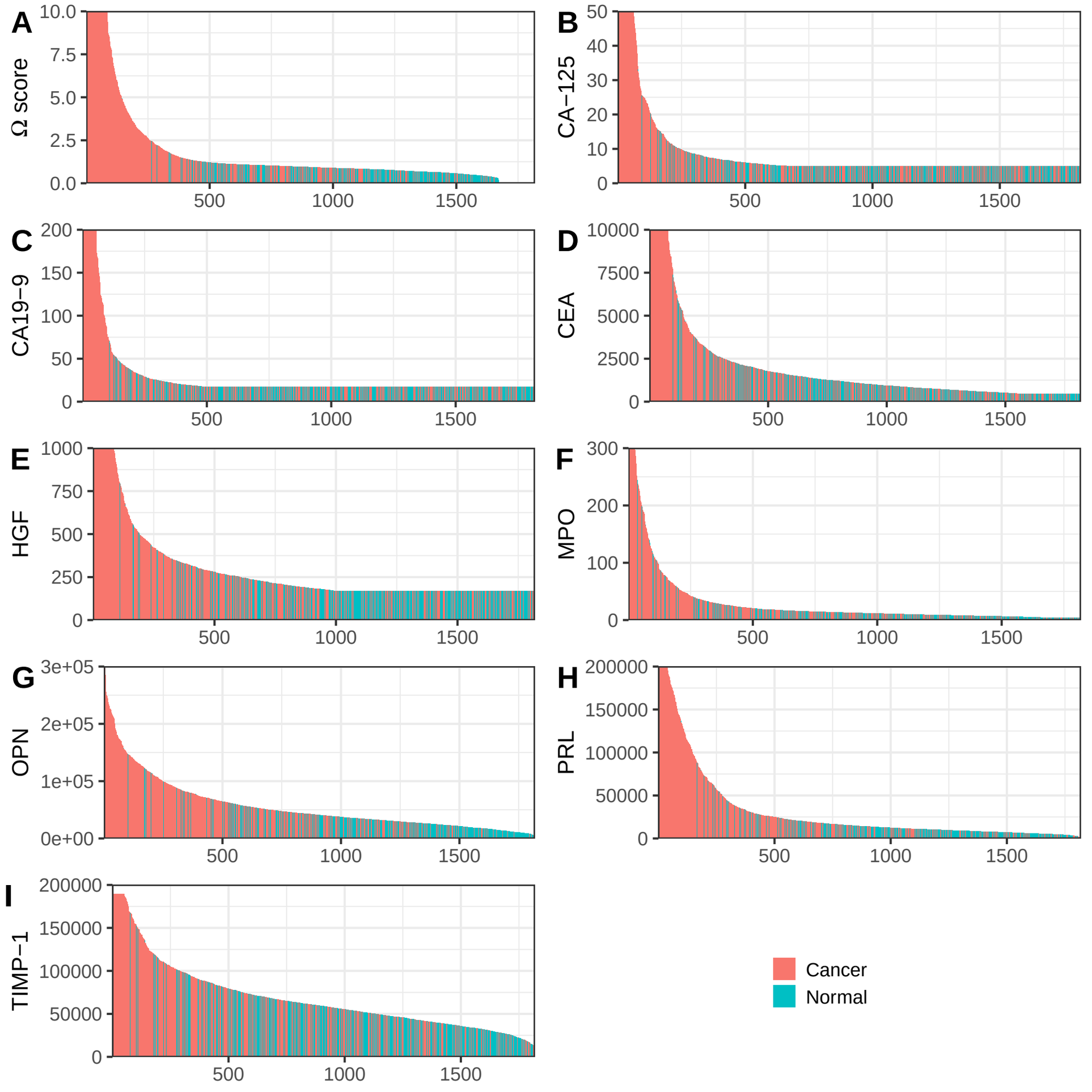
A
Ω score
10.0
7.5
5.0
2.5
0.0
500
1000
1500
B
CA−125
50
40
30
20
10
0
C
CA19−9
200
150
100
50
0
D
CEA
10000
7500
5000
2500
0
E
HGF
1000
750
500
250
F
MPO
300
200
100
G
OPN
3e+05
2e+05
1e+05
0e+00
H
PRL
200000
150000
100000
50000
I
TIMP−1
Cancer
Normal

Fig. S2. Waterfall plots of the ctDNA and eight protein features used in CancerSEEK illustrate the separation between healthy controls and cancer patients. Values are sorted from high (left) to low (right). Each column represents an individual patient sample (red, cancer patient; blue, healthy control).

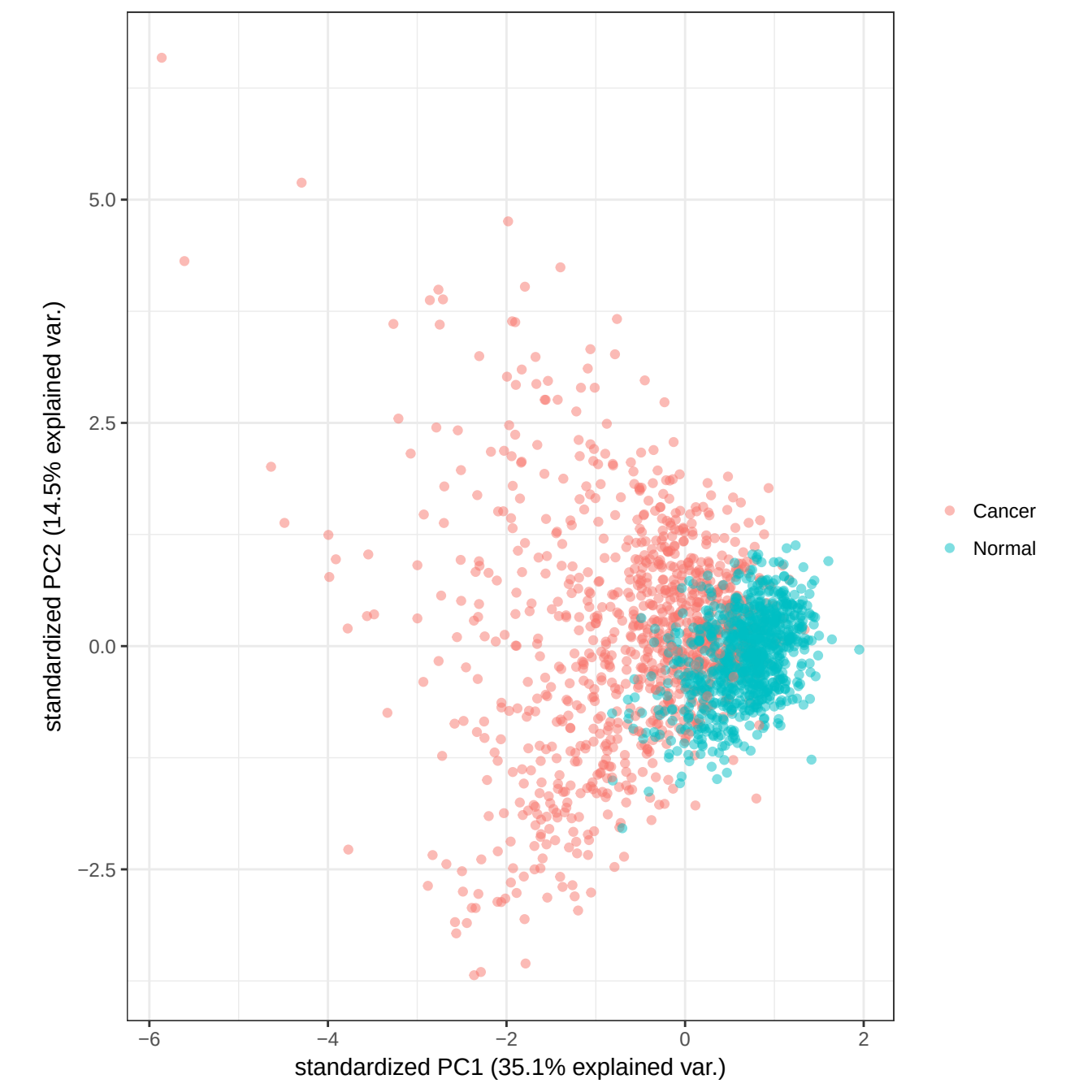
standardized PC2 (14.5% explained var.)
5.0
2.5
0.0
−2.5
−6
−4
−2
0
2
standardized PC1 (35.1% explained var.)
Cancer
Normal

Fig. S3. Principal component analysis of the ctDNA and eight protein features used in CancerSEEK. Each dot represents an individual patient sample (red, cancer patient; blue, healthy control).

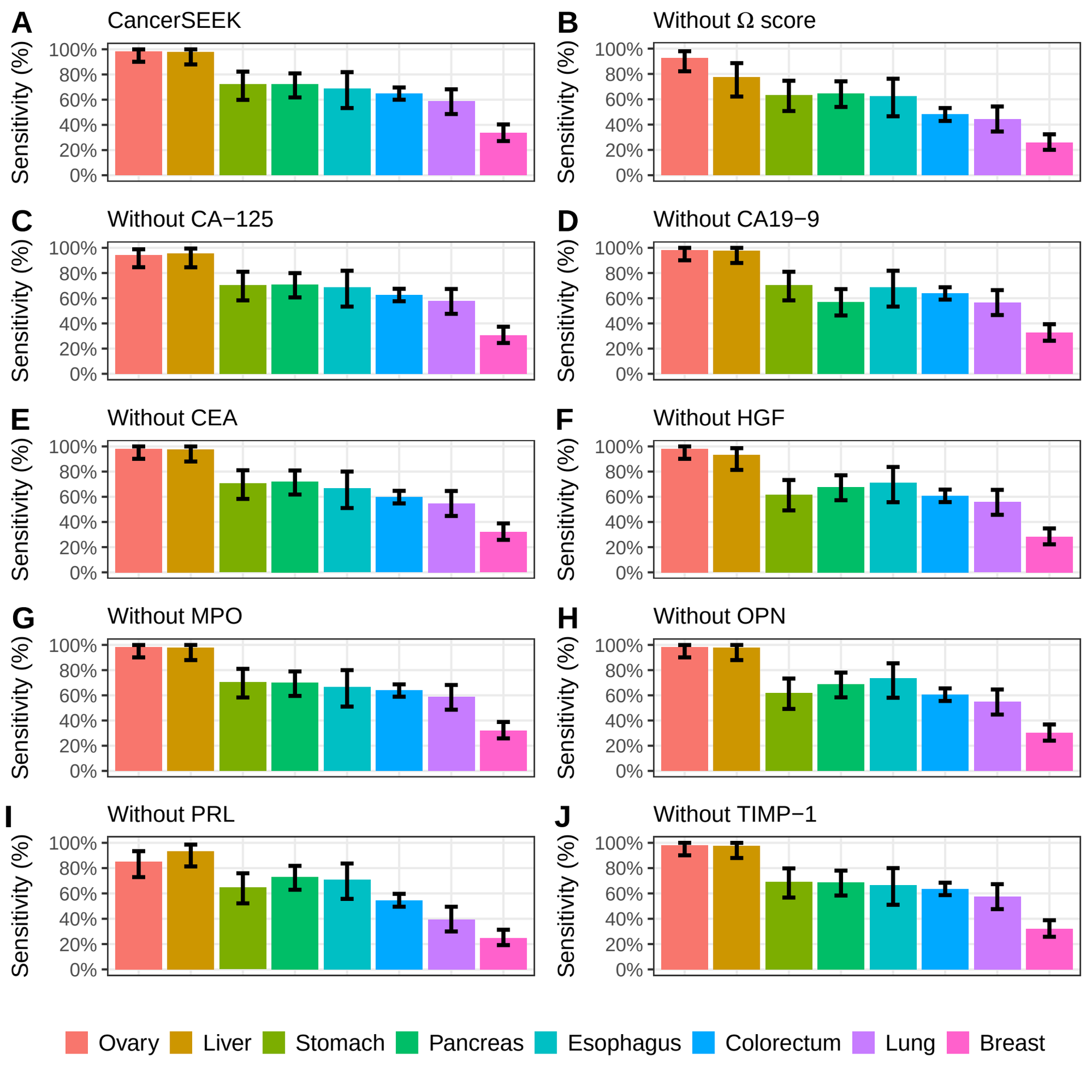
A
CancerSEEK
B
Without Ω score
C
Without CA−125
D
Without CA19−9
E
Without CEA
F
Without HGF
G
Without MPO
H
Without OPN
I
Without PRL
J
Without TIMP−1
Sensitivity (%)
100%
80%
60%
40%
20%
0%
Ovary
Liver
Stomach
Pancreas
Esophagus
Colorectum
Lung
Breast

Fig. S4. Effect of individual CancerSEEK features on sensitivity. (A) Sensitivity of CancerSEEK by tumor type as in Fig. 2C. (B-J) Each panel displays the sensitivity achieved when a particular CancerSEEK feature is excluded from the logistic regression. The difference in sensitivity relative to that achieved by CancerSEEK reflects the relative contribution of each biomarker to the performance of the CancerSEEK test.

References and Notes

1. R. L. Siegel, K. D. Miller, A. Jemal, Cancer Statistics, 2017. *CA Cancer J. Clin.* **67**, 7–30 (2017). doi:10.3322/caac.21387 Medline

2. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz Jr., K. W. Kinzler, Cancer genome landscapes. *Science* **339**, 1546–1558 (2013). doi:10.1126/science.1235122 Medline

3. S. Jones, W. D. Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, A. Traulsen, M. A. Nowak, C. Siegel, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. Willis, S. D. Markowitz, Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283–4288 (2008). doi:10.1073/pnas.0712345105 Medline

4. S. Yachida, C. M. White, Y. Naito, Y. Zhong, J. A. Brosnan, A. M. Macgregor-Das, R. A. Morgan, T. Saunders, D. A. Laheru, J. M. Herman, R. H. Hruban, A. P. Klein, S. Jones, V. Velculescu, C. L. Wolfgang, C. A. Iacobuzio-Donahue, Clinical significance of the genetic landscape of pancreatic cancer and implications for identification of potential long-term survivors. *Clin. Cancer Res.* **18**, 6339–6347 (2012). doi:10.1158/1078-0432.CCR-12-1215 Medline

5. B. Vogelstein, K. W. Kinzler, The Path to Cancer —Three Strikes and You're Out. *N. Engl. J. Med.* **373**, 1895–1898 (2015). doi:10.1056/NEJMp1508811 Medline

6. I. Bozic, J. G. Reiter, B. Allen, T. Antal, K. Chatterjee, P. Shah, Y. S. Moon, A. Yaqubie, N. Kelly, D. T. Le, E. J. Lipson, P. B. Chapman, L. A. Diaz Jr., B. Vogelstein, M. A. Nowak, Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* **2**, e00747 (2013). doi:10.7554/eLife.00747 Medline

7. T. J. Semrad, A. R. Fahrni, I. Y. Gong, V. P. Khatri, Integrating Chemotherapy into the Management of Oligometastatic Colorectal Cancer: Evidence-Based Approach Using Clinical Trial Findings. *Ann. Surg. Oncol.* **22** (Suppl 3), S855–S862 (2015). doi:10.1245/s10434-015-4610-4 Medline

8. C. G. Moertel, T. R. Fleming, J. S. Macdonald, D. G. Haller, J. A. Laurie, C. M. Tangen, J. S. Ungerleider, W. A. Emerson, D. C. Tormey, J. H. Glick, M. H. Veeder, J. A. Mailliard, Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: A final report. *Ann. Intern. Med.* **122**, 321–326 (1995). doi:10.7326/0003-4819-122-5-199503010-00001 Medline

9. A. C. Huang, M. A. Postow, R. J. Orlowski, R. Mick, B. Bengsch, S. Manne, W. Xu, S. Harmon, J. R. Giles, B. Wenz, M. Adamow, D. Kuk, K. S. Panageas, C. Carrera, P. Wong, F. Quagliarello, B. Wubbenhorst, K. D'Andrea, K. E. Pauken, R. S. Herati, R. P. Staupe, J. M. Schenkel, S. McGettigan, S. Kothari, S. M. George, R. H. Vonderheide, R. K. Amaravadi, G. C. Karakousis, L. M. Schuchter, X. Xu, K. L. Nathanson, J. D. Wolchok, T. C. Gangadhar, E. J. Wherry, T-cell invigoration to tumour burden ratio associated with anti-PD-1 response. *Nature* **545**, 60–65 (2017). doi:10.1038/nature22079 Medline

10. P. F. Pinsky, P. C. Prorok, B. S. Kramer, Prostate Cancer Screening - A Perspective on the Current State of the Evidence. *N. Engl. J. Med.* **376**, 1285–1289 (2017). doi:10.1056/NEJMsb1616281 Medline

11. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014). doi:10.1126/scitranslmed.3007094 Medline

12. D. A. Haber, V. E. Velculescu, Blood-based analyses of cancer: Circulating tumor cells and circulating tumor DNA. *Cancer Discov.* **4**, 650–661 (2014). doi:10.1158/2159-8290.CD-13-1014 Medline

13. S. J. Dawson, D. W. Y. Tsui, M. Murtaza, H. Biggs, O. M. Rueda, S.-F. Chin, M. J. Dunning, D. Gale, T. Forshew, B. Mahler-Araujo, S. Rajan, S. Humphray, J. Becq, D. Halsall, M. Wallis, D. Bentley, C. Caldas, N. Rosenfeld, Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013). doi:10.1056/NEJMoa1213261 Medline

14. Y. Wang, S. Springer, C. L. Mulvey, N. Silliman, J. Schaefer, M. Sausen, N. James, E. M. Rettig, T. Guo, C. R. Pickering, J. A. Bishop, C. H. Chung, J. A. Califano, D. W. Eisele, C. Fakhry, C. G. Gourin, P. K. Ha, H. Kang, A. Kiess, W. M. Koch, J. N. Myers, H. Quon, J. D. Richmon, D. Sidransky, R. P. Tufano, W. H. Westra, C. Bettegowda, L. A. Diaz Jr., N. Papadopoulos, K. W. Kinzler, B. Vogelstein, N. Agrawal, Detection of somatic mutations and HPV in the saliva and plasma of patients with head and neck squamous cell carcinomas. *Sci. Transl. Med.* **7**, 293ra104 (2015). doi:10.1126/scitranslmed.aaa8507 Medline

15. T. Forshew, M. Murtaza, C. Parkinson, D. Gale, D. W. Y. Tsui, F. Kaper, S.-J. Dawson, A. M. Piskorz, M. Jimenez-Linan, D. Bentley, J. Hadfield, A. P. May, C. Caldas, J. D. Brenton, N. Rosenfeld, Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci. Transl. Med.* **4**, 136ra68 (2012). doi:10.1126/scitranslmed.3003726 Medline

16. C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. Marie Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers,

V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann, C. Swanton, M. Jamal-Hanjani, C. Abbosh, S. Veeriah, S. Shafi, J. Czyzewska-Khan, D. Johnson, J. Laycock, L. Bosshard-Carter, G. Goh, R. Rosenthal, P. Gorman, N. Murugaesu, R. E. Hynds, G. A. Wilson, N. J. Birkbak, T. B. K. Watkins, N. McGranahan, S. Horswell, M. A. Bakir, E. Grönroos, R. Mitter, M. Escudero, A. Stewart, P. Van Loo, A. Rowan, H. Xu, S. Turajlic, C. Hiley, J. Goldman, R. K. Stone, T. Denner, N. Matthews, G. Elgar, S. Ward, J. Biggs, M. Costa, S. Begum, B. Phillimore, T. Chambers, E. Nye, S. Graca, K. Joshi, A. Furness, A. Ben Aissa, Y. N. S. Wong, A. Georgiou, S. A. Quezada, K. S. Peggs, J. A. Hartley, H. L. Lowe, J. Herrero, D. Lawrence, M. Hayward, N. Panagiotopoulos, S. Kolvekar, M. Falzon, E. Borg, T. Marafioti, C. Simeon, G. Hector, A. Smith, M. Aranda, M. Novelli, D. Oukrif, A. U. Akarca, S. M. Janes, R. Thakrar, M. D. Forster, T. Ahmad, S. M. Lee, D. Papadatos-Pastos, D. Carnell, R. Mendes, J. George, N. Navani, A. Ahmed, M. Taylor, J. Choudhary, Y. Summers, R. Califano, P. Taylor, R. Shah, P. Krysiak, K. Rammohan, E. Fontaine, R. Booton, M. Evison, P. A. Crosbie, S. Moss, F. Idries, L. Joseph, P. Bishop, A. Chaturvedi, A. M. Quinn, H. Doran, A. Leek, P. Harrison, K. Moore, R. Waddington, J. Novasio, F. Blackhall, J. Rogan, E. Smith, C. Dive, J. Tugwood, G. Brady, D. G. Rothwell, F. Chemi, J. Pierce, S. Gulati, B. Naidu, G. Langman, S. Trotter, M. Bellamy, H. Bancroft, A. Kerr, S. Kadiri, J. Webb, G. Middleton, M. Djearaman, D. A. Fennell, J. A. Shaw, J. L. Quesne, D. A. Moore, A. Thomas, H. Walter, J. Riley, L. Martinson, A. Nakas, S. Rathinam, W. Monteiro, H. Marshall, L. Nelson, J. Bennett, L. Primrose, G. Anand, S. Khan, A. Amadi, M. Nicolson, K. Kerr, S. Palmer, H. Remmen, J. Miller, K. Buchan, M. Chetty, L. Gomersall, J. F. Lester, A. Edwards, F. Morgan, H. Adams, H. Davies, M. Kornaszewska, R. Attanoos, S. Lock, A. Verjee, M. MacKenzie, M. Wilcox, H. Bell, N. Iles, A. Hackshaw, Y. Ngai, S. Smith, N. Gower, C. Ottensmeier, S. Chee, B. Johnson, A. Alzetani, E. Shaw, E. Lim, P. De Sousa, M. T. Barbosa, A. Bowman, S. Jordan, A. Rice, H. Raubenheimer, C. Proli, M. E. Cufari, J. C. Ronquillo, A. Kwayie, H. Bhayani, M. Hamilton, Y. Bakar, N. Mensah, L. Ambrose, A. Devaraj, S. Buderi, J. Finch, L. Azcarate, H. Chavan, S. Green, H. Mashinga, A. G. Nicholson, K. Lau, M. Sheaff, P. Schmid, J. Conibear, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, P. Russell, T. Light, T. Horey, S. Danson, J. Bury, J. Edwards, J. Hill, S. Matthews, Y. Kitsanta, K. Suvarna, P. Fisher, A. D. Keerio, M. Shackcloth, J. Gosney, P. Postmus, S. Feeney, J. Asante-Siaw, T. Constantin, R. Salari, N. Sponer, A. Naik, B. G. Zimmermann, M. Rabinowitz, H. J. W. L. Aerts, S. Dentro, C. Dessimoz, C. Swanton; TRACERx consortium; PEACE consortium, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017). doi:10.1038/nature22364 Medline

17. E. Beddowes, S. J. Sammut, M. Gao, C. Caldas, Predicting treatment resistance and relapse through circulating DNA. *Breast* **34** (Suppl 1), S31–S35 (2017). doi:10.1016/j.breast.2017.06.024 Medline

18. J. Phallen, M. Sausen, V. Adleff, A. Leal, C. Hruban, J. White, V. Anagnostou, J. Fiksel, S. Cristiano, E. Papp, S. Speir, T. Reinert, M. W. Orntoft, B. D. Woodward, D. Murphy, S. Parpart-Li, D. Riley, M. Nesselbush, N. Sengamalay, A. Georgiadis, Q. K. Li, M. R.

Madsen, F. V. Mortensen, J. Huiskens, C. Punt, N. van Grieken, R. Fijneman, G. Meijer, H. Husain, R. B. Scharpf, L. A. Diaz Jr., S. Jones, S. Angiuoli, T. Ørntoft, H. J. Nielsen, C. L. Andersen, V. E. Velculescu, Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* **9**, eaan2415 (2017). doi:10.1126/scitranslmed.aan2415 Medline

19. I. A. Cree, L. Uttley, H. Buckley Woods, H. Kikuchi, A. Reiman, S. Harnan, B. L. Whiteman, S. T. Philips, M. Messenger, A. Cox, D. Teare, O. Sheils, J. Shaw; UK Early Cancer Detection Consortium, The evidence base for circulating tumour DNA blood-based biomarkers for the early detection of cancer: A systematic mapping review. *BMC Cancer* **17**, 697 (2017). doi:10.1186/s12885-017-3693-7 Medline

20. J. D. Cohen, A. A. Javed, C. Thoburn, F. Wong, J. Tie, P. Gibbs, C. M. Schmidt, M. T. Yip-Schneider, P. J. Allen, M. Schattner, R. E. Brand, A. D. Singhi, G. M. Petersen, S.-M. Hong, S. C. Kim, M. Falconi, C. Doglioni, M. J. Weiss, N. Ahuja, J. He, M. A. Makary, A. Maitra, S. M. Hanash, M. Dal Molin, Y. Wang, L. Li, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, M. G. Goggins, R. H. Hruban, C. L. Wolfgang, A. P. Klein, C. Tomasetti, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, A. M. Lennon, Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 10202–10207 (2017). doi:10.1073/pnas.1704961114 Medline

21. A. Bardelli, K. Pantel, Liquid Biopsies, What We Do Not Know (Yet). *Cancer Cell* **31**, 172–179 (2017). doi:10.1016/j.ccell.2017.01.002 Medline

22. F. Diehl, M. Li, D. Dressman, Y. He, D. Shen, S. Szabo, L. A. Diaz Jr., S. N. Goodman, K. A. David, H. Juhl, K. W. Kinzler, B. Vogelstein, Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 16368–16373 (2005). doi:10.1073/pnas.0507904102 Medline

23. S. A. Forbes, D. Beare, H. Boutselakis, S. Bamford, N. Bindal, J. Tate, C. G. Cole, S. Ward, E. Dawson, L. Ponting, R. Stefancsik, B. Harsha, C. Y. Kok, M. Jia, H. Jubb, Z. Sondka, S. Thompson, T. De, P. J. Campbell, COSMIC: Somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45** (D1), D777–D783 (2017). doi:10.1093/nar/gkw1121 Medline

24. I. Kinde, J. Wu, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, Detection and quantification of rare mutations with massively parallel sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9530–9535 (2011). doi:10.1073/pnas.1105422108 Medline

25. L. A. Liotta, E. F. Petricoin 3rd, The promise of proteomics. *Clin. Adv. Hematol. Oncol.* **1**, 460–462 (2003). Medline

26. H. Wang, T. Shi, W.-J. Qian, T. Liu, J. Kagan, S. Srivastava, R. D. Smith, K. D. Rodland, D. G. Camp 2nd, The clinical impact of recent advances in LC-MS for cancer biomarker discovery and verification. *Expert Rev. Proteomics* **13**, 99–114 (2016). doi:10.1586/14789450.2016.1122529 Medline

27. E. F. Patz Jr., M. J. Campa, E. B. Gottlin, I. Kusmartseva, X. R. Guan, J. E. Herndon 2nd, Panel of serum biomarkers for the diagnosis of lung cancer. *J. Clin. Oncol.* **25**, 5578–5583 (2007). doi:10.1200/JCO.2007.13.5392 Medline

28. *Treatment of Tuberculosis: Guidelines* (World Health Organization, Geneva, 2010).

29. *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. A. B. Benson 3rd, A. P. Venook, L. Cederquist, E. Chan, Y.-J. Chen, H. S. Cooper, D. Deming, P. F. Engstrom, P. C. Enzinger, A. Fichera, J. L. Grem, A. Grothey, H. S. Hochster, S. Hoffe, S. Hunt, A. Kamel, N. Kirilcuk, S. Krishnamurthi, W. A. Messersmith, M. F. Mulcahy, J. D. Murphy, S. Nurkin, L. Saltz, S. Sharma, D. Shibata, J. M. Skibber, C. T. Sofocleous, E. M. Stoffel, E. Stotsky-Himelfarb, C. G. Willett, C. S. Wu, K. M. Gregory, D. Freedman-Cass, Colon Cancer, Version 1.2017, NCCN Clinical Practice Guidelines in Oncology. *J. Natl. Compr. Canc. Netw.* **15**, 370–398 (2017). doi:10.6004/jnccn.2017.0036 Medline

31. Y. Wang, K. Sundfeldt, C. Mateoiu, IeM. Shih, R. J. Kurman, J. Schaefer, N. Silliman, I. Kinde, S. Springer, M. Foote, B. Kristjansdottir, N. James, K. W. Kinzler, N. Papadopoulos, L. A. Diaz, B. Vogelstein, Diagnostic potential of tumor DNA from ovarian cyst fluid. *eLife* **5**, e15175 (2016). doi:10.7554/eLife.15175 Medline

32. H. Jung, D. Lee, J. Lee, D. Park, Y. J. Kim, W.-Y. Park, D. Hong, P. J. Park, E. Lee, Intron retention is a widespread mechanism of tumor-suppressor inactivation. *Nat. Genet.* **47**, 1242–1248 (2015). doi:10.1038/ng.3414 Medline

33. S. Jaiswal, P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler, B. L. Ebert, Age-related clonal hematopoiesis associated with adverse outcomes. *N. Engl. J. Med.* **371**, 2488–2498 (2014). doi:10.1056/NEJMoa1408617 Medline

34. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010). doi:10.18637/jss.v033.i01 Medline

35. A. Liaw, M. Wiener, Classification and regression by randomForest. *R News* **2**, 18–22 (2001).

36. A. P. Makohon-Moore, M. Zhang, J. G. Reiter, I. Bozic, B. Allen, D. Kundu, K. Chatterjee, F. Wong, Y. Jiao, Z. A. Kohutek, J. Hong, M. Attiyeh, B. Javier, L. D. Wood, R. H. Hruban, M. A. Nowak, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, C. A. Iacobuzio-Donahue, Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer. *Nat. Genet.* **49**, 358–366 (2017). doi:10.1038/ng.3764 Medline

37. I. Kinde, N. Papadopoulos, K. W. Kinzler, B. Vogelstein, FAST-SeqS: A simple and efficient method for the detection of aneuploidy by massively parallel sequencing. *PLOS ONE* **7**, e41162 (2012). doi:10.1371/journal.pone.0041162 Medline