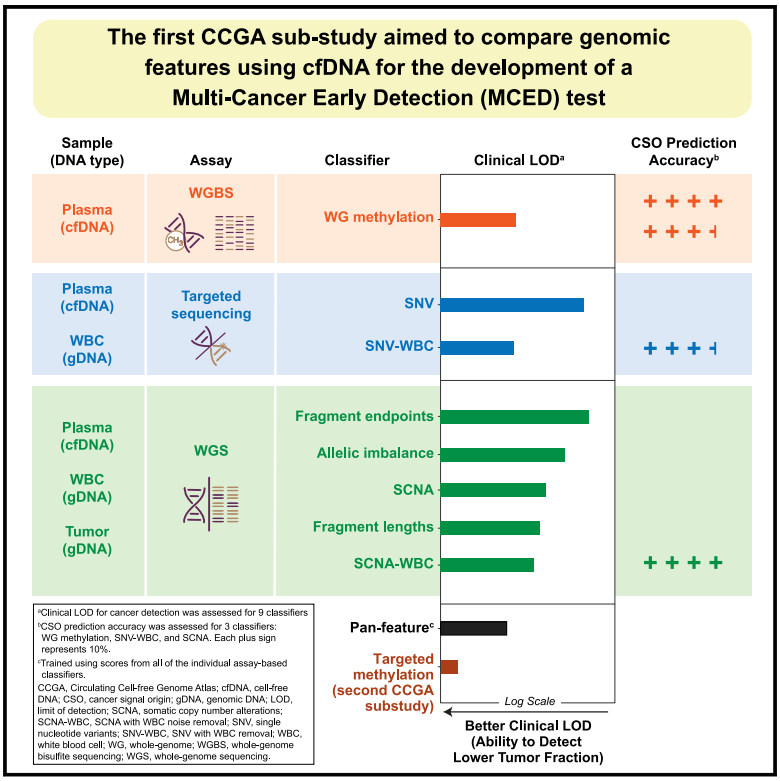
Cancer Cell

Article

Evaluation of cell-free DNA approaches for multi-cancer early detection

Graphical abstract



Highlights

- Clinical LOD is a useful benchmark to assess cfDNA-based test performance
- cTAF accounts for cfDNA cancer signal variation across cancer types and stages
- cfDNA methylation was the most promising genomic feature for cancer signal detection
- The results informed the development of a cfDNA-based multi-cancer early detection test

Authors

Arash Jamshidi, Minetta C. Liu, Eric A. Klein, ..., Alexander M. Aravanis, Michael V. Seiden, Charles Swanton

Correspondence

ajamshidipub@gmail.com (A.J.), ovenn@grailbio.com (O.V.)

In brief

Jamshidi et al. compare several approaches for circulating cell-free DNA (cfDNA)-based multi-cancer early detection (MCED) tests. A whole-genome methylation-based approach has the best performance among those evaluated. In addition, they define a metric—clinical limit of detection (LOD)—based on tumor fraction to enable future comparison of cfDNA-based tests.

Jamshidi et al., 2022, Cancer Cell *40*, 1537–1549
December 12, 2022
https://doi.org/10.1016/j.ccell.2022.10.022

CellPress

Cancer Cell

CellPress
OPEN ACCESS

Article

Evaluation of cell-free DNA approaches for multi-cancer early detection

Arash Jamshidi,[1,*] Minetta C. Liu,[2,18] Eric A. Klein,[3] Oliver Venn,[1,26,*] Earl Hubbell,[1] John F. Beausang,[1] Samuel Gross,[1] Collin Melton,[1] Alexander P. Fields,[1] Qinwen Liu,[1] Nan Zhang,[1] Eric T. Fung,[1] Kathryn N. Kurtzman,[1] Hamed Amini,[1] Craig Betts,[1,19] Daniel Civello,[1,20] Peter Freese,[1,21] Robert Calef,[1] Konstantin Davydov,[1,22] Saniya Fayzullina,[1] Chenlu Hou,[1] Roger Jiang,[1] Byoungsok Jung,[1] Susan Tang,[1,23] Vasiliki Demas,[1,24] Joshua Newman,[1] Onur Sakarya,[1] Eric Scott,[1] Archana Shenoy,[1] Seyedmehdi Shojaee,[1] Kristan K. Steffen,[1] Virgil Nicula,[1] Tom C. Chien,[1] Siddhartha Bagaria,[1] Nathan Hunkapiller,[1] Mohini Desai,[1] Zhao Dong,[1] Donald A. Richards,[4] Timothy J. Yeatman,[5,6] Allen L. Cohn,[7] David D. Thiel,[8] Donald A. Berry,[9] Mohan K. Tummala,[10] Kristi McIntyre,[11] Mikkael A. Sekeres,[12] Alan Bryce,[13] Alexander M. Aravanis,[14] Michael V. Seiden,[15,25] and Charles Swanton[16,17]

[1]GRAIL, LLC, Menlo Park, CA 94025, USA
[2]Department of Oncology, Mayo Clinic, Rochester, MN 55905, USA
[3]Cleveland Clinic, Cleveland, OH 44103, USA
[4]Texas Oncology, Tyler, TX 75702, USA
[5]Gibbs Cancer Center and Research Institute, Spartanburg, SC 29303, USA
[6]Department of Surgery, University of Utah, Salt Lake City, UT 84112, USA
[7]Rocky Mountain Cancer Center, Denver, CO 80218, USA
[8]Department of Urology, Mayo Clinic Florida, Jacksonville, FL 32224, USA
[9]Department of Biostatistics, MD Anderson Cancer Center, Houston, TX 77030, USA
[10]Mercy Clinic Cancer Center, Springfield, MO 65804, USA
[11]TOPA Dallas Presbyterian, Dallas, TX 75231, USA
[12]University of Miami Sylvester Cancer Center, Miami, FL 33136, USA
[13]Mayo Clinic, Phoenix, AZ 85054, USA
[14]Illumina, Inc., San Diego, CA 92122, USA
[15]US Oncology Research, The Woodlands, TX 77380, USA
[16]Francis Crick Institute, London, NW1 1AT, UK
[17]UCL Cancer Institute, CRUK Lung Cancer Centre of Excellence, London, WC1E 6DD, UK
[18]Present address: Natera, Inc., Austin, TX 78753, USA
[19]Present address: Cellanome, Inc., Palo Alto, CA 94303, USA
[20]Present address: Loyal / Cellular Longevity, Inc., Menlo Park, CA 94025, USA
[21]Present address: Meta, Menlo Park, CA 94025, USA
[22]Present address: Aurora Innovation, Inc., Mountain View, CA 94043, USA
[23]Present address: Strateos, Inc., Menlo Park, CA 94025, USA
[24]Present address: Tesseract Health, Guilford, CT 06437, USA
[25]Present address: GRAIL, LLC, Menlo Park, CA 94025, USA
[26]Lead contact
*Correspondence: ajamshidipub@gmail.com (A.J.), ovenn@grailbio.com (O.V.)
https://doi.org/10.1016/j.ccell.2022.10.022

SUMMARY

In the Circulating Cell-free Genome Atlas (NCT02889978) substudy 1, we evaluate several approaches for a circulating cell-free DNA (cfDNA)-based multi-cancer early detection (MCED) test by defining clinical limit of detection (LOD) based on circulating tumor allele fraction (cTAF), enabling performance comparisons. Among 10 machine-learning classifiers trained on the same samples and independently validated, when evaluated at 98% specificity, those using whole-genome (WG) methylation, single nucleotide variants with paired white blood cell background removal, and combined scores from classifiers evaluated in this study show the highest cancer signal detection sensitivities. Compared with clinical stage and tumor type, cTAF is a more significant predictor of classifier performance and may more closely reflect tumor biology. Clinical LODs mirror relative sensitivities for all approaches. The WG methylation feature best predicts cancer signal origin. WG methylation is the most promising technology for MCED and informs development of a targeted methylation MCED test.

INTRODUCTION

Circulating biomarkers have become important indicators of disease in clinical practice. In particular, the discovery that DNA from various tissues in the body exists in the bloodstream and other bodily fluids outside of cells (cell-free deoxyribonucleic acid [cfDNA])[1] has led to a number of blood-based cfDNA tests to interrogate specific genomic abnormalities, such as chromosomal copy number aberrations for non-invasive prenatal testing[2] or actionable tumor-derived mutations for targeted cancer therapy selection.[3] A critical application of blood-based cfDNA or other circulating analytes is the early detection of a shared cancer signal across multiple cancers using a blood test. One type of multi-cancer early detection (MCED) test is currently available,[4] and others are in development,[5,6] although each uses different information from cfDNA and, in some cases, other circulating analytes. To our knowledge, there has been no reported rigorous and systematic comparison of various genomic features from cfDNA for MCED testing until the analysis presented here.

Despite significant advances in cancer treatments and the known benefits of early detection through recommended screening tests, cancer was the first or second leading cause of death in over 100 countries in 2019.[7] In 2020, almost 10 million people died of cancer worldwide.[8] Many cancers are diagnosed at an advanced stage, after metastasis,[9] and so have lower survival compared with those that are diagnosed at earlier stages.[10] Moreover, advanced cancers may lead to considerable increases in health-care costs, which are projected to increase dramatically over the next 20 years.[11] Detecting cancers early is a viable population health strategy to improve outcomes, but the vast majority of cancers do not have recommended screening tests. Indeed, population screening is recommended by the US Preventive Services Task Force (USPSTF) only for breast, cervical, colorectal, lung (in high-risk patients), and—upon discussion with a physician—prostate cancer.[12–16] Although available, these single-cancer screening tests are not perfect[12–16] and are organ specific,[17] such that the false-positive rate of multiple tests is cumulative. For example, after 3 years of routine cancer screening in three different organs (lung, colon/rectum, and prostate or ovaries), the cumulative risk of one false positive is 60% for men and 49% for women.[18] This is not expected to be the case for an MCED test, given the low false-positive rate (<1%) that has been reported.[4,19] Recent modeling work predicted that adding an MCED test to standard care may improve early-stage detection and prevent 39% of all cancer-related deaths within 5 years of diagnosis that would otherwise be expected to occur in individuals with a positive test result.[20] Complementary MCED testing may allow population screening across numerous deadly cancer types at once and potentially a shift toward earlier detection of clinically significant tumors.[17]

Any population screening test for a low annual incidence disease such as cancer[9] requires high specificity to have a favorable benefit-to-harm balance and avoid burdening the population with false-positive results. Blood-based MCED tests face a challenge in this respect because there is an abundant background of non-cancer cfDNA in the blood relative to the genomic material shed from the tumor.[17,21,22] In addition, prevalent somatic biology such as clonal hematopoiesis (CH) is observed in cfDNA, which may confound specific cancer signal detection. Multiple approaches to overcome this signal-to-background ratio have been proposed,[23–26] including sequencing at high depth and breadth, as well as machine-learning techniques. As such, a fundamental evaluation of cfDNA approaches, each using contemporaneous blood samples from the same individuals with cancer who are demographically and geographically matched to non-cancer individuals in large studies may reduce statistical uncertainty and provide robust evidence for an MCED method best suited to population use. The Circulating Cell-free Genome Atlas (CCGA; NCT02889978) study included three prespecified substudies, in part, for this purpose. The first substudy evaluated cfDNA features in prototype assays and prototype machine-learning classifiers to determine the most promising approach for an MCED test with a low false-positive rate and sufficient sensitivity to improve outcomes. That first substudy is reported here.

RESULTS

Participant disposition and demographics

A total of 2,800 participants, 1,628 with cancer and 1,172 without cancer (non-cancer), were included in the first CCGA substudy (Figure 1) and randomly assigned to independent training or validation sets. Of the 2,800 participants, 2,261 (1,414 training; 847 validation) had analyzable results (Figure 1; analyzable results were those samples that met predetermined laboratory quality control standards). Patients with cancer across all clinical stages diagnosed by screening or clinical presentation were enrolled before starting cancer therapy. Demographic and clinical characteristics of participants with analyzable assay data were similar across cancer and non-cancer participants, as well as training and validation sets (Table 1). Importantly, among participants with cancer, mean age was matched across training and validation sets (61 ± 12 years for the training set, 62 ± 12 years for the validation set). Smoking status, clinical cancer stage, sex, and ethnicity were also well-balanced across sets, with slightly more representation from early-stage cancers than later-stage cancers across both sets (Table 1).

Cancer signal detection

Ten different classifiers—nine using different cfDNA features from three broad and deep prototype MCED assays and one using only clinical data (Table 2)—were trained to detect a cancer signal and were independently validated (see STAR Methods). In both training and validation sets, sensitivity at a *post hoc* level of 98% specificity (2% false-positive rate) was reported and statistically compared for each of the 10 classifiers (Table 3 and Figure 2; see STAR Methods; a relatively high specificity level was chosen because it is important for a population screening test not to burden the population with false-positive results). The top classifiers by sensitivity at 98% specificity were whole-genome (WG) methylation (training, 39%; validation, 34%), single nucleotide variant (SNV) with paired white blood cell (WBC) background removal (SNV-WBC; training, 36%; validation, 33%), and the pan-feature classifier (training, 40%; validation, 36%). In the validation set, sensitivity of the WG methylation classifier at 98% specificity was significantly greater

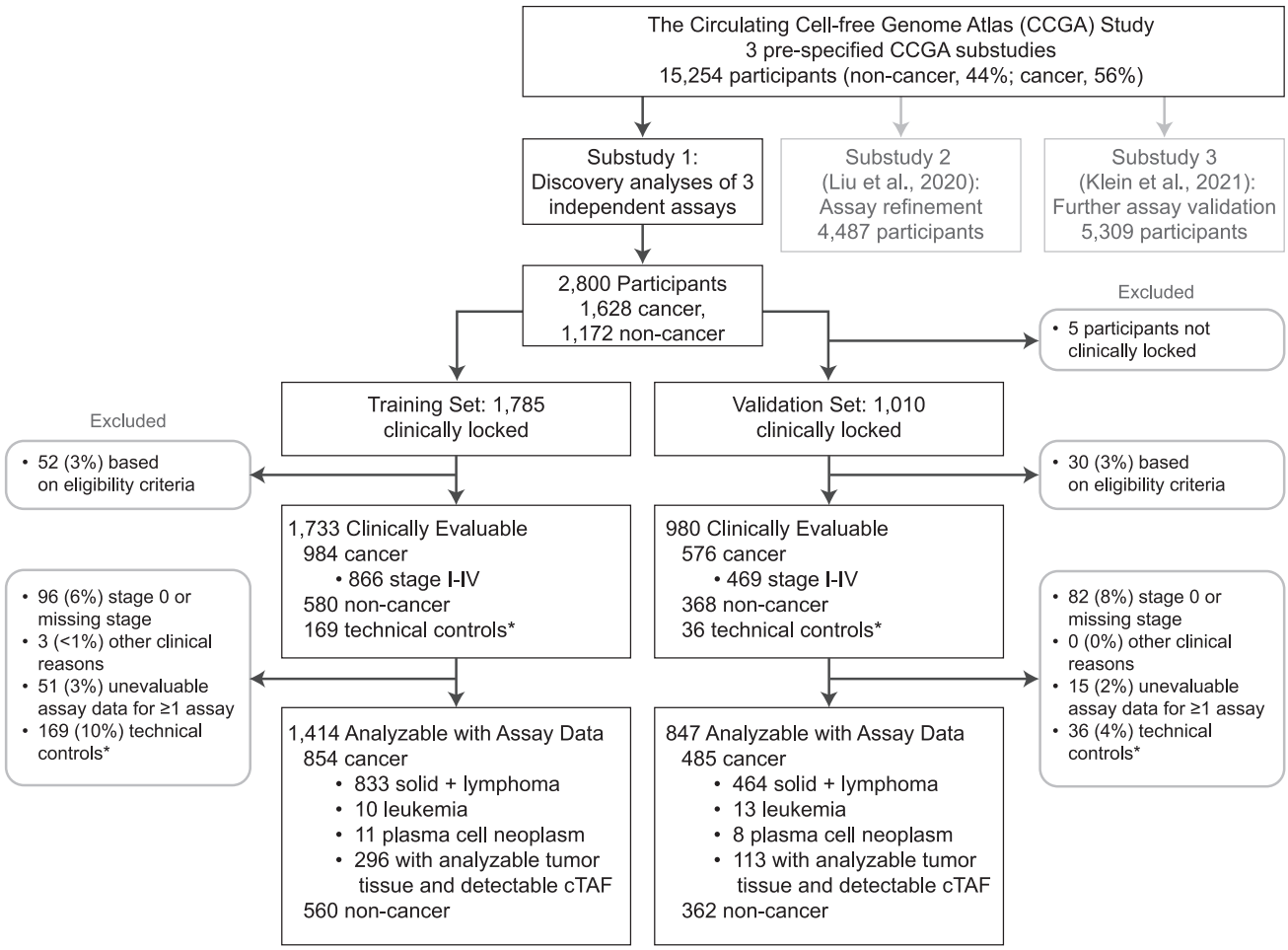


Figure 1. CCGA substudy 1 participant disposition

*Technical controls: non-smoking, non-cancer participants under the age of 35 years. CCGA, Circulating Cell-free Genome Atlas; cTAF, circulating tumor allele fraction.

(McNemar test, $p < 0.01$) than all but the pan-feature and the SNV-WBC classifier. SNV and somatic copy number alteration (SCNA) classifiers showed statistically worse sensitivity than WG methylation (SNV, 16%; SCNA, 27%); to achieve sensitivity similar to WG methylation, the SNV classifier required WBC sequencing to remove biological background due to CH (SNV-WBC, 33%). Accounting for CH for the SCNA classifier improved its sensitivity performance (30%), although the SCNA-WBC classifier was still less sensitive than the top three classifiers. Among the classifiers that were derived from WG sequencing (WGS), the fragment lengths classifier provided sensitivity similar to that of the SCNA-WBC classifier (29% and 30%, respectively). The next-best WGS classifiers by sensitivity were SCNA (27%), allelic imbalance (22%), and fragment endpoints (18%). The classifier containing only clinical features performed poorly by comparison (3%). As expected, the sensitivity of all cancer classifiers increased with clinical cancer stage (Figure S1). Training and validation sets provided generally consistent sensitivity results (Figure 2).

To investigate classifier performance across methods based on circulating tumor allele fraction (cTAF; an estimate of the relative amount of tumor-specific mutations in participant cfDNA), which was strongly associated with cancer signal detection performance (see cTAF section below), we leveraged the concept of a limit of detection (LOD) and created a clinical LOD measure. For this analysis, clinical LOD of a single MCED test was defined as the cTAF at which the probability of detecting a cancer signal was at least 50% while maintaining a 98% specificity (see STAR Methods; note that clinical LOD should not be confused with analytical LOD, which typically indicates the lowest concentration of a specific analyte that can be detected 95% of the time using known replicates over a dilution series).[27] Four hundred nine participants for whom tumor tissue was available and SNVs were detected in both tissue and cfDNA were included (see STAR Methods). Similar to the sensitivity results for the cancer signal detection classifiers, WG methylation, SNV-WBC, and the pan-feature classifiers provided the lowest clinical LODs, but were not significantly different from one another. The SNV classifier required CH background removal to match the LOD performance of the WG methylation classifier (Figures 3 and S2). The next lowest clinical LOD from this study was from the SCNA-WBC classifier, followed by the fragment lengths classifier. For

Table 1. CCGA substudy 1 participant demographics

	Training (n = 1,414)		Validation (n = 847)	
	Cancer[a]	Non-cancer	Cancer[a]	Non-cancer
Total, n	854	560	485	362
Age, mean (SD)	61 (12)	60 (12)	62 (12)	59 (14)
Sex, n (%)				
Female	594 (70%)	436 (78%)	307 (63%)	235 (65%)
Race/ethnicity, n (%)				
White, non-Hispanic	737 (86%)	473 (84%)	400 (82%)	308 (85%)
Black or African American	54 (6%)	46 (8%)	33 (7%)	25 (7%)
Hispanic	43 (5%)	29 (5%)	31 (6%)	22 (6%)
Other	20 (2%)	12 (2%)	21 (4%)	7 (2%)
Age group, n (%)				
≥50 years	710 (83%)	452 (81%)	414 (85%)	274 (76%)
Smoking status, n (%)				
Never smoker	409 (48%)	322 (57%)	243 (50%)	183 (51%)
Body mass index, n (%)				
Normal/underweight	237 (28%)	147 (26%)	139 (29%)	84 (23%)
Overweight	275 (32%)	180 (32%)	160 (33%)	123 (34%)
Obese	341 (40%)	233 (42%)	186 (38%)	154 (43%)
Geographic region, n (%)				
Northeast	46 (5%)	53 (9%)	26 (5%)	25 (7%)
Midwest	150 (18%)	83 (15%)	127 (26%)	64 (18%)
West	167 (20%)	78 (14%)	104 (21%)	89 (25%)
South	491 (57%)	346 (62%)	228 (47%)	184 (51%)
Overall clinical stage, n (%)				
I	289 (34%)	–	163 (34%)	–
II	239 (28%)	–	141 (29%)	–
III	159 (19%)	–	75 (15%)	–
IV	157 (18%)	–	93 (19%)	–
Non-informative/missing	10 (1%)	–	13 (3%)	–
Method of diagnosis, n (%)				
Screening	293 (34%)	–	167 (34%)	–
Clinical presentation[b]	561 (66%)	–	317 (65%)	–

Based on clinically evaluable population with results on all three assays. CCGA, Circulating Cell-free Genome Atlas; SD, standard deviation.

[a]Analyzable cancers by training/validation: breast (339/170), lung (118/46), prostate (69/55), colon/rectum (45/39), kidney (26/13), uterus (27/9), pancreas (26/22), esophagus (24/7), lymphoma (22/18), head and neck (19/12), ovary (17/7), liver/bile duct/gallbladder (13/14), melanoma (9/8), cervix (13/8), plasma cell neoplasm (11/8), leukemia (10/13), thyroid (13/5), bladder (10/1), stomach (11/13), multiple primaries (6/0), anus (7/2), unknown primary/other (19/15).

[b]Clinical presentation includes all cancers detected by a method other than cancer screening. Includes incidental screening findings. Excludes one participant with missing test information.

each cancer signal detection classifier compared here, the clinical LOD mirrored sensitivity performances relative to one another.

To investigate whether combining results across sequencing assays should be expected to improve cancer signal detection, we used the top-three single-assay-based classifiers (WG methylation, SCNA-WBC, and SNV-WBC) to determine whether each identified cancer signals in different samples, or whether they mostly identified the same samples. Note that the pan-feature classifier was a similarly top-performing classifier but was excluded from this analysis because it already included information from all three assays. The majority of correctly detected cancer signals (true positives) across the top single-assay classifiers were from the same participants (Figure S3). Samples that were detected by only two classifiers were more likely to be detected by the WG methylation and the SNV-WBC classifiers. Non-cancer samples that were incorrectly classified as cancer (false positives) from all three assays were almost entirely different across classifiers (Figure S3). See Figure S4 for a visualization of the feature space for all classifiers with cancer signal classification from WG methylation indicated.

Table 2. Mapping of assays to samples, features, and classifiers

Assay	Sample type	Feature	Classifier name	
WGBS (30×)	cfDNA	WG methylation patterns for ≈30 million CpGs	WG methylation	pan-feature[a]
Targeted sequencing assay of 507 genes (60,000× raw depth, 3,000× unique depth)	cfDNA	small somatic variants	SNV	
	WBCs	small somatic mutations to remove noise	SNV-WBC	
WGS (30× for cfDNA and WBCs; 60× target depth for tumor tissue)	cfDNA	somatic copy number alterations	SCNA	
		fragment endpoints	fragment endpoints	
		fragment lengths	fragment lengths	
		allelic imbalance	allelic imbalance	
	WBCs	somatic copy number alterations to remove noise	SCNA-WBC	
	tumor tissue	variant calling for allele fraction estimation[b]	–	–
None	clinical data	age, smoking history, family history of breast/ovarian cancer	clinical data[c]	

cfDNA, cell-free deoxyribonucleic acid; SCNA, somatic copy number alterations; SNV, single nucleotide variants; WBC, white blood cell; WG, whole-genome; WGBS, whole-genome bisulfite sequencing; WGS, whole-genome sequencing.
[a]The pan-feature classifier was trained using scores from all of the individual assay-based classifiers.
[b]Variant allele fraction was used to estimate circulating tumor allele fraction, not in a classifier.
[c]The clinical data classifier used only clinical data (no assay data).

Cancer signal origin prediction

One top-performing cancer signal detection feature from each of the three assays (WG methylation, SCNA, and SNV-WBC) was chosen to create three cancer signal origin (CSO) classifiers. Each classifier was evaluated using the validation set of cancer samples that were jointly detected by all three corresponding cancer signal detection classifiers (n = 127; see STAR Methods). Among the three CSO classifiers, the WG methylation CSO classifier accurately predicted CSO for 75% (95/127) of jointly detected cancer samples, whereas SCNA and SNV-WBC accurately predicted CSO for 41% (52/127) and 35% (44/127) of cancer samples, respectively (Figure 4). WG methylation predicted CSO with significantly more accuracy than either SCNA (McNemar test, $p = 8 \times 10^{-9}$) or SNV-WBC ($p = 6.5 \times 10^{-12}$). The CSO prediction accuracy of the SCNA classifier versus the SNV-WBC classifier was not statistically different ($p = 0.35$). Similar accuracy for all CSO classifiers resulted when each classifier was evaluated using the full set of validation cancer samples detected by that particular cancer signal detection classifier rather than the jointly detected set (see STAR Methods for a description of validation sets; Figure S5).

cTAF

cTAF varied by orders of magnitude within and across cancer types and clinical stages (Figure S6). In cancer types with sufficient samples to analyze correlative trends (breast, colon/rectum, and lung; Figure 5), cTAF increased with clinical cancer stage (Spearman rank correlation test: breast $p = 1.4 \times 10^{-10}$, colon/rectum $p = 2.1 \times 10^{-10}$, lung $p = 7.8 \times 10^{-3}$, and remaining cancer types $p = 3.2 \times 10^{-15}$). Although not powered to draw definitive conclusions, notable differences in cTAF within clinical stage and across cancer types were observed (Figure S6). For example, early-stage prostate cancer showed samples with lower cTAF, whereas early-stage lung and colon/rectum cancers showed samples with higher cTAF. cTAF was an important covariate of cancer signal detection and accounted for 72% of the variance in WG methylation cancer signal detection classifier scores (Figure S7). Further, in a multivariate analysis of cancer signal detection versus log10(cTAF) and cancer type and clinical stage, cTAF was the only significant predictor of classifier performance ($p < 5 \times 10^{-17}$; Table S1). Biological differences in shedding rate and the strong influence of cTAF on classifier performance may explain the differences in sensitivity between cancer types compared by clinical stage and support that cTAF is the main driver of cancer signal detection. If cTAF was omitted from the multivariate analysis, then clinical stage was significant ($p < 0.0002$). Taken together, these results indicate that much of the predictive power of clinical stage for signal detection is mediated through cTAF differences by stage within cancer type and motivate the utility of clinical LOD for assessing cancer signal detection performance.

CH somatic variants

The frequency of recurrent SNVs among participant-matched WBCs in the CCGA population was calculated to determine how often the same CH mutations were observed. WBC-matched SNVs accounted for 97% (7,275/7,504) of all SNVs detected in cfDNA from participants without cancer and 69% (11,125/16,169) of the SNVs in cfDNA from participants with solid cancers. Importantly, 90% (16,571/18,400) of all WBC-matched SNVs were detected at <1% plasma variant allele fraction (Figure S8A). Approximately 80% (14,607/18,400) of these SNVs were non-synonymous, with the majority of them (93% = 11,705/12,601) unique to individual participants (Figure S8B).

Clinical LOD from the second CCGA substudy

Based on the data reported here from the first CCGA substudy, WG methylation was selected (see the discussion) for further

Table 3. Performance metrics at 98% specificity for prototype cancer signal detection classifiers

		Training set		Validation set	
		Sensitivity at 98% specificity[a]		Sensitivity at 98% specificity[a]	
Assay	Classifier	% (95% CI)	TP/total cancer samples, n	% (95% CI)	TP/total cancer samples, n
WGBS	WG methylation	39% (36%–43%)	328/833	34% (30%–39%)	158/464
Targeted sequencing	SNV	19% (16%–22%)	159/833	16%[b] (13%–20%)	75/464
	SNV-WBC	36% (33%–39%)	299/833	33% (29%–38%)	155/464
WGS	SCNA	33% (29%–36%)	271/833	27%[b] (23%–31%)	125/464
	SCNA-WBC	33% (30%–37%)	278/833	30%[c] (26%–34%)	139/464
	fragment endpoints	22% (19%–25%)	181/833	18%[b] (15%–22%)	84/464
	fragment lengths	28% (25%–32%)	236/833	29%[c] (25%–34%)	136/464
	allelic imbalance	25% (22%–28%)	210/833	22%[b] (18%–26%)	101/464
All three	pan-feature	N/A[d]	N/A[d]	36% (31%–40%)	165/464
None	clinical data[e]	2.7% (1.7%–4.1%)	22/815	2.6%[b] (1.4%–4.5%)	12/457

CI, confidence interval; N/A, not available; SCNA, somatic copy number alteration; SCNA-WBC, somatic copy number alterations with correction for clonal hematopoiesis noise; SNV, single nucleotide variant; SNV-WBC, single nucleotide variants with correction for clonal hematopoiesis noise; TP, true positive; WG, whole-genome.

[a]At the target specificity of 98%, the observed specificity for the training set was 97.9% (true negatives/total non-cancer samples: 548/560) for all classifiers except the clinical data classifier (98.0% [540/551[e]] for the clinical data classifier). The observed specificity for the validation set was 97.8% for all classifiers (354/362 for all but clinical data; 350/358[e] for the clinical data classifier).

[b]$p < 0.0001$. The p values were computed only for the validation set and represent paired McNemar analysis versus WG methylation.

[c]$p < 0.01$.

[d]Performance of the pan-feature classifier is not reported for the training dataset. See text for details.

[e]The clinical data classifier used fewer cases in the training and validation sets than the other classifiers because of a missing clinical variable for some participants.

development and improvement using a targeted methylation approach. The clinical LOD for the refined, targeted methylation MCED test used in the second CCGA substudy[19] was 3.1×10^{-4} cTAF at the 99.3% specificity reported for that test (n = 641 participants with available tumor tissue from the second CCGA substudy validation set; see STAR Methods). To allow direct performance comparison with the prototype tests evaluated here, we also calculated the targeted methylation clinical LOD at 98% specificity using only solid cancers from the second substudy validation set[19] to match the specificity and cancer-type subset here (see STAR Methods). The targeted methylation clinical LOD at 98% specificity was 1.3×10^{-4} cTAF (n = 559 participants with solid cancer and available tumor tissue from the second CCGA substudy validation set; Figures 3 and S2). Both clinical LODs for the targeted methylation MCED test were almost an order of magnitude better than the top classifiers evaluated here.

DISCUSSION

To our knowledge, this initial CCGA substudy is the only comprehensive genome-wide comparison of cfDNA-based approaches. The rigorous multi-center study design included balanced demographics, clinical stage distribution, and cancer types; careful randomization of samples to avoid batch effects; and independent training and validation sets to provide robust data with which to evaluate different detection methods using the same set of participants. The general terms "liquid biopsy" and "circulating tumor DNA (ctDNA)" have been indiscriminately used to describe MCED tests,[28] but there are a number of genomic features that can be derived from cfDNA that have different performance characteristics, as investigated here. As such, it is critical to describe in detail the specific features and approaches used to allow proper comparison of results between studies.

A key finding of this study was that most of the variation in cfDNA cancer signal across cancer types and clinical stages could be attributed to the relative amount of tumor-specific genomic features in circulation, cTAF, and in particular, more than could be attributed to cancer type and clinical stage. Tumor biopsy sequencing to estimate cTAF across cancer types and stages provided a survey of the distribution of shedding across diagnosed cancers. In general, cTAF increased with increasing clinical stage, compatible with a general relationship of increased tumor shedding with advancing stage within cancer types.[29] Importantly, extensive variation in cTAF was observed between cancer types as well as within single stages, which indicated that stage alone may not be the sole predictor of the amount of tumor-specific genomic features and, further, that cancer types may have dramatically different shedding rates even after controlling for stage. As expected,[19] cancer signal detection improved for each cancer with increasing stage and increasing cTAF, although not all stage IV cancer signals were detected. This may be explained by molecular factors in undetected stage IV cancers such as low mitotic and metabolic activity,[30,31] or physical factors such as low surface area and low microscopic tumor extent (i.e., access to the blood supply)[30]; all of which have been associated with lower tumor DNA shedding and lower cTAF. Clinical staging may not completely capture tumor behavior. Conversely, cancer signals were detected from stage I–III cancers that showed higher cTAF, which may

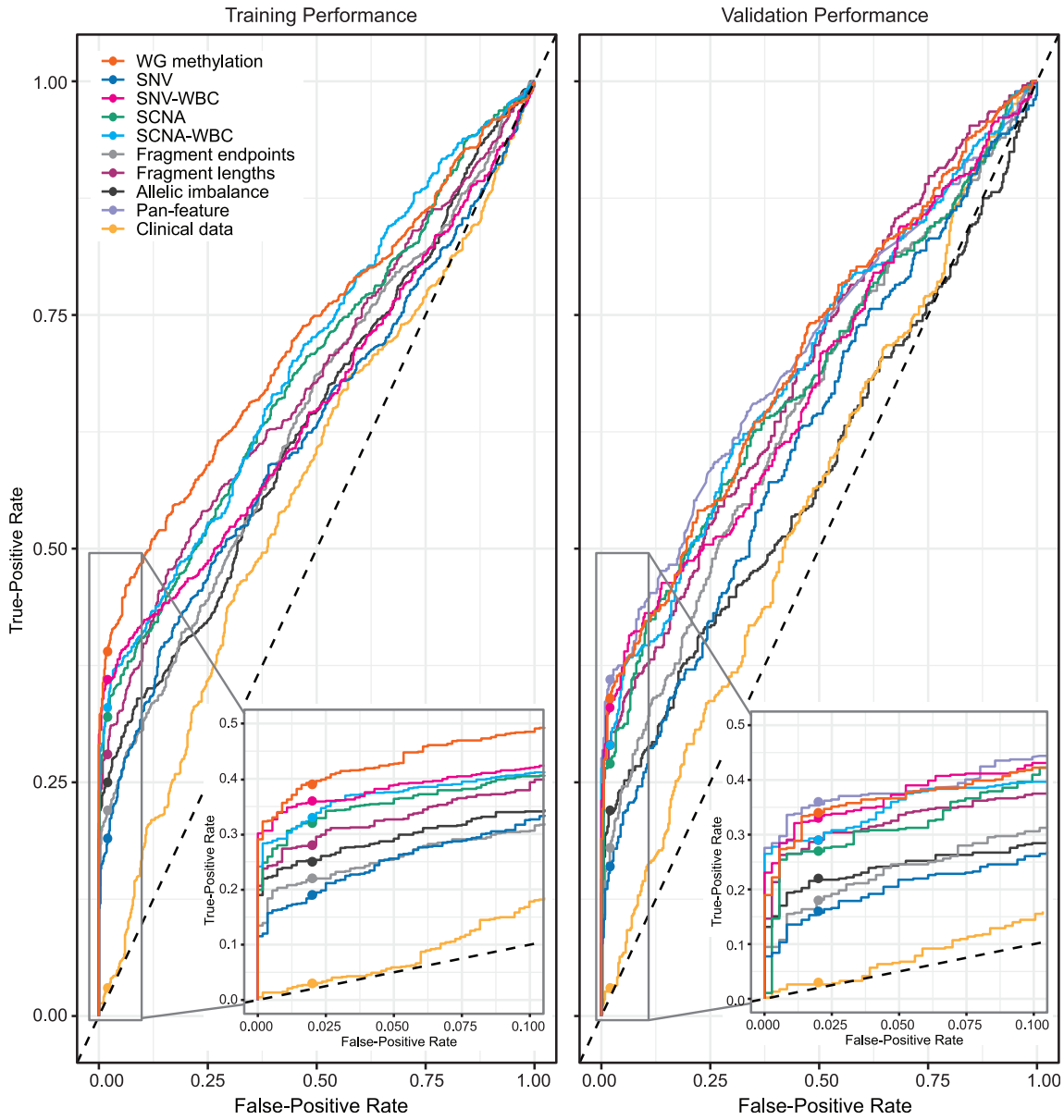


Figure 2. Performance of different classifiers for cancer signal detection

Receiver-operating characteristic curves demonstrate the relationship between the false-positive and the true-positive rates for each trained classifier. Greater area under the curve indicates better performance, and the dotted black line indicates random chance. After each classifier was trained on the training set data, its performance was assessed on the independent validation set data. Performance achieved on the training set by cross-validation (left) was similar to that achieved on the validation set (right), indicating that the classifiers were generalizable and not overfit to the training data. The sensitivity (true-positive rate) of each classifier at a target *post hoc* specificity cutoff of 98% (false-positive rate of 2%) is indicated by a circle on each line in the zoomed-in plots (bottom). See Table S2 for partial areas under the receiver-operating characteristic curves from the validation set. SCNA, somatic copy number alterations; SCNA-WBC, somatic copy number alterations with white blood cell noise removal; SNV, single nucleotide variants; SNV-WBC, single nucleotide variants with white blood cell noise removal; WG, whole-genome.

reflect active proliferation and high tumor DNA shedding.[16,17] In addition, in a model attributing classifier cfDNA cancer signal detection to cancer type, clinical stage, and cTAF, only cTAF was a statistically significant contributor. Coupled with previous research indicating that more aggressive cancers tend to shed more DNA into the bloodstream,[29] it is plausible that cfDNA cancer signals and cTAF are correlated with more aggressive tumor biology; thus, cfDNA assays may preferentially detect clinically significant cancers. Indeed, cancer prognosis has been shown to be better with lower cTAF.[32–35] Further, cancers detected by cfDNA tests have shown survival compatible with that reported in the Surveillance, Epidemiology, and End Results (SEER) program, wherein those cancers that were not detected by cfDNA showed significantly better survival.[29] Altogether, these data suggest that cTAF may be a more direct and accurate measure of the underlying tumor biology driving cfDNA cancer signal

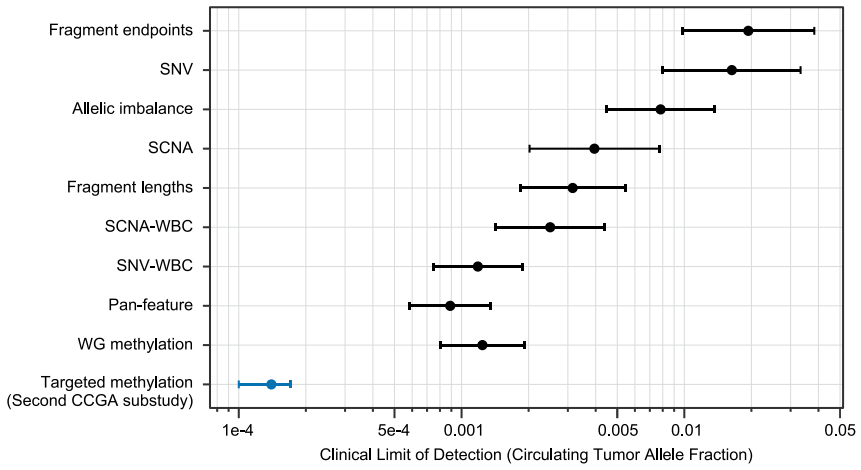


Figure 3. Clinical LOD for each cancer signal detection classifier

The cTAF was used to determine a clinical LOD at 98% specificity for each cancer signal detection classifier. The clinical LOD, defined as the cTAF that corresponded to 50% cancer signal detection, was estimated using the subset of samples that had corresponding cancer tissue sequencing available. Results from the validation set are shown here for simplicity. See Figure S2 for training set results. The blue data point indicates the estimated clinical LOD at 98% specificity for the improved targeted methylation assay that was used in the second CCGA substudy.[19] Note that the clinical LOD for the targeted methylation MCED test used in the second CCGA substudy was 3.1×10^{-4} cTAF at the 99.3% specificity reported for that test.[19] For all black data points, n = 113 samples with available tumor tissue from the first CCGA substudy validation set. For the blue data point, n = 559 samples with available tumor tissue from the second CCGA substudy validation set. Error bars indicate 95% confidence intervals. CCGA, Circulating Cell-free Genome Atlas; cTAF, circulating tumor allele fraction; LOD, limit of detection; SCNA, somatic copy number alterations; SCNA-WBC, somatic copy number alterations with white blood cell noise removal; SNV, single nucleotide variants; SNV-WBC, single nucleotide variants with white blood cell noise removal; WG, whole-genome.

detection than current prognostic indicators such as stage and cancer type, and that strategies for optimization of MCED tests should include efforts to improve detection at lower cTAF levels. The strong relationship between classifier signal detection and cTAF in this study motivated the creation of a clinical LOD using cTAF to provide a metric for classifier optimization that accounted for the extensive shedding variation within stage and between cancer types. The clinical LOD presented here may enable direct comparison of cancer signal detection performance between different cfDNA-based assays and studies, provided the methods are compared at equivalent levels of test specificity using tumor-biopsy-verified features to estimate cTAF.

Among all cfDNA features explored in this study, WG methylation was among the most sensitive of methods, did not require WBC sequencing, displayed one of the lowest clinical LODs, and had the highest CSO prediction accuracy. The non-methylation classifiers we developed (SCNA, fragment endpoints, fragment lengths, and allelic imbalance) were similar to other work,[36–38] but assessed on a held-out validation cohort using samples collected contemporaneously from the same participants as for the methylation assay. In the assessment, fragmentomic-type features distinguished between cancer and non-cancer but predominantly at higher tumor fractions compared with methylation, which achieved an order of magnitude better clinical LOD in a targeted methylation assay. In terms of sensitivity, paired WBC sequencing was required for classifiers that used SNVs and SCNAs to match or approach the clinical LOD and sensitivity performance of the other top classifiers. This may be because SCNA and SNV features are confounded by biological phenomena such as CH, whereas methylation patterns are easier to distinguish from background. Coupled with the observation that most WBC CH variants were unique to each participant, it is likely necessary to sequence WBCs in addition to cfDNA on an individual level to achieve high enough specificity for any SNV- or SCNA-based MCED test. Such a need may complicate clinical implementation by way of increased costs and test complexity.[39,40] Clinical LODs mirrored the relative sensitivity results for each classifier, further supporting WG methylation as the most promising approach. Importantly, there was little to no complementarity of other genomic features when combined in a pan-feature classifier to improve clinical LOD or sensitivity above WG methylation, which was most likely due to the shared level of ctDNA. As such, there may not be any significant added value from the complexity of including multiple cfDNA genomic features in an MCED test. The WG bisulfite sequencing (WGBS) CSO classifier (based on WG methylation) was significantly more accurate than the classifiers representing WGS and targeted sequencing (TS), which further indicated that WG methylation was the most promising approach. Accurate CSO predictions have the potential to help inform the downstream diagnostic workup. The fact that WG methylation was the most promising overall option in this study across classifier sensitivity, clinical LOD, and CSO prediction may be, in part, because it is a pervasive signal across the genome (≈30 million CpGs). Also, methylation patterns along each fragment contain a robust tumor-specific signal and can be readily identified above normal genomic background variation, which may enable the methylation signal to be detected at lower cTAF levels than the other features of the cancer genome tested here.

The WGBS assay, which generated WG methylation features, also had the most potential for improvement among the three assays and, as noted, was selected for further development. In this study, WGS and TS were performed with significant sequencing depth and breadth (30× and 60,000× covering 507 genes, respectively). Both technical noise removal (using unique molecular identifiers) and CH suppression (using WBCs) were applied to these data, and so the results here likely represent the upper limit of performance for WGS and TS assays for a practical MCED test. However, for the bisulfite sequencing assay that extracted WG methylation patterns, targeting the most informative CpG-containing regions predicted to contain cancer- and tissue-specific methylation patterns in cfDNA would allow greatly increased effective sequencing depth while controlling complexity to pursue improvements toward better clinical LOD. Because of its simplicity, superior overall performance, and potential for assay

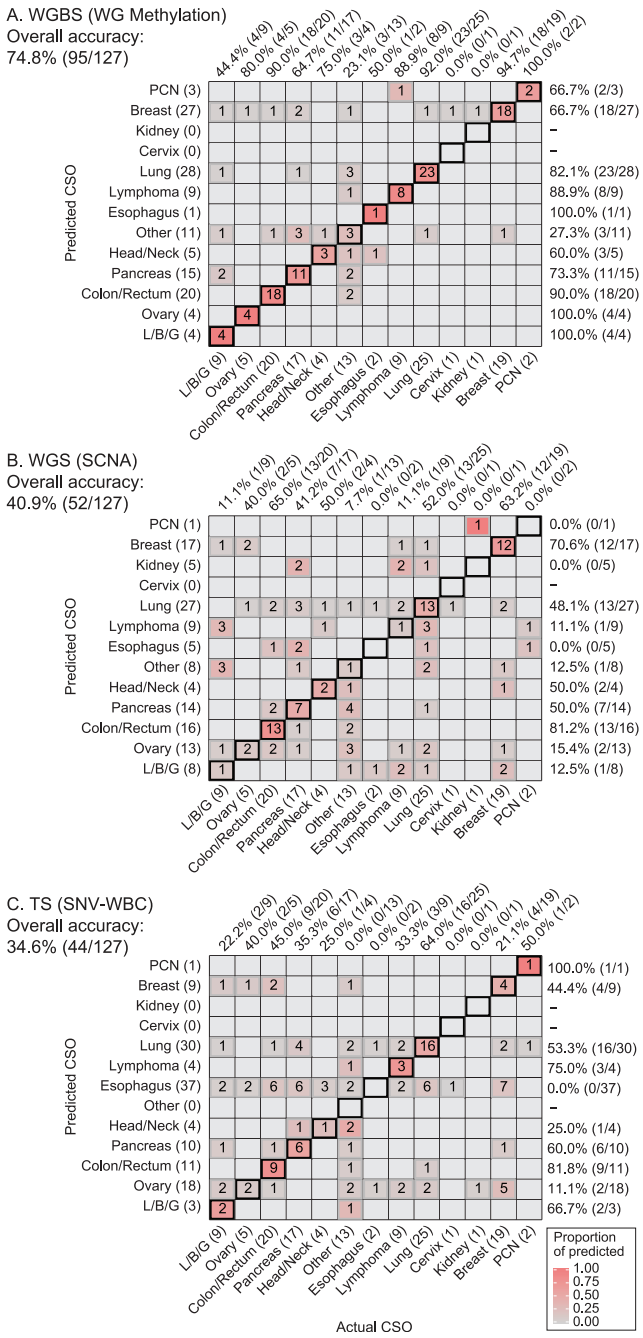


Figure 4. Accuracy of cancer signal origin prediction by prototype assay using the jointly detected validation set of cancer samples

Confusion matrices representing the accuracy of CSO prediction. Agreement between the true (x axis) and the predicted (y axis) CSO per sample using experimental CSO classifiers from each of the three assays is presented. Color corresponds to the proportion of predicted CSOs (y axis), which were correct (x axis). Included participants (n = 127) are those predicted as having cancer at 98% specificity by all three corresponding cancer signal detection classifiers. Row rates represent the probability of yielding a correct localization for CSO predictions, column rates represent the prediction accuracy for cancers within each CSO. Sample sizes are indicated in parentheses in the figure. CSO, cancer signal origin; L/B/G, liver/bile duct/gallbladder; PCN, plasma cell neoplasm; SCNA, somatic copy number alterations; SNV-WBC, single nucleotide variants with white blood cell noise removal; TS, targeted sequencing; WGBS, whole-genome bisulfite sequencing; WG, whole-genome; WGS, whole-genome sequencing.

optimization, a methylation-based approach was chosen for further development. Indeed, the clinical LOD for the targeted methylation classifier that was validated in the second CCGA substudy[19] showed almost an order-of-magnitude improvement relative to WG methylation. Importantly, performance improvements (specificity, sensitivity, and CSO accuracy) were observed in subsequent, recently reported large-scale clinical validations of the targeted-methylation-based assay and classifier, including large-scale validation that supported the clinical implementation of the Galleri® MCED test.[4,19] Our findings add to the existing literature on the use of cfDNA methylation patterns for cancer detection.[41,42]

Limitations of the study

This study was not without limitations. First, the sensitivities reported here were based on sensitivity across various cancers and clinical stages represented in the study population. Therefore, sensitivity is skewed, as sensitivity is expected to improve with increased stage. This renders direct comparison of sensitivity values in this study population to a real-world screening population complex, but the relative performance of the features tested here would likely remain similar. It is notable that cancer samples were from participants with already diagnosed cancer and that cTAF may vary between preclinical and recently diagnosed cancers. Although sensitivity is vulnerable to differences in population, clinical LOD was designed precisely to allow generalization to other populations by accounting for variations in cTAF, which we observed may be a critical biological property of test performance. Clinical LOD allowed robust comparison between cfDNA approaches here and, compared with sensitivity, should enable fairer comparisons of cfDNA approaches in different populations. Although one-third of participants with cancer in this study were diagnosed through screening, and were presumed asymptomatic at the time of that screening, it was not known whether clinical symptoms had developed before blood draw. Second, participants who were enrolled without known cancer did not receive diagnostic testing to demonstrate the absence of cancer, so clinical truth was based on a negative cancer screening test result; a non-cancer participant with a positive signal could not be definitively known to be a false positive. Participants are currently undergoing long-term follow-up, which may further clarify their cancer status. Other ongoing studies, including an interventional study, are currently addressing MCED test performance in asymptomatic populations (NCT04241796, NCT03085888, NCT03934866). Third, the small sample sizes for some cancers and resulting bias toward cancers with more samples may have led to errors in the accuracy of predicted CSOs. Another potential reason for these errors is the heterogeneous biology of cancers, which could have led to confusion between CSO labels. Of note, these errors are reduced in the second[19] and third[4] CCGA substudies with the targeted methylation assay. Fourth, the study population of this study included a slightly larger female and white population than is reflective of the general US population, although, as described,

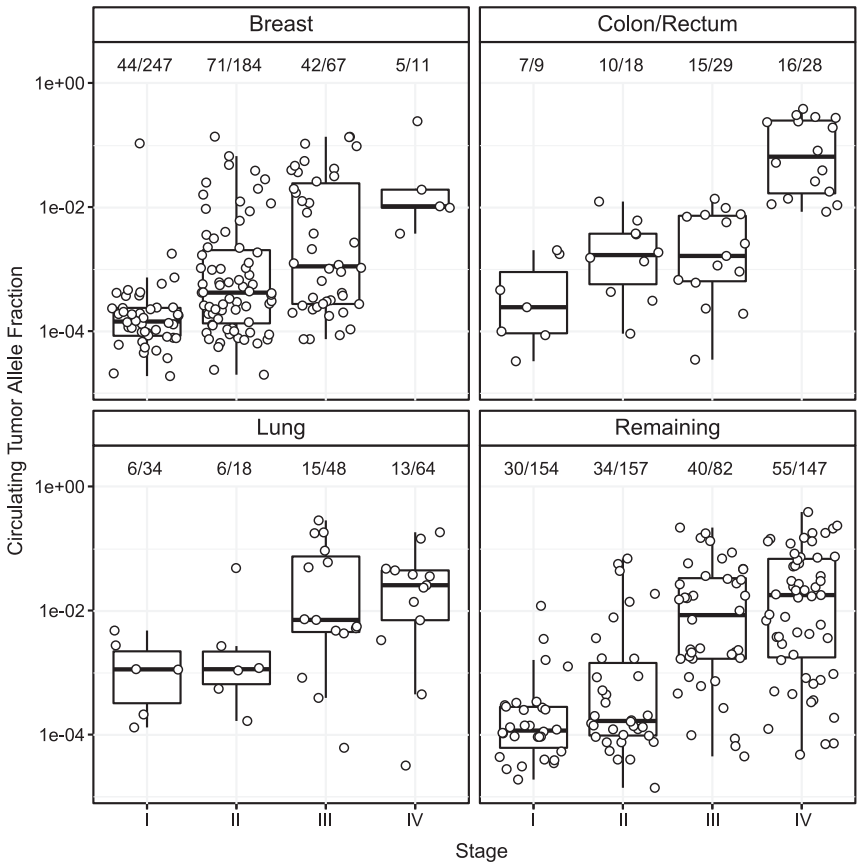


Figure 5. Distribution of cTAF by cancer stage

Each dot in the box plots represents the cTAF measurement from one participant on a log10 scale. Text insets above each box indicate the following numbers: participants with analyzable tumor tissue/total participants with cancer from the analyzable set of cancer cases. The total number of participants with analyzable tumor tissue for this analysis was n = 409. The total number of analyzable cfDNA samples from participants with cancer was n = 1,297. Boxes represent the interquartile range and whiskers represent the largest and smallest values no more than 1.5 × IQR. cfDNA, cell-free deoxyribonucleic acid; cTAF, circulating tumor allele fraction.

the clinical LOD metric is expected to generalize to a real-world population. Effort should be taken to improve representation in future studies. Fifth, the data presented here are also restricted to one blood draw at a single point in time, because multiple time points were out of scope of this study, although the second CCGA substudy[19] does include a second blood draw for non-cancer participants. Open questions related to this include how each method performs upon repeated use in individuals, and how each performs over time as cancer progresses and/or is treated and in remission. Longer-term follow-up data from the CCGA study may address such questions in the future. Last, in this study, sensitivity was assessed at only 98% specificity, but higher specificity is clinically preferred for a screening population. Even higher sensitivity at specificity >99% was reported in subsequent CCGA substudies.[4,19]

Conclusions

This first CCGA substudy showed that clinical LOD is a useful benchmark to assess classifier performance and that cTAF accounts for most of the variation in cfDNA cancer signal across cancer types and clinical stages, more so than clinical staging or typing information. For these reasons, clinical LOD is an attractive candidate to enable comparison between studies, as long as specificity and detection probabilities are equivalent. Further, WG methylation from cfDNA used in a prototype MCED test provided the best performance among the approaches characterized for cancer signal detection and CSO prediction, without requiring additional sequencing to correct for WBC background. These results informed the design and performance of the recently reported targeted-methylation-based cfDNA MCED test,[19] which showed substantial improvements relative to the top-performing prototype tests evaluated here and formed the basis for the Galleri® MCED test.[4]

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODELS AND SUBJECT DETAILS
  - Study design and participants
  - Inclusion/exclusion criteria
  - Human subject consent
- METHOD DETAILS
  - Patient data collection and oversight
  - WGBS: WG methylation classifier - Brief description
  - WGBS: WG methylation classifier - Detailed description
  - TS: SNV and SNV-WBC classifiers - Brief description

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2022.10.022.

ACKNOWLEDGMENTS

The authors would like to thank all individuals who participated in the CCGA study, as well as all study staff. The following former employees of GRAIL, LLC, are also acknowledged for important contributions to this work while they were employed at GRAIL, LLC: Tara Maddala, Darya Filippova, M. Cyrus Maher, Anton Valouev, Shilpen Patel, Jessica Yecies, Jonathan Newman, Xiao Yang, Jonathan M. Toung, Karthik Jagadeesh, Shivani Nautiyal, Cosmos Nicolaou, Richard Rava, Jeanne Yue, Hui H. Xu, Ling Shen, Sante Gnerre, Ravi Vijaya Satya, Mark Lee, Amy Sehnert, Alexander Blocker, Richard T. Williams, Richard D. Klausner, Chandra Subhachandra, Anne-Renee Hartman, Meredith Halks Miller, and Catalin Barbacioru. Chris Chang, an employee of GRAIL, LLC, is acknowledged for contributions to the work. Rita Lopatin, an employee of GRAIL, LLC, is also acknowledged for assistance in reviewing the manuscript.

This work was supported by GRAIL, LLC (Menlo Park, CA). Medical writing, logistical, and editorial support was funded by GRAIL, LLC. Medical writing support was provided by Alexis Fedorchak, PhD; Grace Wang, PharmD; Ruhi Ubale, PhD; and Megan Hall, PhD, from GRAIL, LLC. Logistical support was provided by LeAnn Berens and Rebecca Myers from Prescott Medical Communications Group (Chicago, IL). Editorial support for text and figures was provided by Erin Spohr from ENGAGE Labs, LLC (Oak Ridge, NJ), and Kristi Whitfield, PhD, from PosterDocs (Oakland, CA).

AUTHOR CONTRIBUTIONS

Conceptualization, A.J., O.V., E.H., N.Z., A.M.A., and C.S.; methodology, A.J., O.V., E.H., J.F.B., S.G., C.M., A.P.F., Q.L., H.A., C.B., D.C., P.F., R.C., K.D., C.H., R.J., B.J., V.D., J.N., J.N., O.S., E.S., S.S., K.K.S., V.N., T.C.C., N.H., and M.D.; software, A.J., O.V., E.H., J.F.B., S.G., C.M., A.P.F., Q.L., H.A., P.F., R.C., K.D., J.N., O.S., E.S., S.S., V.N., and T.C.C.; validation, A.J., O.V., E.H., J.F.B., S.G., C.M., A.P.F., Q.L., H.A., P.F., R.C., K.D., J.N., O.S., E.S., S.S., V.N., and T.C.C.; formal analysis, A.J., O.V., E.H., J.F.B., S.G., C.M., A.P.F., Q.L., H.A., P.F., R.C., K.D., J.N., O.S., E.S., S.S., V.N., T.C.C., C.B., C.H., R.J., B.J., V.D., K.K.S., N.H., M.D., N.Z., K.N.K., S.F., S.T., A.S., S.B., and Z.D., investigation, N.H., M.D., K.N.K., S.F., D.C., M.C.L., E.A.K., E.T.F., D.A.R., T.J.Y., A.L.C., D.D.T., D.A.B., M.K.T., K.M., M.A.S., A.B., and M.V.S; writing – original draft, A.J., O.V., and E.H.; writing – review & editing, A.J., O.V., E.H., S.S., C.B., C.H., R.J., B.J., V.D., K.K.S., N.H., M.D., K.N.K., S.F., D.C., M.C.L., E.A.K., E.T.F., D.A.R, T.J.Y., A.L.C., D.D.T., D.A.B., M.K.T., K.M., M.A.S., A.B., M.V.S., J.F.B., S.G., C.M., A.P.F., Q.L., H.A., P.F., R.C., K.D., J.N., O.S., E.S., V.N., T.C.C., N.Z., S.T., A.S., S.B., Z.D., A.M.A., and C.S.; visualization, A.J., O.V., E.H., and J.F.B.; supervision, A.J. and O.V.

DECLARATION OF INTERESTS

A.J., O.V., E.H., J.F.B., S.G., Q.L., N.Z., E.T.F., K.N.K., H.A., C.B., D.C., K.D., S.F., C.H., R.J., B.J., S.T., C.M., V.D., J.N., O.S., E.S., A.S., S.S., K.K.S., V.N., A.P.F., T.C.C., S.B., N.H., M.D., Z.D., and M.P.H. are employees of GRAIL, LLC, with equity in Illumina, Inc. C.M. also holds stock in Novartis, Clovis, Cara, Gilead, and Bluebird. M.C.L. is an uncompensated consultant for GRAIL, LLC. The Mayo Clinic was compensated for M.C.L.'s and D.D.T.'s advisory board activities for GRAIL, LLC. E.A.K. is a consultant for GRAIL, LLC. D.A.R. is a consultant for Ipsen. M.A.S. is a consultant for Celgene, Millennium, and Syros Pharmaceuticals. A.M.A. was previously employed by GRAIL, LLC; has equity in Illumina, Inc.; is currently employed by Illumina, Inc.; and is an advisor to and an equity holder in Foresite Labs and Myst Therapeutics. M.V.S. is an employee of and holds stock in McKesson Corporation, and is a clinical adviser for GRAIL, LLC. D.A.B. is a co-owner of Berry Consultants, LLC. A.H.B. is a consultant for Pfizer, Merck, Bayer, and Astellas Pharmaceuticals. C.S. holds stock in Illumina, Inc., Epic Biosciences, and Apogen Biotech; receives grants from Pfizer and AstraZeneca; receives honoraria or consultant fees from Roche Ventana, Celgene, Pfizer, Novartis, Genentech, and BMS; and is a co-founder of Achilles Therapeutics. S.G., O.V., A.P.F., A.J., K.D., V.N., J.F.B., C.M., E.H., Q.L., N.Z., P.F., and O.S. are inventors on pending patent applications related to this work, for which GRAIL, LLC, has ownership rights. GRAIL, LLC, a subsidiary of Illumina, Inc., is currently held separate from Illumina, Inc., under the terms of the Interim Measures Order of the European Commission dated October 29, 2021. C.S. recieved grant support from AstraZeneca, Boehringer-Ingelheim, Bristol Myers Squibb, Pfizer, Roche-Ventana, Invitae (previously Archer Dx Inc), and Ono Pharmaceutical; is an AstraZeneca Advisory Board member and Chief Investigator for the AZ MeRmaiD 1 and 2 clinical trials and is also Co-Chief Investigator of the NHS Galleri trial funded by GRAIL and a paid member of GRAIL's Scientific Advisory Board (SAB); received consultant fees from Achilles Therapeutics (also SAB member), Bicycle Therapeutics (also a SAB member), Genentech, Medicxi, Roche Innovation Centre – Shanghai, Metabomed, and the Sarah Cannon Research Institute; received honoraria from Amgen, AstraZeneca, Pfizer, Novartis, GlaxoSmithKline, MSD, Bristol Myers Squibb, Illumina, and Roche-Ventana; had stock options in Apogen Biotechnologies and GRAIL until June 2021, and currently has stock options in Epic Bioscience, Bicycle Therapeutics, and Achilles Therapeutics; and is a co-founder of Achilles Therapeutics; holds patents relating to assay technology to detect tumour recurrence (PCT/GB2017/053289), targeting neoantigens (PCT/EP2016/059401), identifying patent response to immune checkpoint blockade (PCT/EP2016/071471), determining HLA LOH (PCT/GB2018/052004), predicting survival rates of patients with cancer (PCT/GB2020/050221), and identifying patients who respond to cancer treatment (PCT/GB2018/051912); holds US patent relating to detecting tumour mutations (PCT/US2017/28013), methods for lung cancer detection (US20190106751A1); and holds both a European and US patent related to identifying insertion/deletion mutation targets

(PCT/GB2018/051892). C.S. is a Royal Society Napier Research Professor (RSRP\R\210001) and has received funding from the Francis Crick Institute that receives its core funding from Cancer Research UK (CC2041), the UK Medical Research Council (CC2041), and the Wellcome Trust (CC2041); Cancer Research UK (TRACERx [C11496/A17786], PEACE [C416/A21999], and CRUK Cancer Immunotherapy Catalyst Network); Cancer Research UK Lung Cancer Centre of Excellence (C11496/A30025); the Rosetrees Trust, Butterfield and Stoneygate Trusts; NovoNordisk Foundation (ID16584); Royal Society Professorship Enhancement Award (RP/EA/180007); National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre; the Cancer Research UK-University College London Centre; Experimental Cancer Medicine Centre; the Breast Cancer Research Foundation (US) (BCRF-22-157); Cancer Research UK Early Detection and Diagnosis Primer Award (Grant EDDPMA-Nov21/100034); The Mark Foundation for Cancer Research Aspire Award (Grant 21-029-ASP); Stand Up To Cancer-LUNGevity American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (Grant Number: SU2C-AACR-DT23-17); and an ERC Advanced Grant (PROTEUS) from the European Research Council under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 835297).

Received: October 6, 2021
Revised: August 3, 2022
Accepted: October 26, 2022
Published: November 17, 2022

REFERENCES

1. Mandel, P., and Metais, P. (1948). Comptes rendus des séances de la Société de biologie et de ses filiales [Nuclear acids in human blood plasma]. C. R. Seances Soc. Biol. Fil. *142*, 241–243.
2. Lo, Y.M. (2000). Fetal DNA in maternal plasma: biology and diagnostic applications. Clin. Chem. *46*, 1903–1906.
3. Oliveira, K.C.S., Ramos, I.B., Silva, J.M.C., Barra, W.F., Riggins, G.J., Palande, V., Pinho, C.T., Frenkel-Morgenstern, M., Santos, S.E.B., Assumpcao, P.P., et al. (2020). Current perspectives on circulating tumor DNA, precision medicine, and personalized clinical management of cancer. Mol. Cancer Res. *18*, 517–528. MCR-19-0768. https://doi.org/10.1158/1541-7786.
4. Klein, E.A., Richards, D., Cohn, A., Tummala, M., Lapham, R., Cosgrove, D., Chung, G., Clement, J., Gao, J., Hunkapiller, N., et al. (2021). Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. Ann. Oncol. *32*, 1167–1177. https://doi.org/10.1016/j.annonc.2021.05.806.
5. Cohen, J.D., Li, L., Wang, Y., Thoburn, C., Afsari, B., Danilova, L., Douville, C., Javed, A.A., Wong, F., Mattox, A., et al. (2018). Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science *359*, 926–930. https://doi.org/10.1126/science.aar3247.
6. Lennon, A.M., Buchanan, A.H., Kinde, I., Warren, A., Honushefsky, A., Cohain, A.T., Ledbetter, D.H., Sanfilippo, F., Sheridan, K., Rosica, D., et al. (2020). Feasibility of blood testing combined with PET-CT to screen for cancer and guide intervention. Science *369*, eabb9601. https://doi.org/10.1126/science.abb9601.
7. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. *71*, 209–249. https://doi.org/10.3322/caac.21660.
8. Ferlay, J., Ervik, M., Lam, F., Colombet, M., Mery, L., Piñeros, M., et al. (2020). Global Cancer Observatory: Cancer Today (International Agency for Research on Cancer). https://gco.iarc.fr/today.
9. Siegel, R.L., Miller, K.D., and Jemal, A. (2020). Cancer statistics, 2020. Cancer J. Clin. *70*, 7–30. https://doi.org/10.3322/caac.21590.
10. Hawkes, N. (2019). Cancer survival data emphasise importance of early diagnosis. BMJ *364*, l408. https://doi.org/10.1136/bmj.l408.
11. American Cancer Society (2021). Cancer Facts & Figures 2021 (American Cancer Society). https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2021/cancer-facts-and-figures-2021.pdf.
12. Bibbins-Domingo, K., Grossman, D.C., Curry, S.j., Davidson, K.W., Epling, J.W., Jr., Garcia, F.A.R., Gillman, M.W., Harper, D.M., Kemper, A.R., Krist, A.H., et al. (2016). Screening for colorectal cancer: US preventive Services Task Force recommendation statement. JAMA *315*, 2564–2575. https://doi.org/10.1001/jama.2016.5989.
13. Curry, S.J., Krist, A.H., Owens, D.K., Barry, M.J., Caughey, A.B., Davidson, K.W., Doubeni, C.A., Epling, J.W., Jr., Kemper, A.R., Kubik, M., et al. (2018). Screening for cervical cancer: US preventive Services Task Force recommendation statement. JAMA *320*, 674–686. https://doi.org/10.1001/jama.2018.10897.
14. US Preventive Services Task Force, Grossman, D.C., Curry, S.J., Owens, D.K., Bibbins-Domingo, K., Caughey, A.B., Davidson, K.W., Doubeni, C.A., Ebell, M., Epling, J.W., Jr., et al. (2018). Screening for prostate cancer: US preventive Services Task Force recommendation statement. JAMA *319*, 1901–1913. https://doi.org/10.1001/jama.2018.3710.
15. Moyer, V.A.; US Preventive Services Task Force (2014). Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. Ann. Intern. Med. *160*, 330–338. https://doi.org/10.7326/M13-2771.
16. Siu, A.L.; US Preventive Services Task Force (2016). Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann. Intern. Med. *164*, 279–296.
17. Ahlquist, D.A. (2018). Universal cancer screening: revolutionary, rational, and realizable. NPJ Precis. Oncol. *2*, 23. https://doi.org/10.1038/s41698-018-0066-x.
18. Croswell, J.M., Kramer, B.S., Kreimer, A.R., Prorok, P.C., Xu, J.-L., Baker, S.G., Fagerstrom, R., Riley, T.L., Clapp, J.D., Berg, C.D., et al. (2009). Cumulative incidence of false-positive results in repeated, multimodal cancer screening. Ann. Fam. Med. *7*, 212–222. https://doi.org/10.1370/afm.942.
19. Liu, M.C., Oxnard, G.R., Klein, E.A., Swanton, C., and Seiden, M.V.; CCGA Consortium (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann. Oncol. *31*, 745–759. https://doi.org/10.1016/j.annonc.2020.02.011.
20. Hubbell, E., Clarke, C.A., Aravanis, A.M., and Berg, C.D. (2021). Modeled reductions in late-stage cancer with a multi-cancer early detection test. Cancer Epidemiol. Biomarkers Prev. *30*, 460–468. https://doi.org/10.1158/1055-9965.EPI-20-1134.
21. Cescon, D.W., Bratman, S.V., Chan, S.M., and Siu, L.L. (2020). Circulating tumor DNA and liquid biopsy in oncology. Nat. Can. (Que.) *1*, 276–290. https://doi.org/10.1038/s43018-020-0043-5.
22. Heitzer, E., Auinger, L., and Speicher, M.R. (2020). Cell-free DNA and apoptosis: how dead cells inform about the living. Trends Mol. Med. *26*, 519–528. https://doi.org/10.1016/j.molmed.2020.01.012.
23. Chan, H.T., Chin, Y.M., Nakamura, Y., and Low, S.K. (2020). Clonal hematopoiesis in liquid biopsy: from biological noise to valuable clinical implications. Cancers *12*, 2277. https://doi.org/10.3390/cancers12082277.
24. Chabon, J.J., Hamilton, E.G., Kurtz, D.M., Esfahani, M.S., Moding, E.J., Stehr, H., Schroers-Martin, J., Nabet, B.Y., Chen, B., Chaudhuri, A.A., et al. (2020). Integrating genomic features for non-invasive early lung cancer detection. Nature *580*, 245–251. https://doi.org/10.1038/s41586-020-2140-0.
25. Li, M., Xie, S., Lu, C., Zhu, L., and Zhu, L. (2021). Application of data science in circulating tumor DNA detection: a promising avenue towards liquid biopsy. Front. Oncol. *11*, 692322. https://doi.org/10.3389/fonc.2021.692322.
26. Keller, L., Belloum, Y., Wikman, H., and Pantel, K. (2021). Clinical relevance of blood-based ctDNA analysis: mutation detection and beyond. Br. J. Cancer *124*, 345–358. https://doi.org/10.1038/s41416-020-01047-5.
27. Sesler, C.L., and Grigorenko, E.V. (2018). Analytical validation of qPCR-based multivariate index assays in a clinical laboratory: practical

challenges and limitations. J. Appl. Lab. Med. *3*, 267–281. https://doi.org/10.1373/jalm.2017.025924.

28. Bettegowda, C., Sausen, M., Leary, R.J., Kinde, I., Wang, Y., Agrawal, N., Bartlett, B.R., Wang, H., Luber, B., Alani, R.M., et al. (2014). Detection of circulating tumor DNA in early- and late-stage human malignancies. Sci. Transl. Med. *6*, 224ra24. https://doi.org/10.1126/scitranslmed.3007094.
29. Chen, X., Dong, Z., Hubbell, E., Kurtzman, K.N., Oxnard, G.R., Venn, O., Melton, C., Clarke, C.A., Shaknovich, R., Ma, T., et al. (2021). Prognostic significance of blood-based multi-cancer detection in plasma cell-free DNA. Clin. Cancer Res. *27*, 4221–4229. https://doi.org/10.1158/1078-0432.CCR-21-0417.
30. Bredno, J., Lipson, J., Venn, O., Gross, S., Fields, A.P., Beausang, J.F., Liu, Q., Brooks, J.D., Chen, X., Lopatin, R., et al. (2020). Tumor area and microscopic extent of invasion to determine circulating tumor DNA fraction in plasma and detectability of colorectal cancer (CRC). J. Clin. Oncol. *38*, 243. https://doi.org/10.1200/JCO.2020.38.4_suppl.243.
31. Kustanovich, A., Schwartz, R., Peretz, T., and Grinshpun, A. (2019). Life and death of circulating cell-free DNA. Cancer Biol. Ther. *20*, 1057–1067. https://doi.org/10.1080/15384047.2019.1598759.
32. Unseld, M., Belic, J., Pierer, K., Zhou, Q., Moser, T., Bauer, R., Piringer, G., Gerger, A., Siebenhüner, A., Speicher, M., et al. (2021). A higher ctDNA fraction decreases survival in regorafenib-treated metastatic colorectal cancer patients. Results from the regorafenib's liquid biopsy translational biomarker phase II pilot study. Int. J. Cancer *148*, 1452–1461.
33. Stover, D.G., Parsons, H.A., Ha, G., Freeman, S.S., Barry, W.T., Guo, H., Choudhury, A.D., Gydush, G., Reed, S.C., Rhoades, J., et al. (2018). Association of cell-free DNA tumor fraction and somatic copy number alterations with survival in metastatic triple-negative breast cancer. J. Clin. Oncol. *36*, 543–553.
34. Chen, L., Zhang, Y., Cheng, Y., Zhang, D., Zhu, S., and Ma, X. (2018). Prognostic value of circulating cell-free DNA in patients with pancreatic cancer: a systemic review and meta-analysis. Gene *679*, 328–334. https://doi.org/10.1016/j.gene.2018.09.029.
35. Ocaña, A., Díez-González, L., García-Olmo, D.C., Templeton, A.J., Vera-Badillo, F., José Escribano, M., Serrano-Heras, G., Corrales-Sánchez, V., Seruga, B., Andrés-Pretel, F., et al. (2016). Circulating DNA and survival in solid tumors. Cancer Epidemiol. Biomarkers Prev. *25*, 399–406. https://doi.org/10.1158/1055-9965.EPI-15-0893.
36. Mouliere, F., Chandrananda, D., Piskorz, A.M., Moore, E.K., Morris, J., Ahlborn, L.B., Mair, R., Goranova, T., Marass, F., Heider, K., et al. (2018). Enhanced detection of circulating tumor DNA by fragment size analysis. Sci. Transl. Med. *10*, eaat4921. https://doi.org/10.1126/scitranslmed.aat4921.
37. Cristiano, S., Leal, A., Phallen, J., Fiksel, J., Adleff, V., Bruhm, D.C., Jensen, S.Ø., Medina, J.E., Hruban, C., White, J.R., et al. (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. Nature *570*, 385–389. https://doi.org/10.1038/s41586-019-1272-6.
38. Mathios, D., Johansen, J.S., Cristiano, S., Medina, J.E., Phallen, J., Larsen, K.R., Bruhm, D.C., Niknafs, N., Ferreira, L., Adleff, V., et al. (2021). Detection and characterization of lung cancer using cell-free DNA fragmentomes. Nat. Commun. *12*, 5060–5064. https://doi.org/10.1038/s41467-021-24994-w.
39. Adalsteinsson, V.A., Ha, G., Freeman, S.S., Choudhury, A.D., Stover, D.G., Parsons, H.A., Gydush, G., Reed, S.C., Rotem, D., Rhoades, J., et al. (2017). Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nat. Commun. *8*, 1324. https://doi.org/10.1038/s41467-017-00965-y.
40. Hu, Y., Ulrich, B.C., Supplee, J., Kuang, Y., Lizotte, P.H., Feeney, N.B., Guibert, N.M., Awad, M.M., Wong, K.-K., Jänne, P.A., et al. (2018). False-positive plasma genotyping due to clonal hematopoiesis. Clin. Cancer Res. *24*, 4437–4443. https://doi.org/10.1158/1078-0432.CCR-18-0143.
41. Moss, J., Magenheim, J., Neiman, D., Zemmour, H., Loyfer, N., Korach, A., Samet, Y., Maoz, M., Druid, H., Arner, P., et al. (2018). Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat. Commun. *9*, 5068. https://doi.org/10.1038/s41467-018-07466-6.
42. Shen, S.Y., Singhania, R., Fehringer, G., Chakravarthy, A., Roehrl, M.H.A., Chadwick, D., Zuzarte, P.C., Borgida, A., Wang, T.T., Li, T., et al. (2018). Sensitive tumour detection and classification using plasma cell-free DNA methylomes. Nature *563*, 579–583. https://doi.org/10.1038/s41586-018-0703-0.
43. Razavi, P., Li, B.T., Brown, D.N., Jung, B., Hubbell, E., Shen, R., Abida, W., Juluru, K., De Bruijn, I., Hou, C., et al. (2019). High-intensity sequencing reveals the sources of plasma circulating cell-free DNA variants. Nat. Med. *25*, 1928–1937. https://doi.org/10.1038/s41591-019-0652-7.
44. Jung, M., and Pfeifer, G.P. (2015). Aging and DNA methylation. BMC Biol. *13*, 7. https://doi.org/10.1186/s12915-015-0118-4.
45. Jones, M.J., Goodman, S.J., and Kobor, M.S. (2015). DNA methylation and healthy human aging. Aging Cell *14*, 924–932. https://doi.org/10.1111/acel.12349.
46. Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. J. Stat. Software *33*, 1–22.
47. Kingma, D.P., and Ba, J. (2014). Adam: a method for stochastic optimization (v9). Preprint at arXiv. https://doi.org/10.48550/arXiv.1412.6980.
48. Filippova, D., Larson, M.H., Maher, M.C., Calef, R., Pimentel, M., Zhou, Y., Newman, J., Gross, S., Nicula, V., Liu, T.-C., et al. (2019). The Circulating Cell-free Genome Atlas (CCGA) study: size-selection of cell-free DNA (cfDNA) fragments. J. Clin. Oncol. *37*, 3103. https://doi.org/10.1200/JCO.2019.37.15_suppl.3103.
49. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McVean, G.A., McCarthy, S., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74. https://doi.org/10.1038/nature15393.
50. Loh, P.R., Danecek, P., Palamara, P.F., Fuchsberger, C., A Reshef, Y., K Finucane, H., Schoenherr, S., Forer, L., McCarthy, S., Abecasis, G.R., et al. (2016). Reference-based phasing using the haplotype reference consortium panel. Nat. Genet. *48*, 1443–1448. https://doi.org/10.1038/ng.3679.
51. Chen, T., and Guestrin, C. (2016). XGBoost: a scalable tree boosting system KDD '16. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. https://doi.org/10.1145/2939672.2939785.
52. Riedmiller, M., and Braun, H. (1992). RPROP - a fast adaptive learning algorithm. In Proc. of ISCIS VII, Universitat.
53. Stuart, J.M., Weinstein, J.N., Collisson, E.A., Mills, G.B., Shaw, K.R.M., Ozenberger, B.A., Ellrott, K., Shmulevich, I., and Sander, C.; Cancer Genome Atlas Research Network (2013). The Cancer Genome Atlas pan-cancer analysis project. Nat. Genet. *45*, 1113–1120. https://doi.org/10.1038/ng.2764.
54. Allaoui, M., Kherfi, M.L., and Cheriet, A. (2020). Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study. ICISP *5*, 317–325. https://doi.org/10.1007/978-3-030-51935-3_34.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Critical commercial assays		
QIAamp® Circulating Nucleic Acid Kit	Qiagen	Cat #: 55114
EZ DNA Methylation-Lightning™ MagPrep Kit	Zymo Research	Cat #: D5046/D5047
Accel-NGS™ Methyl-Seq DNA Library Kit	Swift BioSciences	Cat #: 30024/30096
KAPA™ Library Quantification Kit	Kapa Biosystems	Cat #: KK4873
AccuClear® Ultra High Sensitivity dsDNA Quantitation Kit	Biotium	Cat #: 31029
DNeasy® Blood & Tissue Kit	Qiagen	Cat # 69504/69506
Chemicals, peptides, and recombinant proteins		
Agencourt AMPure XP magnetic beads	Beckman Coulter	Cat #: A63882
Deposited data		
Processed, de-identified genomic data and code to recreate the figures and tables	This study	https://github.com/grailbio-publications/Jamshidi_CCGA1LOD_2022
Software and algorithms		
Glmnet package	Friedman et al., 2010[46]	
Adam optimizer	Kingma and Ba, 2014[47]	
Eagle2	Loh et al., 2016[50]	
XGBoost	Chen and Guestrin, 2016[51]	
Rprop	Riedmiller and Braun, 1992[52]	
Uniform Manifold Approximation and Projection	Allaoui et al., 2020[54]	
Other		
NanoDrop™	Thermo Scientific	Cat #: ND-1
E220 focused-ultrasonicator	Covaris	Cat #: 500239
HiSeq X flow cell	Illumina	Cat #: 15072585
NovaSeq 6000 S2 flow cell	Illumina	Cat #: 20015845

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Oliver Venn (ovenn@grailbio.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Assay summary statistics, demographic data, and code to reproduce the figures and tables in this paper have been deposited at Github and are publicly available. A link to the Github repository is listed in the key resources table.
- The genomic sequencing data used in this study have not been deposited in a public repository because we consider it personal and protected patient data.

EXPERIMENTAL MODELS AND SUBJECT DETAILS

Study design and participants

The CCGA study (NCT02889978) is a multi-center, observational, case-controlled study that included three pre-specified substudies and is undergoing prospective longitudinal follow-up. The first substudy to comprehensively analyze multiple potential approaches

for an MCED test is reported here (Figure 1). The second substudy developed a targeted methylation assay and classifier,[19] and the third substudy validated the targeted methylation approach developed in the second substudy.[4] Enrollment in the CCGA study was monitored to ensure representation of a broad set of cancer types. For the first CCGA substudy, participants with cancer and without cancer (controls) were randomized to training and validation sets in batches.

Participants with cancer and non-cancer controls were enrolled in the CCGA study from the same centers or, when not possible, geographic regions, to ensure consistency between cancer and control groups among race, ethnicity, and body mass index, and to control for pre-analytic factors. Characteristics of the enrolled population were monitored for each center and for the study overall to ensure, within reason, that a diverse and fair representation of participants was contributed from each location. Age was frequency-matched between cancer and non-cancer participants selected for the training and validation sets. Cancer type and stage were monitored to fulfill a distribution of multiple cancer types. The number of participants allowed to be enrolled from a single site was limited to no more than 40% of the overall study. In line with the initial focus of this substudy, the participants selected for the initial sequencing (first four batches) included only women. Men were included in subsequent sequencing batches to achieve a broader set of target cancers. Men were distributed equally across subsequent batches, resulting in a higher proportion of men in the validation set than in the training set.

Non-cancer participants were enrolled at an approximate ratio of three non-cancer participants for every seven cancer participants within each individual clinical research center also enrolling cancer participants. Non-cancer participants were enrolled from mammography and colonoscopy clinics as well as general medical clinics, hematology clinics, cancer risk clinics, and blood draw areas. Potential participants were screened utilizing procedures available at approved enrollment centers. Enrollment of non-cancer participants was monitored to ensure a distribution that reflected demographic characteristics within each individual research center of the cancer population (eg, by age, gender, ethnicity, and smoking status).

Clinical information, demographics, and medical data relevant to cancer status were collected on all participants and their medical record at baseline (time of biospecimen collection). Annual updates through medical record review for up to 5 years were mandatory; contacting investigators for additional clinical insights from existing medical records was allowed and was not intended to alter clinical care. A future blood collection may be requested from study participants during the follow-up period, but is not a scheduled event. The work reported here is based on data obtained at baseline.

Inclusion/exclusion criteria

Non-cancer participants were required to be aged $\geq$20 years and able to provide written informed consent. Potential non-cancer participants were enrolled if there was no cancer diagnosis at the time of enrollment, although no verification (eg, by PET-CT) was done and no 1-year follow-up was done to exclude non-cancer participants with subsequent cancer diagnosis. Cancer participants were required to be aged $\geq$20 years, able to provide written informed consent, and have either of the following: (1) a confirmed cancer diagnosis of any stage (I-IV or carcinoma *in situ*) within 90 days prior to study blood draw, based on assessment of a pathological specimen (including, but not limited to, biopsy from primary tumor site, lymph node, or metastatic lesion; or cytology specimen, or bone marrow or blood specimen for hematological malignancies); or (2) a high suspicion for a cancer diagnosis by clinical and/or radiological assessment, with planned biopsy or surgical resection to establish a definitive diagnosis within 6 weeks (42 days) after study blood draw.

Potential cancer and non-cancer participants were excluded for a known current or prior diagnosis of cancer (with the exception of participants with a history of non-melanoma skin cancer [eg, basal cell carcinoma or squamous cell carcinoma] that was effectively and exclusively managed by local or focal therapies such as surgical resection, radiation therapy, cryotherapy, or topical therapy); for cancer participants, this was separate from the confirmed or suspected cancer diagnosis associated with study enrollment. Additionally, potential cancer participants were excluded for current or prior receipt of any of the following therapies for treatment of their current cancer: surgical management of the cancer beyond that required to establish the cancer diagnosis; local, regional, or systemic chemotherapy, including chemoembolization; targeted therapy; immunotherapy, including cancer vaccines; hormone therapy, or radiation therapy. Potential cancer and non-cancer participants were also excluded for oral or intravenous corticosteroid use in the 14 days prior to blood draw; pregnancy (by self-report of pregnancy status); current febrile illness, acute exacerbation or flare of an inflammatory condition requiring escalation in medical therapy within 14 days prior to blood draw; recipient or organ transplant; prior non-autologous (allogeneic) bone marrow of stem cell transplant; poor health status, or inability to tolerate blood draw. For cancer participants, submission of a pathological tumor specimen to GRAIL, LLC was not required for study enrollment.

Human subject consent

All CCGA-eligible participants provided informed consent as documented via a language-appropriate written informed consent form that was approved by the governing Institutional Review Board/Ethics Committee, and met all of the inclusion criteria and none of the exclusion criteria.

Approval from local institutional review boards to recruit participants was obtained by 142 sites in the United States and one in Canada. Sites were selected based on their ability to enroll appropriate participants and to perform the study in accordance with the principles of the International Conference on Harmonization for Good Clinical Practice; the provisions specified in Title 21 Parts 50, 54, 56, and 812 of the U.S. Code of Federal Regulations; and, where applicable, all federal, provincial, state, and local laws of pertinent regulatory authorities. The Investigators agreed to abide by these regulations, as well as the study protocol.

METHOD DETAILS

Patient data collection and oversight

Sample collection, processing, and preparation

De-identified blood samples were collected from participants at sites in the United States and Canada as previously described.[4,19] Eight 10 mL tubes of peripheral blood were collected using Streck Cell-Free DNA BCT® for each participant and shipped to BioStorage Technologies, Inc. (Indianapolis, IN). Whole blood was isolated into plasma and buffy coat, and stored at −80 °C at BioStorage Technologies, Inc. The median time between blood collection and plasma isolation was <2 days (range, 1–5 days). Cell-free DNA was extracted from the plasma of each patient using a modified QIAamp Circulating Nucleic Acid Kit (Qiagen; Germantown, MD). Two tubes of plasma were used for each of the WGBS and WGS cfDNA libraries. One tube of buffy coat was used for participant-matched WGS libraries of WBC genomic DNA. As discussed below, the same DNA library was used for TS and WGS. Selected patient plasma and buffy coat samples were transferred to GRAIL, LLC and processed.

Tumor tissue is typically collected for standard-of-care diagnostic and treatment purposes. When a pretreatment tumor tissue biopsy was available for enrolled participants, study sites submitted formalin-fixed, paraffin-embedded (FFPE) cell or tissue blocks or a series of stained and unstained glass slides to GRAIL, LLC for research use. Sites were instructed to prepare one 5 μm-thick slide stained with hematoxylin & eosin (H&E), then a series of 10 unstained slides at 10 μm-thickness, followed by a final H&E slide at 5 μm, for a total set of 12 slides. Blocks that were received at GRAIL, LLC were sectioned and stained in a similar manner. The H&E slides were evaluated by board-certified anatomic pathologists using upright brightfield microscopes. Areas of tumor were identified on the H&E slides and circled with a marking pen to define a region of interest (ROI) that was at least 80% tumor area (if possible) and with an area of at least 1 mm^2. Pathologists also calculated the nuclear volume within the ROI by estimating the percent of the ROI area that was occupied by tumor nuclei. Based on these calculations, instructions were given to histotechnologists to dissect tissue from 1 to 10 unstained slides (depending on ROI area and percent nuclear volume) and place it in a micronics tissue vial so that the resulting nuclear volume was between 0.5 and 1.5 mm^3.

Sequencing

For the first CCGA substudy, genomic DNA was extracted from scrapings of FFPE tissue specimens by Hudson Alpha Institute for Biotechnology (Huntsville, AL). To calculate the LOD for the second CCGA substudy[19] genomic DNA was extracted from scrapings of FFPE tissue in-house (GRAIL, LLC; Menlo Park, CA) using the AllPrep FFPE DNA/RNA kit (Qiagen; Germantown, MD).

A cfDNA blood sample from each participant was processed using each of the three following sequencing methods: WGBS, TS, and WGS. When tumor tissue was available, it was sequenced using WGS.

WGBS

Paired-end WGBS at a depth of 30X was performed using sequencing libraries prepared from plasma cfDNA subjected to bisulfite conversion. Up to 75 ng of plasma cfDNA was subjected to bisulfite conversion using the EZ DNA Methylation-Lightning MagPrep Kit (Zymo Research, D5046/D5047). Converted cfDNA was used to prepare dual-indexed sequencing libraries using Accel-NGS Methyl-Seq DNA Library Kits (Swift BioSciences; Ann Arbor, MI) and constructed libraries were quantified using KAPA Library Quantification Kit for Illumina Platforms (Kapa Biosystems; Wilmington, MA). Four libraries along with 10% PhiX v3 library (Illumina, FC-110-3001) were pooled and clustered on an Illumina NovaSeq 6000 S2 flow cell followed by 150-bp paired-end sequencing (30X).

TS

Paired-end TS at a raw depth of 60,000X (3,000X unique depth) was performed using amplified WGS libraries prepared from plasma cfDNA and buffy coat WBC genomic DNA that were enriched for a panel of 507 cancer-related genes (2.13 Mb; GRAIL, LLC; Menlo Park, CA).[43] Genomic DNA (gDNA) from buffy coat was extracted using Qiagen DNeasy Blood and Tissue Kit, and was quantified using NanoDrop (Thermo Scientific; Waltham, MA). Extracted gDNA was fragmented using Covaris E220 focused-ultrasonicator (Woburn, MA), and was size-selected using Agencourt AMPure XP magnetic beads (Beckman Coulter; Beverly, MA). Plasma cfDNA (up to 75 ng) and buffy coat gDNA (75 ng) were used for NGS library construction. The adapter included a set of 218 unique molecular identifier (UMI) sequences to reduce assay and sequencing errors. A fraction of amplified libraries (4 μL of 25 μL) were diluted and quantified using AccuClear Ultra High-Sensitivity dsDNA Quantitation kit (Biotium; Fremont, CA); the remainder was used in the whole-genome sequencing protocol (see below). Up to 3.5 μg of each library underwent hybridization-based capture. The gene panel included full exons except for the telomerase reverse transcriptase gene and intronic regions, which were included for rearrangement detection of 28 genes and copy number aberration detection of 42 genes. The enriched libraries were quantified using AccuClear Ultra High Sensitivity dsDNA Quantitation kit. Three or four enriched libraries were normalized, pooled, clustered on a flow cell, and sequenced on Illumina HiSeq X (150-bp paired-end sequencing, 60,000X).

WGS

WGS was performed at a depth of 30× using next-generation sequencing (NGS) libraries (see TS section above). Three or four diluted libraries were normalized, pooled, clustered on a flow cell, and sequenced on an Illumina HiSeq X (30X).

WGS of tumor tissue

WGS was performed at a target depth of 60X for DNA prepared from FFPE tumor tissue. FFPE tumor tissue sections were scraped and sent to the Genome Services Lab at Hudson Alpha Institute for Biotechnology (Huntsville, AL), where DNA was extracted from the scrapings and converted into NGS libraries for whole-genome sequencing on an Illumina HiSeq X (60X target depth). For each tissue scraping, one tube of corresponding buffy coat was shipped to Hudson Alpha for extraction, library preparation, and WGS on Illumina HiSeq X at a target depth of 30X. Sequencing data were transferred to GRAIL, LLC for analysis.

Classification of cancer versus non-cancer

The original primary objectives of this CCGA substudy were to develop classification models to distinguish invasive breast cancer from non-cancer and other cancers in women. Based on evaluation of the training phase, and prior to unblinding the validation set, it was determined that achieving the prespecified goal of specificity and sensitivity for detecting invasive breast cancer was very unlikely. Therefore, the primary objectives were updated to prespecify the primary comparison of interest in the validation phase as all cancer versus non-cancer.

Machine-learning classifiers for determining cancer versus non-cancer were developed and uniquely customized using the following cfDNA genomic features (one classifier per feature): WG methylation patterns, SNV, SNV-WBC, SCNA, SCNA-WBC, fragment endpoints, fragment lengths, allelic imbalance, and a combination of all scores from the previous classifiers (pan-feature). One additional classifier was trained using only clinical characteristics. Table 2 shows the mapping from the three assays (WGBS, TS, WGS) to samples to features to classifiers. WG methylation data were generated from the WGBS assay, and SNV data were generated from the TS assay. SCNA, fragment endpoints, fragment lengths, and allelic imbalance data were each generated from the WGS assay. WBC biological noise data was generated from TS and WGS as indicated below. All cancer signal detection classifiers were trained and validated using the same set of solid cancer (carcinoma, sarcoma, lymphoma) samples across all stages to evaluate detection performance. This allowed parallel and comparative analysis of potential classification of cancer signals using each feature or multiple features simultaneously. The same ten-fold cross-validation strategy was used during training for all classifiers. Both brief and detailed descriptions for each classifier are included below.

For initial cancer signal detection for all ten classifiers, we restricted the samples to solid cancers, which were not expected to have circulating cancer background as a potential component of the WBC fraction. For CSO prediction, three classifiers (WG methylation, SCNA, and SNV-WBC) were later rerun with the previously restricted hematologic cancers (plasma cell neoplasm and leukemia).

WGBS: WG methylation classifier - Brief description

For the WG methylation classifier, a data-reduction step was done first to remove fragments (sequenced molecules) with methylation states commonly found in non-cancer samples using a Markov model trained on a set of reference non-cancer samples. After this reduction, the remaining fragments were further filtered to retain those that were either mostly hyper- or mostly hypo- methylated and included at least 5 CpG (5′-cytosine-phosphate-guanine-3′) sites. Across the genome, a probability model was trained to generate a likelihood that an extreme fragment at that location would be found in non-cancer samples or cancer samples. Finally, for each sample, a set of the likelihoods of the most cancer-defining (highly ranked) hypo- and hyper- methylated fragments were retained as features. These cancer-defining features were then used in a kernel logistic regression classifier to predict whether a cancer signal was present.

WGBS: WG methylation classifier - Detailed description

For each sample, the WGBS fragment set was reduced to a subset of unusual fragments of extreme methylation status (UFXM). Fragments occuring at high frequency in individuals without cancer, or that have unstable methylation, are unlikely to produce highly discriminatory features for classification of cancer status. We therefore produced a statistical model of typical fragments using an independent reference set of 108 participants without cancer (median age: 58 ± 14 years, 79 [73%] women and 45% ever-smokers) from the main CCGA study. These samples were independent from the under-35 non-smokers used as technical controls by the WGS and targeted assays below, as methylation status is known to change with age in healthy individuals.[44,45] These samples were used to train a Markov chain model (order 3) estimating the likelihood of a given sequence of CpG (5′-cytosine-phosphate-guanine-3′) methylation statuses within a fragment. This model was demonstrated to be calibrated within the normal fragment range ($p \geq 0.001$) and was used to reject fragments with a p value from the Markov model $p \geq 0.001$ as insufficiently unusual.

A further data reduction step selected only fragments with at least 5 CpGs covered, and average methylation per fragment either ≥ 0.9 (hyper-methylated) or ≤ 0.1 (hypo-methylated). This procedure resulted in a median (range) of 2,800 (1,500–12,000) UFXM fragments for participants without cancer in training, and a median (range) of 3,000 (1,200–220,000) UFXM fragments for participants with cancer in training. Because this data reduction procedure only used reference set data, this stage was only required to be applied to each sample once.

At selected loci within the genome, an approximate log-ratio score for informativeness for cancer status was constructed separately for both hyper- and hypo-methylated UFXM. First, for each sample at the locus, a binary feature was generated: 0 if no UFXM fragment overlapped that locus within that sample; 1 if there existed a UFXM fragment overlapping the locus. The number of positive values (1s) in samples were then counted from participants with (C_c) and without cancer (C_{nc}). The log-ratio score was then constructed as: $\log(C_c+1)-\log(C_{nc}+1)$, adding a regularization term to the counts, and discarding the normalization term relating to the total number of samples within each group (N_c and N_{nc}) as it is constant ($\log[N_{nc}+2]-\log[N_c+2]$). Scores were constructed at the locations of all CpG sites within the genome, resulting in approximately 25 million loci with assigned scores: one score for UFXM hyper-methylated fragments and one score for UFXM hypo-methylated fragments.

Given a locus-specific log-ratio score, UFXM fragments in a sample were scored by taking the maximum of all log-ratio scores for loci within the fragment and matching the methylation category of either hyper- or hypo-methylated. This resulted in one score per UFXM fragment within a sample.

These fragment-level scores within a sample were reduced to a small set of features per sample by taking the scores of a subset of extreme-ranked fragments within each sample, separately for both hyper- and hypo-methylated fragments. In this way, information

for the most informative fragments in each sample was captured using a small set of useful features. In a low cTAF sample, only a minority of fragments were expected to be unusually informative.

In each category of fragments, the rank 1, 2, 4, …, 64 (2^i, i in 0:6) largest scores were selected for fragments within each category of hyper- and hypo-methylated UFXM, resulting in 14 features (7 and 7). To adjust for sample sequencing depth, the ranking procedure was treated as a function mapping ranks to scores, and we interpolated between the observed scores to obtain scores corresponding to adjusted ranks. The ranks were adjusted in linear proportion to relative sample depth: if the relative sample depth was x, interpolated scores were taken at x multiplied by the original ranks (eg, for x = 1.1, we took scores computed at ranks floor(1.1), floor(2.2), …, floor($x \cdot 2^i$)). Every sample was then assigned a set of 14 adjusted extreme-rank scores to be used in further classification.

Given the feature vector, a kernel logistic regression classifier was used to capture potential non-linearities in predicting cancer/non-cancer status from the features. Specifically, a regularized kernel logistic regression (KLR) classifier was trained using the isotropic radial basis function (power exponential 2) as the kernel with scale parameter gamma, and L2 regularization parameter lambda (adjusted by dividing by m^2, where m is the number of samples so lambda scales naturally with the amount of training data). Gamma and lambda were optimized for holdout log-loss using internal cross-validation within specified training data, and were optimized using grid-search over the range 1–0.01 (gamma), 1000–10 (lambda) in 7 multiplicative steps, starting at the maximum value and halving the parameter at each step. The median optimal parameters over internal cross-validation folds were 0.125 for gamma and 125 for lambda.

To evaluate performance of this extreme-rank score (ERS) classifier procedure on the first CCGA substudy dataset, cross-validation was applied to the training set, dividing the samples into 10 folds. Each fold was held out and the ERS classifier was trained on the remaining 9/10 of the data (using internal cross-validation within those folds to optimize gamma and lambda). The log-ratio scores used in featurization only accessed data from training folds. Output scores from each held-out fold were pooled and used to construct a ROC curve for performance.

For evaluating the validation set, the entire training dataset was used to construct scores and a single KLR classifier, which was then applied to the validation dataset.

Although the methylation classifiers did not use WBC sequencing, we synchronized the training and validation procedure with the other two assays.

TS: SNV and SNV-WBC classifiers - Brief description

For the SNV-based classifiers, after preprocessing (including quality filtering of variants), gene-level features were computed based on the frequency of variants expected to disrupt gene function (non-synonymous and splice), as well as gene-level relative copy number extracted from TS read depth. Variant presence per gene was analyzed with elastic net binomial logistic regression (using glmnet)[46] to derive a cancer signal versus non-cancer signal classification score. To assess the impact of controlling for WBC interference, genomic DNA from matched WBCs were sequenced to the same depth using the TS assay. Classification was performed with variants filtered for WBC interference and with variants not filtered for WBC interference. These two classifiers are the SNV-WBC (with WBC filter) and SNV (without WBC filter) classifiers.

TS: SNV and SNV-WBC classifiers - Detailed description

Within the 507 genes and intergenic regions pulled down by the targeted panel fragments with identical UMIs and endpoint positions were used to define groups of likely polymerase chain reaction (PCR) duplicates, which were collapsed (mean collapsed coverage: 3,000X) and combined to high-accuracy fragment sequences.[43] Candidate variants were generated using a de Bruijn graph assembler[43] and were scored by a noise model trained on data from 169 non-smoking participants younger than age 35 years without a diagnosis of cancer (technical controls), used to measure technical variation from the assay; data from 36 participants allocated for measuring technical variation in the validation set were reserved for future analyses (Figure 1). The noise model provided a calibrated quality score estimated on the support for each variant, allowing for filtering of candidate variants to a high-quality subset of variants unlikely to occur purely by technical variation. Candidate variants were further filtered against DNA damage artifacts that clustered near the ends of reads and occurred in a subset of samples. Variants that were estimated to have a Phred score of 60 or higher and were unlikely to be technical artifacts were retained.

Biological events that were unrelated to solid tumors and were related to the normal aging process were accounted for by applying the same targeted panel to buffy coat–extracted cells from the same blood sample. This allowed exclusion of hematopoietic sources for a large number of variants (SNV-WBC classifier). To accomplish this, a joint model of the probability of observing variant alleles in only one of the two samples was built, and any variants potentially derived from hematopoietic sources were excluded. This compensated for normal within-person accumulation of somatic variants in the blood.

After deriving a list of somatic mutations unlikely to be technical variation or clonal hematopoiesis, a fixed-length feature vector was constructed for every sample by assigning every gene the maximum allele fraction of any non-synonymous variant (including SNVs and indels) occurring within that region, or 0 if no variants were detected above noise within that region in that sample. Splicing variation and 5-prime upstream variation within telomerase reverse transcriptase (TERT) were treated as non-synonymous variants for this purpose. In this way, a variable length list of mutations was transformed into a real-valued vector.

Elastic net regression (using the glmnet package)[46] with binomial outcomes was used to process the above feature vector into a cancer or non-cancer classification score. Hyperparameters (alpha, the parameter governing ridge versus lasso, and lambda, the

penalty strength) were optimized using cross-validation within training folds, and then used to provide scores for held-out samples. Gene regions that artifactually appeared to be more frequently mutated in non-cancers than in cancers were suppressed by requiring non-negative coefficients.

Because WBCs are the main signal source of somatic mutations of certain liquid cancer types (eg, plasma cell neoplasm and leukemia), we included somatic copy number features derived from TS to compensate for the potential loss of sensitivity when WBC filtering was applied.

Read counts were utilized within each targeted region to estimate the presence of copy number aberrations within each gene. Read counts were log-transformed and normalized with respect to guanine-cytosine (GC), and systematic effects were removed using principal components derived from the same individuals used in training the noise model described above. To further remove noise specific to the targeted assay, two rounds of z-score normalization were applied to the normalized, transformed read counts using the same set of technical control samples described above. The final transformed and z-score normalized read counts were turned into a feature vector with one entry per gene. This set of features allowed for sensitivity to liquid cancers that may resemble clonal hematopoiesis by having matched variants present in the buffy coat. Similarly to above, both feature vectors were supplied to the elastic net regression to train and use in scoring samples.

The SNV classifier without WBC filtering proceeded as described above for the SNV-WBC classifier: gene-based feature vectors were constructed based on non-synonymous variants, and elastic net regression was used to predict cancer or non-cancer. The effect of removing deep sequencing of individually matched buffy coat samples was approximated by removing germline variants in each participant, but all other noise model–passed variants, which were originally filtered using individually matched WBC sequencing results, were retained.

WGS: SCNA and SCNA-WBC classifiers - Brief description

For the SCNA-based classifiers, the number of sequencing reads (read depth) in 100 kb genome bins was adjusted for technical variation using principal components derived from technical controls, reduced to normalized log-ratios, and corrected for variation in GC base composition (GC bias). This was performed in parallel in both cfDNA and WBC data that were sequenced to the same depth using WGS. A convolutional neural network was then trained on either the cfDNA normalized log-ratios or a combination of the cfDNA log-ratios and the matched WBC log-ratios to account for hematologic sources of SCNA. These two classifiers are the SCNA-WBC (with WBC log-ratios) and SCNA (without WBC log-ratios) classifiers.

WGS: SCNA and SCNA-WBC classifiers - Detailed description

For each sample in the first CCGA substudy, whole-genome sequencing was performed on cfDNA to an average depth of 30X. The genome was divided into bins of 100 kb, and the number of sequencing reads within or partially overlapping each bin was counted. Restricting the analysis to the autosomes, we transformed the raw sequencing reads by generalized logarithm, subtracted a baseline computed from non-cancer, non-smokers, who were aged <35 years (technical controls), GC corrected using a smooth spline function, and finally removed systematic effects using the first 5 principal components of remaining variation in the controls. Bins with unusually low depth or high variability among the controls were filtered, resulting in a vector of approximately 25,000 values per sample.

A convolutional neural net was trained on these vectors to distinguish cancer from non-cancer status. Treating each autosome as a channel, the network was structured alternating a convolutional layer (rectified linear unit activation) followed by a max-pooling layer. In this way, the data size was successfully reduced while allowing the network to learn what features needed to be retained.

Specifically, the first convolutional layer used a kernel of length 22 with 2 filters per channel and tiled across the bins with stride 4. This was followed by a max-pooling layer with length 3 and stride 3, so that after the first convolution and max-pooling layer group the output size on bins dimension was reduced by a factor of 12. This structure was repeated with another convolutional layer of length 11 stride 3 with one filter per channel, followed by another max-pooling layer of length 2 stride 2, reducing the output size on bins dimension again by a factor of 6. A final softmax node was densely connected to the output of all chromosome channels, to classify cancer or non-cancer status.

The Adam optimizer[47] was used with a staircase exponential decay for the learning rate, starting at 0.0005, with decay rate 0.95 after the first 20 training epochs. The network was trained for a total of 180 epochs with batch size 128. The network weights were initialized with truncated normal distributions with a standard deviation of 0.09. The network was regularized with L2 regularization loss on convolutional weights using a regularizing factor of 0.03.

Cross-validation was used to train classifiers for the other assays, dividing the samples into 10 folds, training on 9 folds, and applying the results to the held-out fold. A further normalization step was done within the cross-validation, in which additional systematic effects were removed by principal components analysis (PCA) as described above. In this case, half of the training-fold samples labeled as non-cancer were used to derive the normalizing components, and applied to the other half of the training fold before training the neural network. The full calling procedure on the held-out fold was then to apply the normalization procedure using the additional components learned from training, and then execute the network. None of the data from the held-out fold was used in either training the additional normalization or the network.

Finally, the same procedure was used to train on the entirety of the training set, and applied to the validation set.

The SCNA classifier used bin counts without information from WBC sequencing. To explore the potential effect of clonal hematopoietic and germline copy number aberrations on classification performance, sequencing results from matched WGS sequencing of

WBC were used to create an SCNA-WBC classifier. Similarly to cfDNA, bin counts from WBC sequencing were corrected for technical variation (GC bias, PCA correction). A statistical model for determining if bin values were derived from WBC aberrations was built (estimates of per-bin variation applied to determine significantly non-zero log-ratios in each WBC bin, similarly applied to segments), and used to supply a masking vector indicating if bin values were likely derived from either germline or hematopoietic aberrations. The neural network training was performed on the augmented dataset, supplying both masked bin-counts from cfDNA and matched WBC to parallel networks, combining at the last step.

WGS: Fragment endpoints classifier - Brief description

For the fragment endpoints classifier, cancer-enriched endpoint positions in the genome were identified using hierarchical Bayesian modeling of fragments that terminated at genome locations in non-cancer individuals to establish the base rate per position. Normalized counts of cancer-enriched endpoints in fragments that were likely from solid tumors due to short length (50–140 bp) were then input to a logistic regression to predict cancer or non-cancer signal status.

WGS: Fragment endpoints classifier - Detailed description

To generate endpoint features for classification, genome-wide positions that are significantly enriched in the endpoints of cancer fragments as compared to non-cancer fragments were determined. Given the underlying variation in cTF, only cancer samples with estimated high-tumor cTF (>5%) were used in the cancer training set. Fragments in both the cancer and non-cancer sets were in silico-filtered to shorter (50–140 nt) fragments because cell-free tumor-derived fragments are shorter than healthy cell-free DNA fragments.[48] The endpoint features were generated under 10-fold cross-validation as follows: the training fold's cancer and non-cancer sample fragment endpoints were aggregated into two respective count vectors each the length of the genome. On a per-chromosome basis, the significance of each position's cancer versus non-cancer count was assessed by the statistical procedure detailed below, resulting in a p value associated with each genomic position. Significant tiers of positions were defined by different p value thresholds decreasing by orders of magnitude from 10^{-1} to 10^{-20} (eg, all positions with $p < 10^{-1}$; positions with $p < 10^{-2}$; etc.). For each of the fold's held-out samples, the sample's number of endpoints at the significant positions was computed for each tier and used as a feature for classification. Using the held-out sample counts for all folds, the selected p value threshold was the significance tier that maximized sensitivity at 98% specificity. This yielded a p value threshold of 10^{-12}.

A hierarchical Bayes approach was taken to model the uncertainty in the rate of non-cancer fragments at each position given the single non-cancer count observation at each position from the aggregated set of non-cancers. The generative model used was that non-cancer fragments at a particular position are generated from a Poisson(ƛ) distribution, where ƛ is the true rate of non-cancer fragments at that position. Performed separately for each chromosome, all non-cancer counts were used to fit a gamma distribution for ƛ, which served as the prior for all positions on the chromosome. This gamma distribution was parameterized by alpha (shape) and beta (rate). The fit alpha and beta determined ƛ's mean and variance of endpoint counts across different positions on the chromosome, because the gamma distribution's mean is alpha/beta and variance is alpha/(beta^2). For each chromosome position, the position's observed non-cancer endpoints count (x_{NC}) was used along with the prior to fit a position-specific posterior: $P_{posterior}(ƛ) \sim$ gamma (posterior-alpha, posterior-beta), where posterior-alpha = alpha + x_{NC} and posterior-beta = beta + 1. Given the model that new observations are drawn from a Poisson(ƛ) distribution where ƛ is gamma-distributed according to $P_{posterior}(ƛ)$, the probability of the new observation follows a negative binomial distribution, with mean = (posterior-alpha)/(posterior-beta) and variance = [posterior-alpha * (posterior-beta + 1)/posterior-beta^2]. The p value of the cancer endpoints count at the position (x_C) was computed as the probability of observing that number or greater cancer counts under this negative binomial model: $P(x \geq x_C)$. Finally, the gamma's posterior-alpha and posterior-beta parameters were scaled before being used in the negative binomial model to account for the global difference in fragment read depth between the total cancer counts and total non-cancer counts over the entire chromosome; posterior-beta was divided by the ratio of the total cancer endpoint counts on the chromosome to the total non-cancer endpoint counts on the chromosome to achieve the scaled gamma distribution.

WGS: Fragment lengths classifier - Brief description

For the fragment lengths classifier, fragment length data was split into 100 kb genome bins, with each bin summarized as the geometric mean of the per-fragment length likelihood ratio of cancer signal fragments to non-cancer signal fragments. These ratios were normalized for technical variation found in non-cancer technical control samples as well as corrected for GC bias. Normalized likelihood ratios were processed using logistic regression on principal components of the bins to classify cancer signal versus non-cancer signals.

WGS: Fragment lengths classifier - Detailed description

For each cfDNA participant in a training set of cancer participants with small variant based cTF estimates greater than 5%, the fragment length distribution was computed per participant. Similarly, a fragment length distribution was computed per participant for a set of non-cancer participants. For each cancer participant, the cancer fragment length distribution was inferred using the following formula:

$$P(\text{length}|\text{sample}) = \text{cTF} * P(\text{length}|\text{cancer}) + (1 - \text{cTF}) * P(\text{length}|\text{noncancer})$$

The inferred cancer fragment length distribution was summarized across all the cancer participants using the median density per fragment length and renormalizing to ensure the density across all fragment lengths summed to 1. The likelihood ratio of the cancer fragment length distribution to the non-cancer fragment length distribution was used as a fragment level feature. Fragment level features were then summarized per 100 kb genomic bin using the geometric mean. These per-region features were normalized to an independent set of non-cancer reference samples and corrected for GC bias. GC bias correction was performed using spline regression of per-bin GC content as well as by regressing out principal components derived from technical controls. Classification was performed using principal component logistic regression with 5 principal components.

WGS: Allelic imbalance classifier - Brief description

For the allelic imbalance classifier, allelic imbalance data was first processed into 100 kb genomic bins. Each bin was summarized as the deviation of the observed from the expected probability of observing the allelic counts for each SNV given the phased haplotypes expected in that bin. Phased haplotypes were computed using Eagle2 using the 1000 Genomes reference set.[49,50] Bins were further robustly summarized into chromosome-arm level features using the median absolute value over bins with at least 50 phased single nucleotide polymorphisms. These final chromosome-arm level features were used to train a logistic regression to predict cancer signal versus non-cancer signal.

WGS: Allelic imbalance classifier - Detailed description

For each participant, WGS data was phased using Eagle2 with the 1000 Genomes reference set.[49,50] For each 100 kb bin, the following metric was computed:

$$(\text{haplotype 1 counts} - \text{expected haplotype 1 counts}) \,/\, \text{depth}$$

Expected haplotype 1 counts were computed after adjusting for sequencing/alignment-related bias and down-weighting fragment counts from single nucleotide polymorphisms (SNPs) found to reproducibly deviate from a binomial distribution centered near 0.5. For each chromosomal arm, a feature was computed as the median per-bin feature for bins with $\geq$50 heterozygous SNPs. Classification was then performed using logistic regression on arm-level features.

Pan-feature classifier - Brief description

Each of the classifiers described above produced a continuous score to predict cancer signals versus non-cancer signals using only the features analyzed in that classifier. The pan-feature classifier was trained using output scores from all of the cfDNA feature-based classifiers (excluding the clinical classifier) as inputs to a gradient-boosted decision-tree classifier (XGBoost).[51] Such a classifier allowed for capture of any nonlinear relations between scores that may allow improved prediction for some samples.

Pan-feature classifier - Detailed description

Individual feature models produced a continuous score predicting cancer versus non-cancer using only the features analyzed in that model. We combined these predicted scores as inputs to a machine-learning classifier to predict cancer versus non-cancer. To avoid assumptions of linearity and address the potential for interaction between scores, a gradient boosted decision tree classifier, XGBoost[51] was used to combine the scores. Hyperparameters were optimized by random search on the training data, followed by a final model trained on the whole training set. This final model was then used to predict cancer versus non-cancer on the validation set.

Clinical classifier - Brief description

The clinical data classifier used logistic regression incorporating age, smoking status, and family history of breast and ovarian cancers. This classifier was intended to provide a benchmark independent of cfDNA that reflected well-established risk factors for cancer.

Clinical classifier - Detailed description

The clinical baseline classifier was trained similarly to the other classifiers: the training set used 10-fold cross-validation to assess performance and a single model was trained on the full training dataset and locked prior to assessing on the validation set. The logistic regression classifier was implemented using R's glm with the true state of cancer or non-cancer regressed against age, smoking status, and family history of breast and ovarian cancers:

cancer_status $\sim$ age + heavy_smoker + current_smoker + family_history

Age was binned into three categories: <50, 50–64 and 65+. Heavy_smoker and current_smoker were two boolean factors: the first indicated if the participant was a heavy smoker (11+ pack-years) and the second indicated if the participant was a current smoker. Family history of breast and ovarian cancer was included as a risk score (0–6) determined as the number of affirmative responses to the following questions:

- Had positive test for hereditary breast syndrome (participant or immediate family member)
- $\geq$1 relative had breast cancer before age 50 years
- $\geq$2 immediate family members diagnosed at any age

- ≥1 relative had cancer in both breasts before age 50 years
- ≥1 male relative had breast cancer
- Immediate family member had ovarian cancer

CSO prediction - Brief description

Informed by the cancer signal detection classification performance, one feature type was chosen from each assay method (WGBS-based WG methylation, WGS-based SCNA, and TS-based SNV-WBC) to create classifiers predicting a likely anatomic location or histological characteristic of the cancer signal in individuals where a positive cancer signal was predicted (CSO). Note that WBC correction was not used for SCNA for CSO prediction because most of the interference from WBCs generated false positives, which would not have an actual cancer origin, and the SCNA-WBC classifier did not show dramatically improved performance over the SCNA classifier. The SNV classifier, however, was dramatically improved when accounting for WBC interference in the SNV-WBC classifier, so WBC correction was used for CSO prediction using SNVs. Because a CSO prediction is only returned in the case where a cancer signal occurs, a common set of cancer samples for evaluation of CSO predictions was generated using only the samples where all three representative cancer signal detection classifiers detected a cancer signal at 98% specificity (see below). For completeness, each CSO classifier was also run on the full set of cancer samples detected in the complete analyzable validation set by its corresponding cancer signal detection classifier (Figure S5).

The CSO classifiers extended the principles for cancer signal detection to predict multiple CSO labels. For the WGBS-based WG methylation CSO classifier, locations where extreme fragments were indicative not merely of cancer presence but of particular types of cancer origins were identified, and the presence or absence of fragments at those locations were used as inputs to a penalized multinomial logistic regression. Each regression was trained to predict a fixed set of CSO labels. The TS-based SNV-WBC CSO classifier substituted an elastic net multinomial regression for the previous binomial regression and used the same feature set of potentially disrupted genes as the cancer signal detection classifier. For the WGS-based SCNA CSO classifier, read depths for bins across the genome were projected onto principal components for SCNA that were defined for 21 cancers in The Cancer Genome Atlas (TCGA). These principal components were then used in a multinomial logistic regression to predict CSO.

CSO prediction - Detailed descriptions

Post hoc determination of CSO was accomplished for each of the sequencing approaches as described below. Reported results reflect the predictions of the described models on the validation set. In particular, CSO performance was reported on detectable cancers (solid and hematologic) as defined by the set of cancer samples that were correctly detected at 98% specificity by all three of the cancer versus non-cancer signal detection classifiers (WG methylation, SNV, and SCNA-WBC). The following set of labels was defined for CSO classification: breast; cervix; colon/rectum; esophagus; head/neck; liver/bile duct/gallbladder; lung; lymphoma; plasma cell neoplasm; other; ovary; pancreas; kidney. The label "other" included anus, cancer of unknown primary, melanoma, stomach, thyroid, and uterus cancers, and was counted as a correct result in the calculation of accuracy and precision for the CSO classifiers.

WGBS: WG methylation CSO prediction

To classify CSO from WGBS data, the genome was first divided into 1 kb regions. Fragments derived from the training set samples were grouped by their training label (non-cancer or a particular cancer type) and used to train a probabilistic model of fragment-level methylation for each label in each region. Each of these models took the form of a three-component Bernoulli mixture model, where each of the three components was parameterized by a set of methylation probabilities, one for each CpG in the region (each with a value between 0 and 1), and a non-negative mixture fraction, the values of which were constrained to sum to 1 across the three components. Model parameters were learned by iterative gradient-based optimization using the rprop algorithm.[52] These models were used to define a set of boolean-valued features, one for each cancer type C, region R, and threshold value T from the set {1, 2, 3, 4, 5, 6, 7, 8, 9}. Each such feature assigned a numerical value of 1 to any sample that included at least one fragment in region R whose log-likelihood assigned by the Bernoulli mixture model for cancer type C was at least T greater than the log-likelihood assigned to that fragment by the non-cancer Bernoulli mixture model. All other samples (ie, those lacking any such fragment in region R) were assigned the value 0 for that feature. Each feature was assessed for its ability to distinguish between training set samples labeled as the feature's positive cancer type and each other (negative) cancer type, pairwise, quantified by the mutual information between the boolean feature values and the pair of cancer labels. For each label pair (ie, one positive cancer type and a distinct negative cancer type), the 256 features with the highest mutual information for that label pair were selected for inclusion in the final feature set, with the additional constraint that each of the 256 must derive from a distinct genomic region. The full set of features was then constructed from the union of the pairwise lists. The final set of feature values was used to train a multinomial logistic regression classifier with ridge penalty, using a nested cross-validation strategy to determine an appropriate ridge penalty coefficient value.

TS: SNV CSO prediction

To predict CSO from TS data, the same technique employed for cancer or non-cancer classification, using multinomial regression in glmnet,[46] was employed. In this case, coefficients were allowed to be either positive or negative, as a mutation in a particular gene region may increase or decrease the likelihood of a sample being a specific cancer type. This was only expected to be accurate in those samples where both the cancer/non-cancer classification score was sufficiently high, and the mutations were characteristic of a particular cancer; the classification coefficients reflected this reality. In addition to using gene somatic mutations as features, the

CSO classifier was trained as a stacking-ensemble classifier. A classifier using copy number alterations (CNAs) estimated from targeted panel regions was trained in a similar fashion as the SNV classifier, and then the probability scores from both SNV and CNA classifiers were used as the input features to train a final multinomial classifier in glmnet.[46] The scores generated by the final stacking ensemble classifier were used for CSO prediction.

WGS: SCNA-WBC CSO prediction

To predict CSO from WGS data, copy number filters were pre-trained on TCGA.[53] This involved converting TCGA array-based copy number counts to a compatible format to CCGA copy number counts based on WGS. This required interpolation from a less dense to a more dense representation of the genome, which was not expected to be problematic because copy number aberrations in cancer are often much longer than the spacing between array-based markers (eg, chromosome arm scale).

Once copy number data were prepared as described above, we then set out to learn copy number filters that would be useful for differentiating among cancer types. PCA reduced the dimensionality of the classification problem by producing features that, on inspection, corresponded to weightings of bins highlighting particular amplification and deletion events.

To increase the likelihood that PCA would identify copy number aberrations that differ among cancer types, the top 50 principal components (PCs) were kept for each pair of the 21 TCGA cancers matching the first CCGA substudy cancer types, and additionally for each of these 21 cancers against non-cancer, resulting in a total of 231 sets of 50 PCs each.

Despite being fit independently on each pair of diagnostic classes, PCs were by design generated without explicit knowledge of the cancer types in each pair. To address this limitation, we determined which of the PCs were informative for cancer signal of origin classification. To reduce the set of 11,550 PCs, within TCGA data, L1 regularized logistic regression was performed to classify each of the 231 pairs, and only PCs with non-zero coefficients were retained. This resulted in an average of 6.5 PCs per pair, reducing the total to 1,502 PCs retained. These PCs were interpreted as the copy number filters that best differentiated diagnostic classes.

For CSO prediction, features for each sample were computed by projecting onto the 1,502 TCGA vectors described above. An additional 20 cfDNA-specific vectors were generated by PCA on the cfDNA training data. Lastly, these features were catenated together and passed into an elastic net-regularized logistic regression. This model was fit on the training set, then applied to the validation set to produce CSO probabilities. Validation set data were preprocessed identically to those in the training set, with no knowledge of validation set labels.

Overview of circulating tumor fraction and clinical LOD

A key exploratory objective of this study was to examine the relationship between classification performance and cfDNA tumor fraction. cTAF is defined as the expected fraction of tumor-specific mutant alleles present in a cfDNA sample across a set of mutated loci. cTAF was computed under a statistical model that integrates over the somatic variants (tumor mutations) identified in a participant's tumor biopsy sample. Tumor mutation candidates were identified from WGS of pre-treatment tumor tissue biopsies and matched WBCs to remove germline polymorphisms from the tumor biopsy somatic mutation candidates. Consequently, cTAFs are only available for cancer cases with tumor biopsy and WBC WGS. For each participant's tumor-specific mutation candidates, the observed frequency of mutant alleles in cfDNA TS reads was used to calculate the fraction of tumor genomes in a cfDNA sample denoted circulating tumor fraction (cTF, see calculation of cTF and cTAF below). The estimated cTF was then multiplied by the median biopsy mutant allele frequency to estimate cTAF (see calculation of cTF and cTAF below). For some samples a cTAF could not be estimated because no tumor-informative variants overlapped with the regions targeted by TS in cfDNA. Samples with TS-overlapping mutations that did not have any mutant allele support in TS were not included in the cTAF analysis. cTAF allowed the assessment of cfDNA shedding behavior across cancer types and the estimation of a clinical LOD for each cancer signal detection classifier. To determine the ability to detect cancer at 98% specificity, the probability of detecting a cancer case was estimated on classifier predictions in cases versus $\log_{10}$(cTAF) under logistic regression, and 95% confidence intervals (CIs) for the clinical LOD were computed using a Gaussian approximation of the slope standard error in logistic regression. Clinical LOD was defined for each classifier as the cTAF where the probability of detecting a cancer signal was 50%. Clinical LOD should not be confused with analytical LOD measurements that typically refer to the limit of analyte detection using known replicates over a dilution series.

Targeted methylation clinical LOD from second CCGA substudy

For comparison, a clinical LOD was also included for the targeted methylation method from the second CCGA substudy. Methods for the second CCGA substudy have been described previously.[19] In order to properly compare with the approaches evaluated here, clinical LOD was calculated for the targeted methylation method at a 98% specificity level, though in the second CCGA substudy the observed specificity for the targeted methylation approach was 99.3%.[19]

Calculation of cTF and cTAF

For each sample, somatic tumor variants were identified from available tumor-tissue biopsy and matched WBC sequencing. For those variants, cTAF was estimated based on the fraction of mutant containing reads contributed from the tumor to the cfDNA sample using the targeted small variants panel (TS). Once computed across participants, cTAF estimation allowed assessment of shedding behavior across cancers and identification of the clinical LOD at the 98% specificity target.

Within each participant, we computed a Bayesian likelihood of observing the cfDNA alternate allele counts corresponding to the alleles found in the tumor for a given value of the fraction of tumor genomes in cfDNA. We refer to the fraction of tumor genomes in cfDNA as cTF, which is the target of the inference. Mutant allele counts were approximated as a Poisson distribution where the rate

was controlled by the local depth of sequencing, cfDNA allele frequency, and tumor biopsy allele frequency. Individual variants were treated as independent for purposes of aggregating the total likelihood. The posterior density was computed using a uniform prior. We define the estimated cTF to be the value at the median of the posterior density. Approximate credible intervals were obtained based on the same density. cTAF was then computed by multiplying the estimated cTF by the median tumor biopsy mutation allele frequency per sample to obtain an expected tumor mutant allele frequency in cfDNA. This provided a simple, robust method for estimating cTAF across samples with multiple orders of magnitude difference in value.

Analysis of CH - Brief description

A *post hoc* analysis of a subset of the CCGA participants in the training set with TS data of paired cfDNA and WBC genomic DNA was performed in order to evaluate the necessity of sequencing individual WBCs in order to remove interference from CH variants. WBC-matched somatic variants (SNVs, indels) were compared with cfDNA variants in participants with and without cancer to identify CH variants.

Analysis of CH - Detailed description

Deep TS was performed on both cfDNA and WBC using the TS panel. A noise model was trained allowing cfDNA candidate variants to be quality scored. After filtering candidate variants for high quality (Phred-scaled quality score 60), as well as for technical artifacts occurring at the ends of reads, a joint calling procedure was performed to establish whether the frequency in cfDNA was incompatible with a hematopoietic origin (99% likely that cfDNA frequency was larger than the WBC frequency, conditional on observed allele counts). Candidates identified as incompatible were removed from WBC-matched analysis.

The remaining variants with frequencies compatible with a WBC origin were further filtered to establish high-quality variants likely to be of hematopoietic origin. Candidate variants likely to be germline (either common or private) were eliminated by requiring observed WBC frequency to be <30%. Candidate variants recurring within technical controls in >5% of the samples were removed as likely artifacts. Variants occurring in repetitive sequence ($\geq$8 repeats) were suppressed, as were candidates in the HLA-A region. At least one alternate allele was also required to be observed in the WBC sequencing results. Finally, biologically derived variants exhibited a range of allele frequencies, while remaining mismapping artifacts either had no occurrence within a sample or had a very tight distribution of frequencies, and variants were removed when the standard deviation of the log(allele frequency) for samples with positive frequency was <0.02; this occurred in $\geq$6 samples. Variants were required to pass all of these filters.

This analysis was performed on all individuals where both targeted panel cfDNA and WBC sequencing passed quality control, including individuals with non-informative clinical staging and those for whom one or more of the other assays failed. Analysis of clonal hematopoietic effects was restricted to individuals with solid tumors. This resulted in 1,438 individuals in the training set where clonal hematopoietic effects could be analyzed independently of cancer type and status.

Site balance analysis

A *post hoc* analysis was performed to investigate the ability of the methylation assay and classifier to perform equivalently on samples from sites not represented in training. The training set was divided into five folds approximately balanced for participants with cancer and without cancer, with the added condition that all samples recruited from a given site were within the same fold. Five corresponding classifiers were trained, each one omitting a fold and hence removing a set of sites from the training for that classifier. Each classifier was then applied to the validation set, generating five replicates, each scored by a classifier trained by leaving out a set of sites. A 98% specificity cutoff was generated for each classifier using only validation set samples where the site was present in the training. This cutoff for each classifier was then applied to determine a cancer versus non-cancer call in all samples scored by that classifier.

A bootstrap analysis resampling by participant was used to account for the fact that the same samples were repeatedly scored. This allowed an estimate of the differential effect due to a site being present in training, conditional on the fact that some samples were easier than others to detect. The validation set contained 847 samples (485 cancer, 362 non-cancer); average sensitivity was 33.4% (range 32.2%–35.1%) across the five classifiers, comparable to the original classifier performance. As each classifier was run on the same set of validation samples, the differential effect of training on the site where a sample was recruited was compared to omitting that site from training. Eighty-four samples (55 cancer, 29 non-cancer) were from sites not used in training, and were therefore not used for differential analysis. The average sensitivity on those samples was 41.8%. The remaining samples were used in a bootstrap analysis to examine if any difference in performance could be attributed to training on sites.

The bootstrap analysis showed that the median sensitivity in the remaining samples was 32.4% over bootstrap replicates, with the median difference in sensitivity being 0% for site included in training versus site not included in training (95% CI: $-$1.2%–1.1%, p = 0.478 for superior performance with site included). Specificity was examined in training with and without sites included. Retaining the originally set specificity thresholds using scores from samples with site included in training, the median false-positive rate for samples with site omitted from training was 1.5% (95% CI: 0.3%–3%, p = 0.745 to reject a false-positive rate $\leq$2% [98% specificity]).

Thus, in this *post hoc* analysis there was no statistically significant effect on sensitivity to omitting sites from training. Specificity was within the expected range of variation for a small sample set at a targeted 98% specificity threshold.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

We analyzed clinically relevant outcomes for statistically significant differences in performance, assessing cancer signal detection with associated clinical LOD and CSO prediction separately. We analyzed sensitivity at a chosen specificity for pairs of classifiers and CSO accuracy for pairs of classifiers using McNemar's two-sided test. A paired test was appropriate for this data because we compared predictions that were generated from the same samples. Finally, for clinical LOD, we compared the overlap of 95% CIs for significance. CIs shown in figures were binomial estimates of the 95% CIs computed using the standard Clopper-Pearson method, except for the CIs for clinical LOD (described above in the section on clinical LOD).

For the validation set, analyses of WG methylation, SNV, SNV-WBC, SCNA, SCNA-WBC, and clinical data cancer signal detection classifiers were double-blinded: none of the personnel involved in performing sequencing or quality control, or classification analyses were aware of clinical information (eg, cancer/non-cancer status) until assay data lock, and none of the team involved in collecting the clinical information were aware of the sequencing, quality control, or classification results until clinical data lock. In particular, for each of the blinded datasets, access to the mapping from sample identification in the laboratory to the clinical data patient identifications was restricted. Additionally, after the validation dataset clinical data lock but before the validation dataset was released for use, a data integrity team reviewed the merged data to ensure completeness across the training and validation datasets; the data integrity team was not involved in classifier development or in determining clinical or assay evaluability. The fragment endpoint, fragment length, allelic imbalance, and pan-feature cancer signal detection classifiers, as well as the CSO classifiers were developed after blinding was lifted.

Performance comparison

The features and classifiers were compared in a number of ways. In order to ensure a method with the best performance was identified at a high specificity to reduce false positives, sensitivity was calculated at a 98% specificity level. A 98% specificity cutoff was determined *post hoc* for the training and validation sets. Clinical LOD was assessed as described above, and CSO prediction accuracy was also assessed for one feature from each of the three sequencing assays.

Additional performance measures

Partial area under the ROC curve (pAUC) values were used to assess cancer signal detection classifier ROC curves. Because population-scale screening is thought to require high specificity, classifiers were compared in the relevant region (98%–100% specificity). A bootstrap technique was used to evaluate differences in pAUC between cancer signal detection classifiers. pAUC was calculated using a pROC R package.

ADDITIONAL RESOURCES

Clinical trial registry number: NCT02889978.

Clinical trial URL: https://clinicaltrials.gov/ct2/show/NCT02889978.

Cancer Cell, Volume *40*

Supplemental information

Evaluation of cell-free DNA approaches for multi-cancer early detection

Arash Jamshidi, Minetta C. Liu, Eric A. Klein, Oliver Venn, Earl Hubbell, John F. Beausang, Samuel Gross, Collin Melton, Alexander P. Fields, Qinwen Liu, Nan Zhang, Eric T. Fung, Kathryn N. Kurtzman, Hamed Amini, Craig Betts, Daniel Civello, Peter Freese, Robert Calef, Konstantin Davydov, Saniya Fayzullina, Chenlu Hou, Roger Jiang, Byoungsok Jung, Susan Tang, Vasiliki Demas, Joshua Newman, Onur Sakarya, Eric Scott, Archana Shenoy, Seyedmehdi Shojaee, Kristan K. Steffen, Virgil Nicula, Tom C. Chien, Siddhartha Bagaria, Nathan Hunkapiller, Mohini Desai, Zhao Dong, Donald A. Richards, Timothy J. Yeatman, Allen L. Cohn, David D. Thiel, Donald A. Berry, Mohan K. Tummala, Kristi McIntyre, Mikkael A. Sekeres, Alan Bryce, Alexander M. Aravanis, Michael V. Seiden, and Charles Swanton

Supplement

Supplementary Tables

Table S1. **Multivariate logistic regression analysis of cTAF, stage, and cancer type with relation to cancer signal detection, related to Figure 5**. Four logistic regression models were fit to different combinations of cancer type, stage, cTAF versus detection by the WG methylation classifier as indicated using stage I of the remaining cancers (not breast, lung, or colon/rectum cancer) as the reference. *p* values represent the significance of each predictor in the model. AIC is used to compare model complexity confirming that cTAF is a good predictor independent of stage. AIC, Akaike's Information Criterion; AIC_{min}, minimum AIC; cTAF, circulating tumor allele fraction. [a]$p<0.1$; [b]$p<0.05$; [c]$p<0.001$. A dash indicates that the variable was not included in the multivariate logistic regression model.

	Regression models incorporating various predictors			
Predictors	**Cancer type**	**Cancer type and stage**	**Cancer type and cTAF**	**Cancer type, stage, and cTAF**
Intercept, *p value*	0.3	$3.9x10^{-8}$[c]	$4.1x10^{-18}$[c]	$5.4x10^{-12}$[c]
Cancer type: breast, *p* value	0.019[b]	0.9	0.53	0.32
Cancer type: lung, *p* value	0.038[b]	0.03[b]	0.62	0.48
Cancer type: colon/rectum, *p* value	0.056[a]	0.046[b]	0.58	0.44
Stage II, *p* value	-	$1.8x10^{-4}$[c]	-	0.48
Stage III, *p* value	-	$2.3x10^{-11}$[c]	-	0.32
Stage IV, *p* value	-	$2.4x10^{-14}$[c]	-	0.32
log_{10}(cTAF), *p* value	-	-	$1.9x10^{-19}$[c]	$4.8x10^{-17}$[c]
AIC	545.9	448.7	201.2	204.6
$AIC-AIC_{min}$	344.7	247.5	0	3.4

Table S2. Area Under the ROC Curve for Each Classifier at High Specificities, related to Figure 2. To account for a range of high specificities, the validation set pAUC for cancer signal detection ROC curves (Figure 2) was calculated for a specificity range 0.98-1 (0%-2% false-positive rate) and is shown here. This range was chosen because high specificity is critical for a population-scale screening test. For the WG methylation classifier, $pAUC_{0.98\text{-}1}$ was greater than those classifiers derived from whole-genome or TS alone (ie, without the addition of WBC sequencing to remove noise). SNV and SCNA methods both needed additional WBC assay data to account for background noise in order to match the pAUC performance of WG methylation. CI, confidence interval; $pAUC_{0.98\text{-}1}$, partial area under the ROC curve from 0.98 - 1 specificity; SCNA, somatic copy number alterations; SCNA-WBC, somatic copy number alterations with WBC noise removal; SNV, single nucleotide variants; SNV-WBC, single nucleotide variants with WBC noise removal; WG, whole-genome.

Classifier	**pAUC0.98-1**	**95% CIs (low, high)**
Pan-feature	0.66	0.64, 0.68
SNV-WBC	0.64	0.62, 0.67
SCNA-WBC	0.64	0.62, 0.66
WG methylation	0.63	0.61, 0.66
Fragment lengths	0.62	0.60, 0.65
SCNA	0.60	0.56, 0.64
Allelic imbalance	0.59	0.57, 0.61
Fragment endpoints	0.56	0.55, 0.59
SNV	0.56	0.54, 0.58
Clinical data	0.50	0.50, 0.51

Supplementary Figures

Figure S1. Sensitivity of Cancer Signal Detection Classifiers by Cancer Stage, related to Table 3 and Figure 2. For all cancer samples with associated clinical staging information, the sensitivity at 98% specificity is shown for each cancer signal detection classifier. Sensitivity increased with clinical cancer stage. Orange indicates the performance of the WG methylation classifier. Data is plotted as mean ± 95% confidence intervals. SCNA, somatic copy number alterations; SCNA-WBC, somatic copy number alterations with WBC noise removal; SNV, single nucleotide variants; SNV-WBC, single nucleotide variants with WBC noise removal; WG, whole-genome.

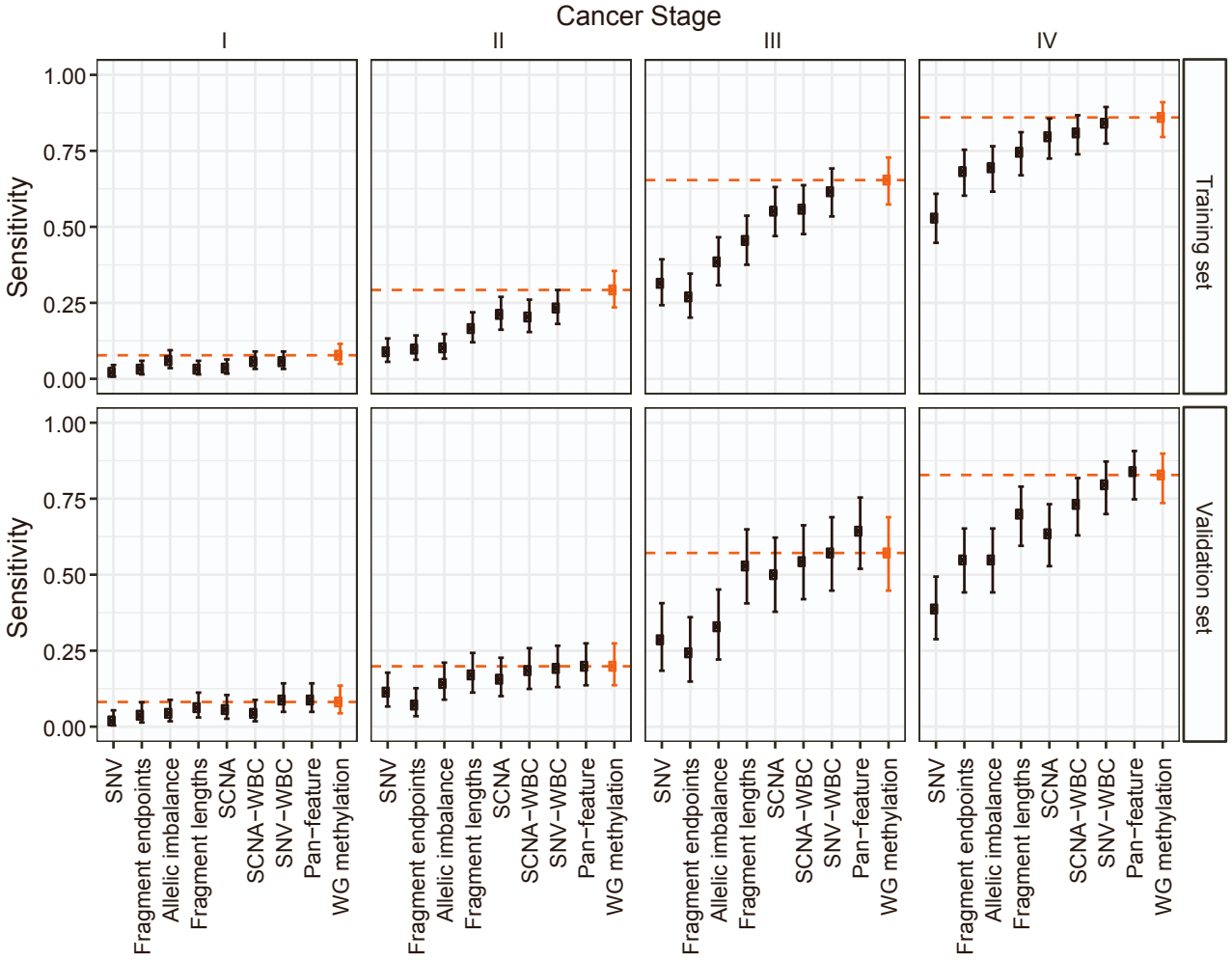


Figure S2. **Clinical Limit of Detection for Each Cancer Signal Detection Classifier: Training and Validation Sets, related to Figure 3.** This plot is similar to Figure 3, but additionally shows clinical LODs at 98% specificity for the training set for the classifiers. The blue data point indicates the clinical LOD at 98% specificity for the improved targeted methylation assay that was used in the second CCGA substudy (n = 559 participants in the second CCGA substudy validation set with available tumor tissue).[1] Error bars indicate 95% confidence intervals. A clinical LOD was only determined for the validation set for the pan-feature classifier because this classifier was a combination of the other cfDNA feature classifiers that had already been trained separately in the first CCGA substudy. LOD, limit of detection; SCNA, somatic copy number alterations; SCNA-WBC, somatic copy number alterations with WBC noise removal; SNV, single nucleotide variants; SNV-WBC, single nucleotide variants with WBC noise removal; WG, whole-genome.

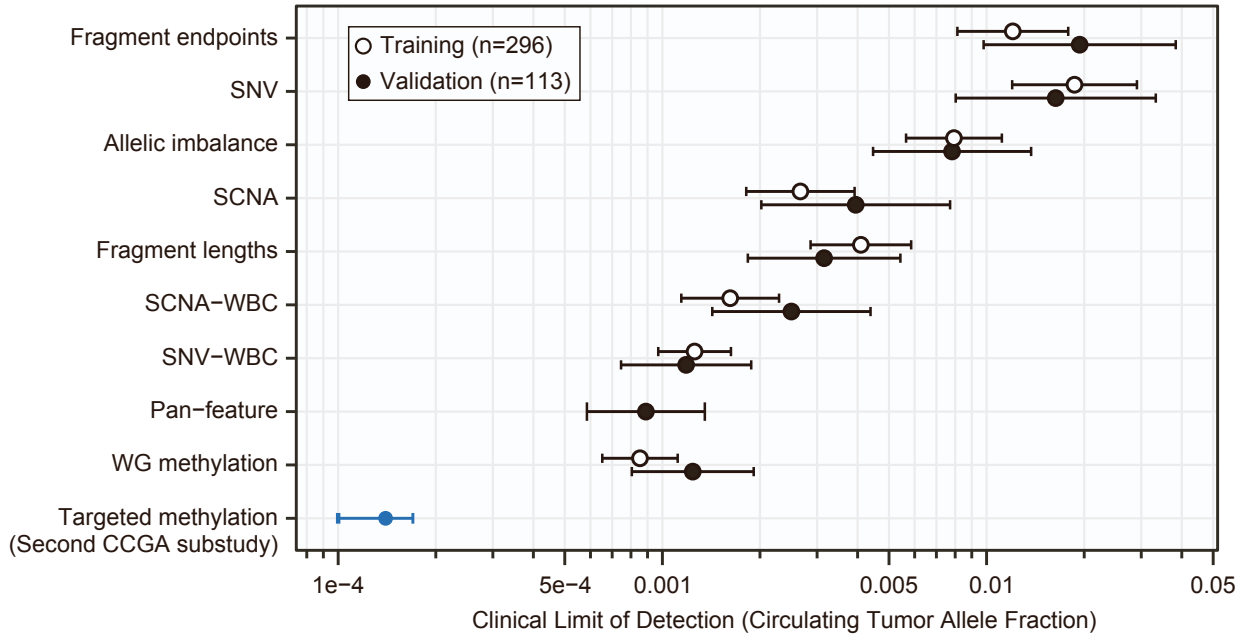


Figure S3. Complementarity of Information Across Classifiers, related to Figure 2, Figure 3, Table 3, and Figure S4. This figure relates to all visual items in the main text that contain results from the pan-feature classifier, which combined all cfDNA cancer signal detection classifiers to assess whether there was added value from combining assay data. Here, all cases that were classified as cancer by classifiers representing each assay were compared across classifiers to determine the overlap of information that each assay provided. True positive results are included in the top row and false positive cancer signal detections are included in the bottom row. The three classifiers that were compared are listed in the three rows at the bottom of each graph and combinations noted along the x-axis. Black dots and connectors along the x- axis indicate which combination of classifiers or single classifiers are shown to have detected the samples indicated by each bar (eg, the first and largest bar in the true positive graphs indicates that most true positive cancer cases were commonly detected by all three classifiers). SCNA-WBC, somatic copy number alterations with paired white blood cell sequencing to remove background noise. SNV-WBC, somatic nucleotide variants with paired white blood cell sequencing to remove background noise. WG, whole-genome.

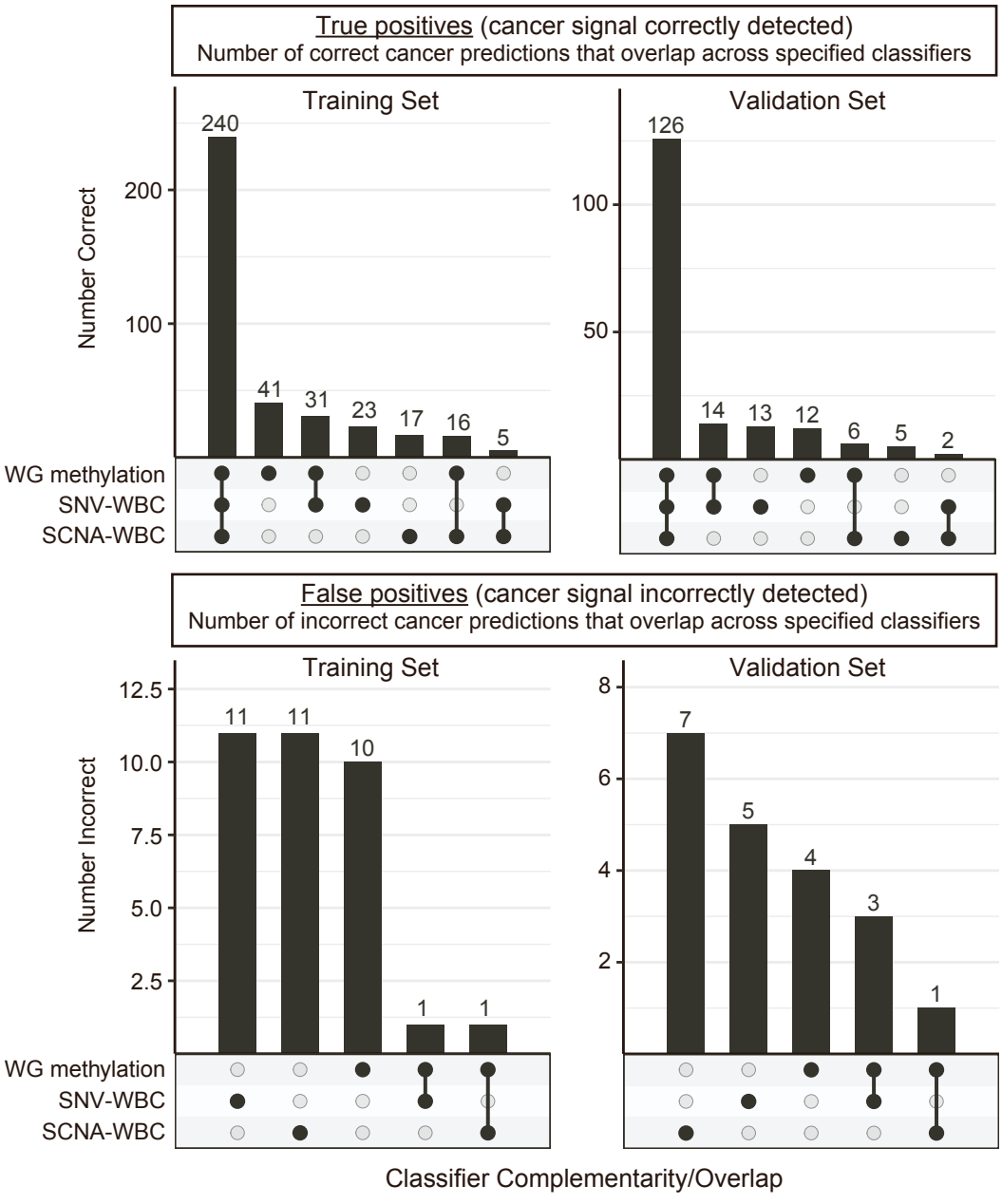


Figure S4. Visualization of the Cancer Signal Detection Feature Space using Dimensionality Reduction, related to Figure 2, Figure 3, Table 3, and Figure S3. To visualize whether the WG methylation cancer signal detection classifier was expected to be improved by combining with information from other classifiers, scores from all classifiers were reduced to a two-dimensional representation using Uniform Manifold Approximation and Projection (UMAP).[2,3] Projection of the data onto the top two dimensions of a UMAP projection of the training scores from all classifiers (all points) are shown here relative to the WG methylation classifier (blue points). Most detected samples (true and false positives) are in the left cluster and undetected samples are in the right cluster (true and false negatives) with both largely overlapping each other, which suggested only small to moderate gains were expected from the pan-feature classifier. This clustering is consistent with the result that sensitivity of the pan-feature classifier was not significantly different from the WG methylation classifier despite using input from nine classifiers across three distinct cfDNA assay technologies. UMAP, Uniform Manifold Approximation and Projection. O indicates non-cancer data. X indicates cancer data.

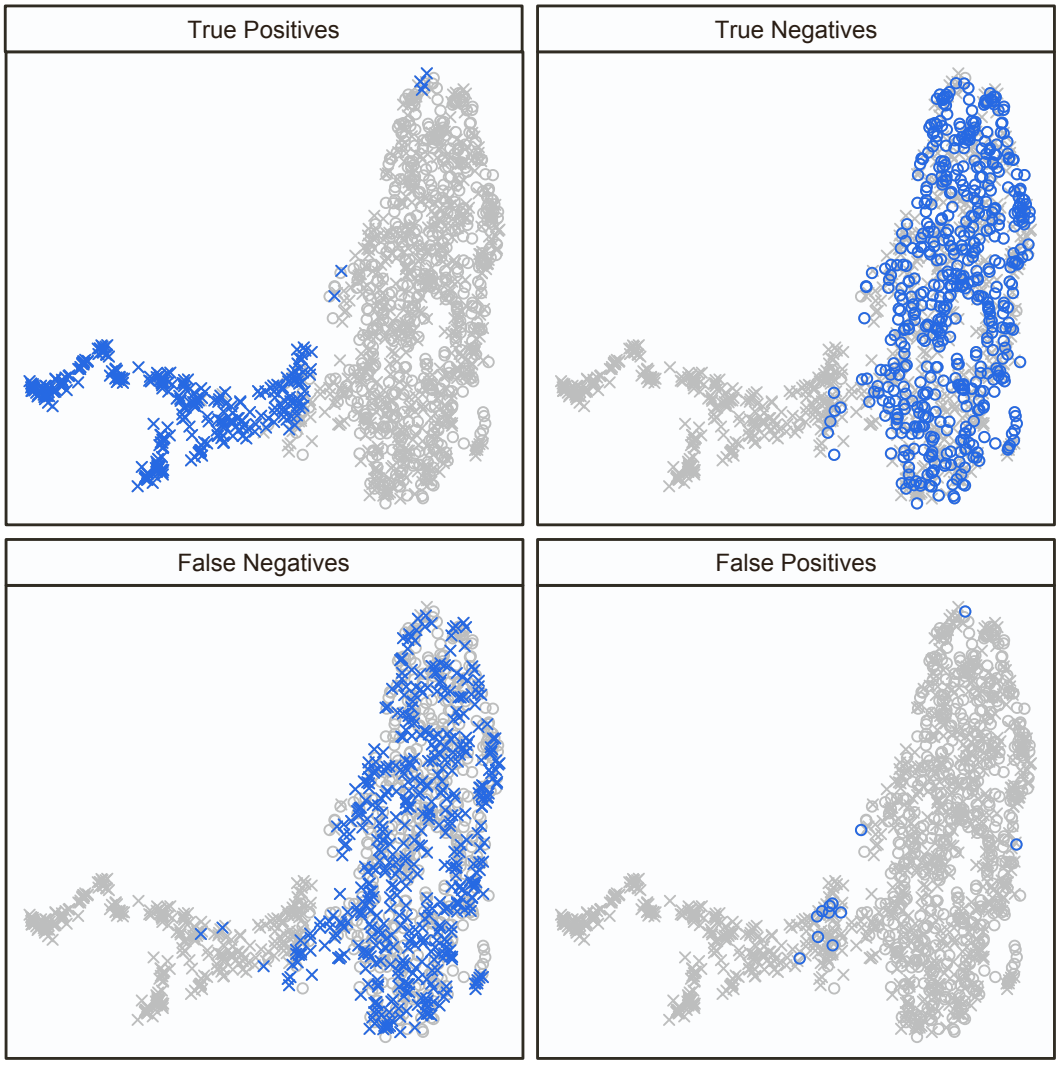


Figure S5. Accuracy of Cancer Signal Origin Prediction by Prototype Assay Using the Full Set of Validation Cancer Samples Detected by Each Particular Classifier, related to Figure 4. Confusion matrices representing the accuracy (along top of each matrix) and precision (along the right side of each matrix) of CSO prediction. Agreement between the true (x-axis) and predicted (y-axis) CSO per sample using experimental CSO classifiers from each of the three assays. Color corresponds to the proportion of predicted CSO (y-axis) which were correct (x-axis). Included participants are those predicted as having cancer at 98% specificity by each separate corresponding cancer signal detection classifier. Sample sizes are indicated in parentheses in the figure. CSO, cancer signal origin; L/B/G, liver/bile duct/gallbladder; PCN, plasma cell neoplasm; SCNA, somatic copy number alterations; SNV-WBC, single nucleotide variants with white blood cell noise removal; TS, targeted sequencing; WGBS, whole genome bisulfite sequencing; WG, whole-genome; WGS, whole-genome sequencing.

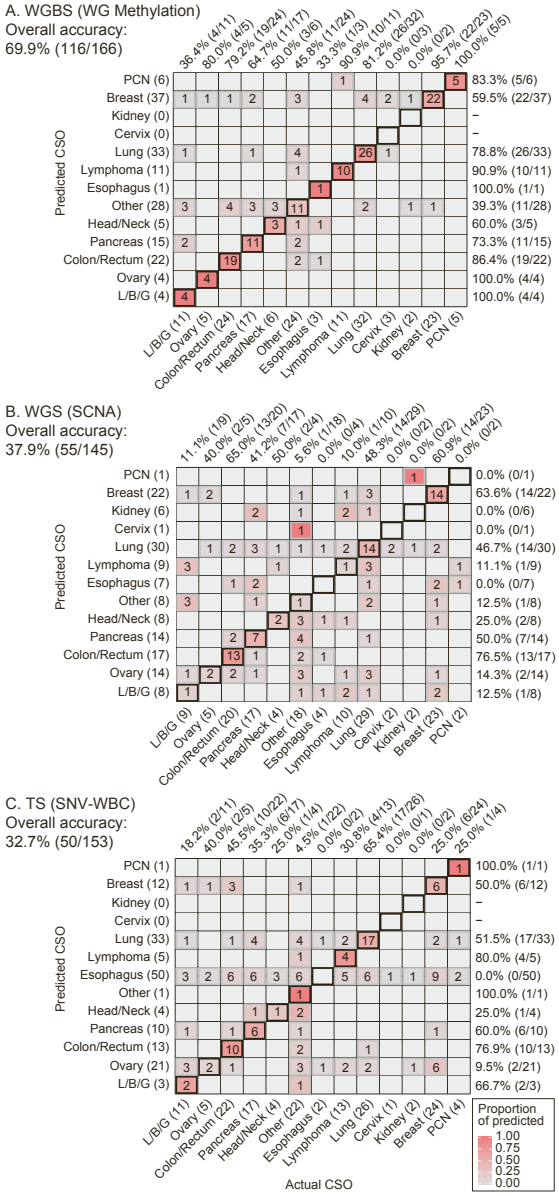


Figure S6. Distribution of cTAF by All Cancer Types and Stage, related to Figure 5. In the box plots, boxes represent 75% confidence intervals and whiskers represent 95% confidence intervals. Each dot represents the cTAF measurement from one participant. Text inserts above each box indicate participants with analyzable tumor tissue / total participants with cancer from the analyzable set of cancer cases. The total number of participants with analyzable tumor tissue for this analysis was n = 409. The total number of analyzable cfDNA samples from participants with cancer was n = 1,297.

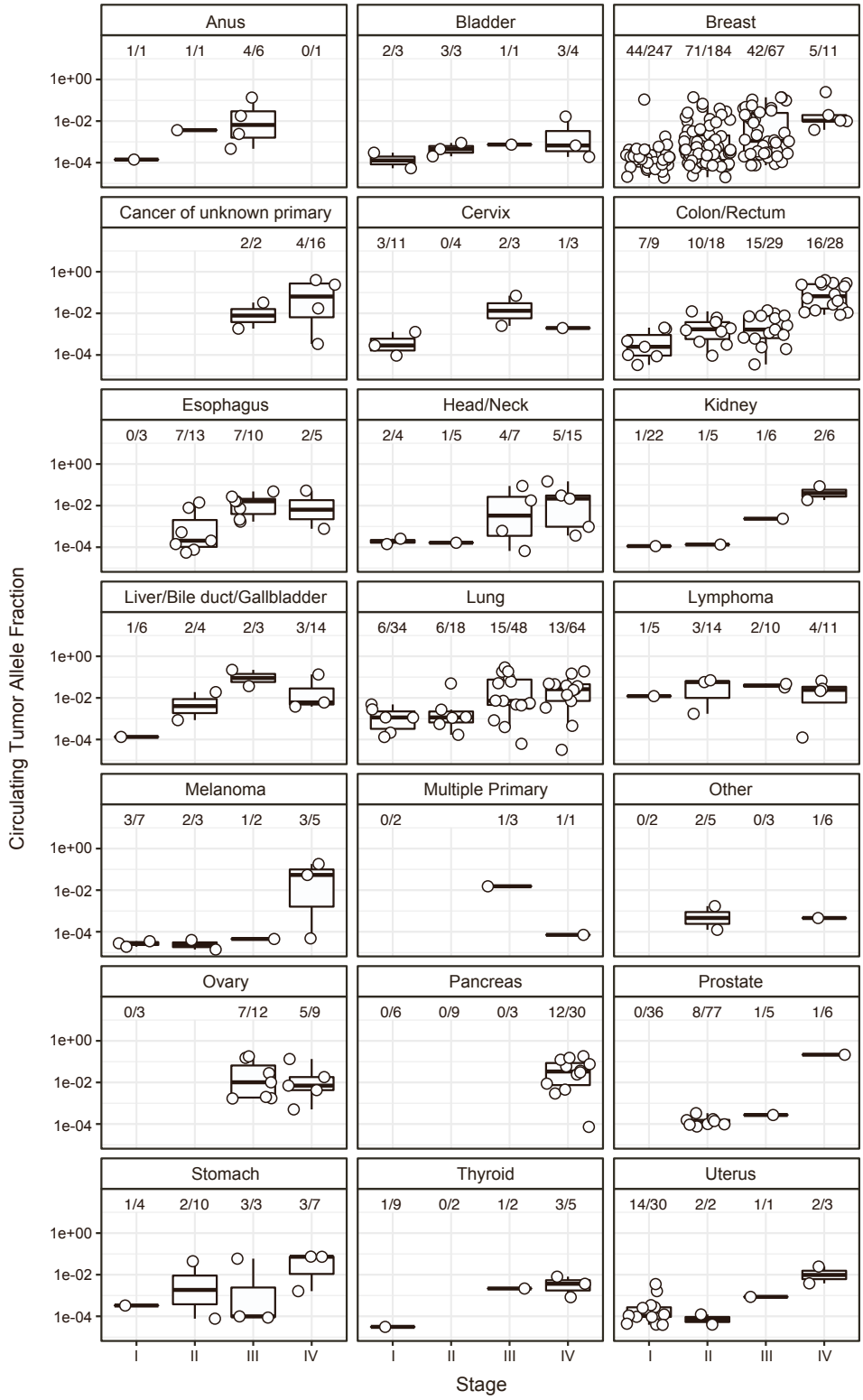


Figure S7. Whole-genome Methylation Cancer Signal Detection Classifier Signal Strength Varied with Circulating Tumor Allele Fraction, related to Figure 5. Cancer signal strength (ie, classifier detection score) on a scale of 0-1 was computed as part of the cancer signal detection classification and is plotted here versus the circulating tumor allele fraction (cTAF). Both have proportional units ranging from 0 to 1. The cTAF explained 72% of the variance in cancer signal strength. Blue indicates that the cancer sample was detected by the WG methylation classifier, and Black indicates that the cancer sample was not detected. The relationship between methylation score and cTAF was visualized (yellow line) using monotonic P-splines fit with a shape-constrained additive model in R (scam package).[4]
n = 113 participants with analyzable tumor tissue and detectable cTAF from the validation set.

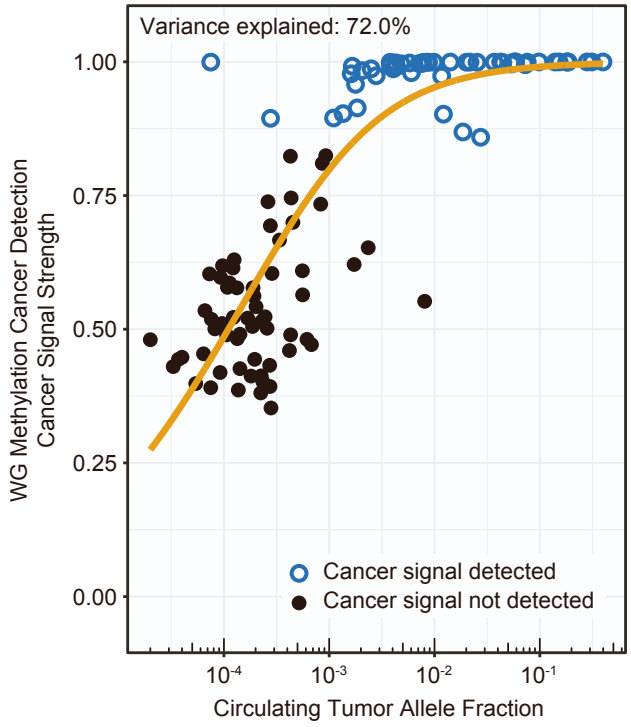
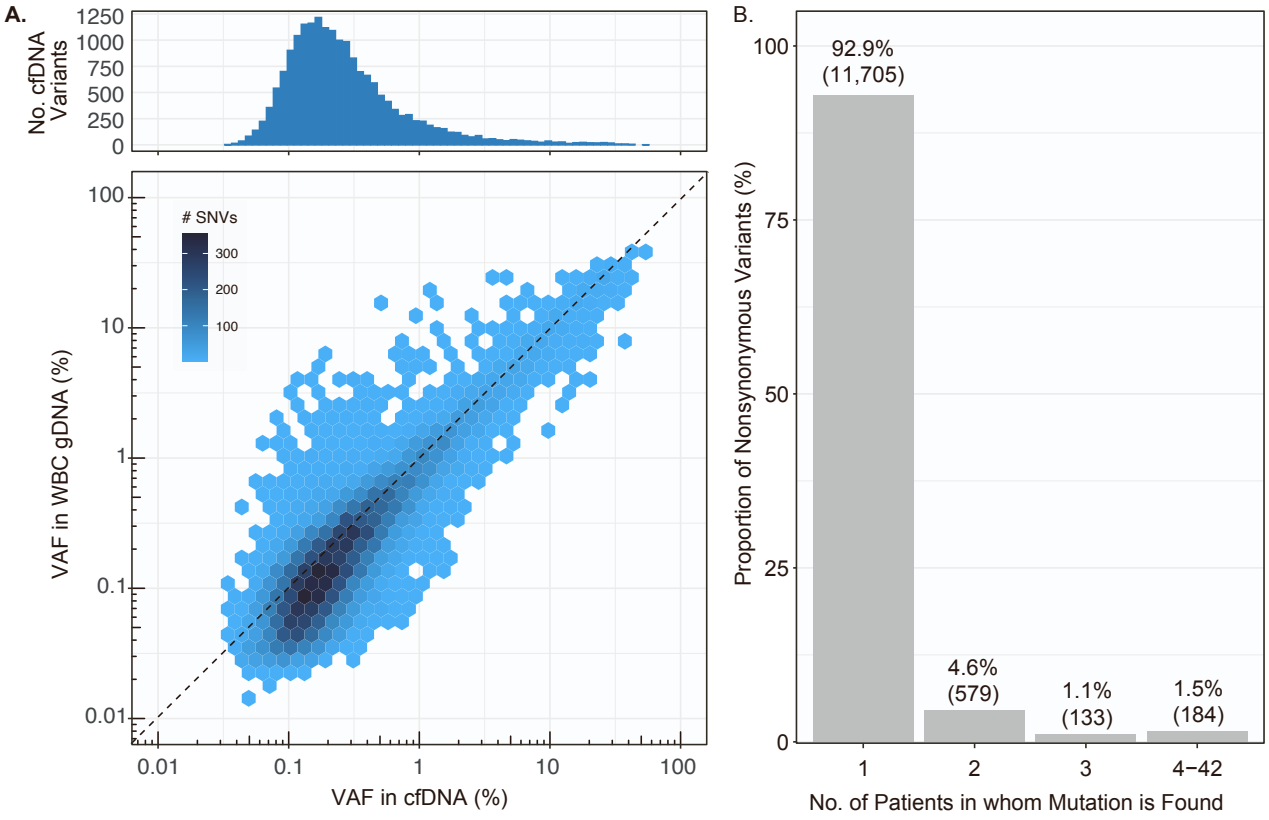


Figure S8. Clonal Hematopoiesis in Cell-Free DNA, related to Figure 2, Figure 3, Figure 4, and Table 3. This figure relates to all figures in the main text that include a classifier with paired WBC sequencing. (A) SNV allele fraction in cfDNA (x-axis, histogram) is plotted against corresponding variant allele fraction in WBC genomic DNA (y-axis). Color indicates the number of SNVs. 90% (16,571 of 18,400) of total SNVs in cfDNA were below 1.0% VAF. (B) 93% (11,705 of 12,601) of unique SNVs in cfDNA that were matching in corresponding participant WBC were specific to an individual participant. cfDNA, cell-free DNA. gDNA, genomic DNA. SNV, single nucleotide variant. VAF, variant allele fraction. WBC, white blood cell.



References

1. Liu M.C., Oxnard G.R., Klein E.A., Swanton C., Seiden M.V., CCGA Consortium. (2020). Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. Ann. Oncol 31,745−759. 10.1016/j.annonc.2020.02.011.
2. Clarke C.A., Hubbell E., Ofman J.J. (2021). Multi-cancer early detection: a new paradigm for reducing cancer-specific and all-cause mortality. Cancer Cell 39, 447−448. 10.1016/j.ccell.2021.02.004.
3. Allaoui M., Kherfi M.L., Cheriet A. (2020). Considerably improving clustering algorithms using UMAP dimensionality reduction technique: a comparative study. ICISP 5, 317–325. doi: 10.1007/978-3-030-51935-3_34.
4. Pya, N., Wood, S.N. (2015). Shape constrained additive models. Stat. Comput. 25, 543–559. 10.1007/s11222-013-9448-7.