CANCER

Cell-free DNA TAPS provides multimodal information for early cancer detection

Paulina Siejka-Zielińska[1,2†], Jingfei Cheng[1,2†], Felix Jackson[1,2,3†], Yibin Liu[1,2‡], Zahir Soonawalla[4], Srikanth Reddy[5], Michael Silva[6], Luminita Puta[1], Misti Vanette McCain[7,8], Emma L. Culver[9], Noor Bekkali[10], Benjamin Schuster-Böckler[1,11], Pier Francesco Palamara[12,13], Derek Mann[7,14,15], Helen Reeves[7,8,16], Eleanor Barnes[9], Shivan Sivakumar[17,18,19]*, Chun-Xiao Song[1,2]*

Multimodal, genome-wide characterization of epigenetic and genetic information in circulating cell-free DNA (cfDNA) could enable more sensitive early cancer detection, but it is technologically challenging. Recently, we developed TET-assisted pyridine borane sequencing (TAPS), which is a mild, bisulfite-free method for base-resolution direct DNA methylation sequencing. Here, we optimized TAPS for cfDNA (cfTAPS) to provide high-quality and high-depth whole-genome cell-free methylomes. We applied cfTAPS to 85 cfDNA samples from patients with hepatocellular carcinoma (HCC) or pancreatic ductal adenocarcinoma (PDAC) and noncancer controls. From only 10 ng of cfDNA (1 to 3 ml of plasma), we generated the most comprehensive cfDNA methylome to date. We demonstrated that cfTAPS provides multimodal information about cfDNA characteristics, including DNA methylation, tissue of origin, and DNA fragmentation. Integrated analysis of these epigenetic and genetic features enables accurate identification of early HCC and PDAC.

INTRODUCTION

Although recent advances in cancer research offer new ways to treat cancer, early detection still represents the best opportunity for curing cancer. Early-stage treatment not only greatly improves patient survival but also costs considerably less. Circulating cell-free DNA (cfDNA)—the free-floating DNA in blood plasma originating from cell death in various healthy and diseased tissues—holds tremendous potential to develop an early cancer detection assay (*1*). Genetic information in cfDNA, such as mutations and copy number variations (CNVs), demonstrate potential utility for monitoring cancer progression and treatment (*2*–*4*). However, genetic alterations are challenging to detect given the low fraction of tumor DNA in early-stage disease (*5*). Furthermore, genetic alterations are weakly informative about the tissue of origin needed to determine the location of malignancy (*6*).

In contrast, widespread epigenetic changes such as DNA methylation of both cancer cells and tumor microenvironment occur early in tumorigenesis (*7*). Recent studies have shown cfDNA methylation to be one of the most promising biomarkers for early cancer detection by providing thousands of methylation changes that can be combined to overcome detection limits and tissue-of-origin information that allows cancer localization with high confidence (*8*–*16*). DNA methylation is best determined by a whole-genome, base-resolution, and quantitative sequencing method, such as bisulfite sequencing. However, bisulfite sequencing is DNA damaging and expensive, so current cfDNA methylation sequencing is limited by being low-depth (*10*, *11*, *14*), targeted (*8*, *12*, *13*, *15*), or low-resolution and qualitative enrichment–based sequencing (*9*), thus imperfectly capturing the cfDNA methylome.

Recently, we developed TET-assisted pyridine borane sequencing (TAPS), a new bisulfite-free DNA methylation sequencing method (*17*). TAPS used mild chemistry to detect DNA methylation directly and showed improved sequence quality, mapping rate, and coverage compared to bisulfite sequencing, while reducing sequencing cost by half (*17*). The combination of direct methylation detection and the nondestructive nature of TAPS makes it ideal not only for DNA methylation analysis but also for simultaneous genetic analysis in cfDNA, which could enhance noninvasive cancer detection by liquid biopsies (*18*). Here, we optimized TAPS for cfDNA (cfTAPS) to deliver high-quality and high-depth whole-genome methylome from only 10 ng of cfDNA.

We applied cfTAPS to hepatocellular carcinoma (HCC) and pancreatic ductal adenocarcinoma (PDAC) cfDNA, two cancer types with particularly poor prognosis mostly due to detection at an advanced disease stage (*19*, *20*). Noninvasive methods for early detection of PDAC and HCC are not available, which contributes to their late diagnosis. For decades, HCC detection has relied on liver ultrasound, combined with serum α-fetoprotein (AFP) measurements (*21*). However, these methods have low specificity and sensitivity (*21*, *22*). There is no blood test to detect or diagnose PDAC. Carbohydrate antigen 19-9 (CA19-9) is used for monitoring PDAC treatment and

[1]Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [2]Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [3]Department of Computer Science, University of Oxford, Oxford, UK. [4]Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [5]Oxford Transplant Centre, Churchill Hospital, Oxford, UK. [6]Department of HPB Surgery, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [7]Newcastle University Translational and Clinical Research Institute, The Medical School, Newcastle University, Newcastle upon Tyne, UK. [8]Newcastle University Centre for Cancer, The Medical School, Newcastle University, Newcastle upon Tyne, UK. [9]Peter Medawar Building and Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford, Oxford, UK. [10]Department of Gastroenterology, John Radcliffe Hospital, Oxford University Hospitals NHS Trust, Oxford, UK. [11]Big Data Institute, University of Oxford, Oxford, UK. [12]Department of Statistics, University of Oxford, Oxford, UK. [13]Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. [14]Newcastle Fibrosis Research Group, Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. [15]Fibrofind, Medical School, Newcastle University, Newcastle upon Tyne, UK. [16]Liver Unit, Freeman Hospital, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. [17]Department of Oncology, University of Oxford, Oxford, UK. [18]Department of Oncology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK. [19]Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK.
*Corresponding author. Email: chunxiao.song@ludwig.ox.ac.uk (C.-X.S.); shivan.sivakumar@oncology.ox.ac.uk (S.S.)
†These authors contributed equally to this work.
‡Present address: Exact Sciences Innovation, Innovation Building, Oxford OX3 7FZ, UK.

development, but its sensitivity and specificity are too low to diagnose or screen for PDAC (*23*). Therefore, novel approaches for PDAC and HCC detection are urgently needed.

Here, we demonstrated that the rich information from cfTAPS enables integrated multimodal epigenetic and genetic analysis of differential methylation, tissue of origin, and fragmentation profiles to accurately distinguish cfDNA samples from patients with HCC and PDAC from controls and patients with precancerous inflammatory conditions.

RESULTS

Adaptation of cfTAPS sequencing

We first optimized the TAPS protocol to work with low-input cfDNA (10 ng, purified from 1 to 3 ml of plasma). Briefly, 10 ng of cfDNA is first ligated to Illumina adapters and 100 ng of carrier DNA is then added to the sample before TET oxidation and pyridine borane (PyBr) reduction steps (Fig. 1A). We found that the addition of carrier DNA improves the recovery of cfDNA during the workflow and results in higher library yields when compared to the standard TAPS protocol (fig. S1A) (*17*). Subsequently, 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC) in cfDNA are oxidized by mTet1CD enzyme to 5-carboxylcytosine (5caC) and reduced to dihydrouracil (DHU), which is amplified as T in the final polymerase chain reaction (PCR) step (Fig. 1A).

We applied cfTAPS to 87 cfDNA samples. Libraries were sequenced to a mean of 360 million read pairs (11.6× mean depth, range 8.2 to 22×) and resulted in high unique mapping rate and unique deduplicated mapping rate of 94.8 and 77.1%, respectively (Fig. 1B and table S1). Among the mapped reads, 99.95% were mapped to the human genome (fig. S1B). In comparison, a recent cfDNA whole-genome bisulfite sequencing (WGBS) study (*24*) sequenced to a similar depth (a mean of 371 million read pairs) and resulted in significantly lower unique mapping rate (63.6%) and unique deduplicated mapping rate (53.9%) (fig. S1C), although it used more cfDNA input (from 5 ml of plasma). This highlights the advantage of cfTAPS to generate higher-quality and more complex data than cfDNA WGBS while requiring less cfDNA input.

Subsequently, we assessed accuracy of cfTAPS to detect 5mC based on spike-in controls that have modified and unmodified cytosines in the known positions. We used CpG-methylated lambda DNA to estimate the conversion of 5mC. Two samples had a low conversion rate below 85% and were excluded from downstream analysis (table S1). The remaining 85 samples had a mean 5mC conversion rate of 97.0% or a false-negative rate (nonconversion rate of 5mC) of 3.0% (Fig. 1C). The false-positive rate (conversion rate of unmodified C), estimated on the basis of unmodified amplicon spike-in, was 0.28%, which confirms that cfTAPS allows highly sensitive and specific detection of 5mC in cfDNA (Fig. 1C). We further confirmed high reproducibility of cfTAPS between technical replicates (fig. S1D).

Whole-genome DNA methylation from cfTAPS

Next, we sought to characterize the cfDNA methylome in the 85 cfDNA samples that passed initial quality control. Our cohort included samples from 21 patients with HCC, 23 with PDAC, 30 noncancer controls, 4 patients with cirrhosis, and 7 with pancreatitis (fig. S2A). Cirrhosis and pancreatitis are precancerous conditions affecting the liver and pancreas, respectively (*25*, *26*). Most PDAC and HCC patients in our cohort were at a nonmetastatic stage, with 52% of

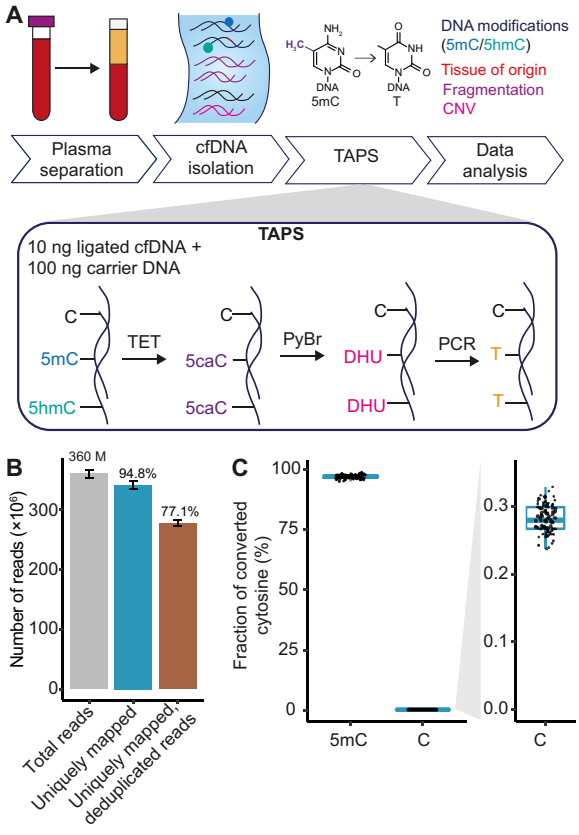

Fig. 1. cfDNA analysis by TAPS. (**A**) Schematic representation of the TAPS approach for cfDNA analysis. CfDNA is isolated from 1 to 3 ml of plasma. cfDNA (10 ng) is ligated to Illumina sequencing adapters and topped up with 100 ng of carrier DNA. Subsequently, 5mC and 5hmC in DNA are oxidized by mTet1CD enzyme to 5caC, reduced by PyBr to DHU, and amplified and detected as T in the final sequencing. Computational analysis of TAPS data allows simultaneous characterization of multiple cfDNA features including DNA methylation, tissue of origin, fragmentation patterns, and CNVs. (**B**) Number of total reads, uniquely mapped reads, and uniquely mapped, PCR deduplicated reads in 87 cfDNA TAPS libraries. Total number of reads, mean percentage of uniquely mapped reads, and deduplicated reads compared to total reads are shown above the bars. Error bars represent SE. (**C**) 5mC conversion rate and false-positive rate in 85 cfDNA TAPS libraries based on spike-in controls with modified or unmodified cytosines at the known positions. Each dot represents an individual sample.

PDAC and 67% HCC patients at stages I and II (Fig. 2A and table S2). Among the 21 HCC patients, only 4 (19%) had elevated levels of AFP (over 20 ng/ml; table S2) (*22*). Among the 18 PDAC patients who had CA19-9 measurement, 16 (89%) had elevated levels of CA19-9 (over 37 U/ml; table S2) (*27*). However, CA19-9 level is often elevated in nonmalignant conditions including inflammatory disease (*23*). Of note, our noncancer controls were collected from an endoscopy clinic and were enriched with gastrointestinal inflammatory conditions such as Crohn's disease and colitis (table S2). While distinguishing these noncancer controls from cancer patients is more challenging than a typically healthy control group, this may provide a more real-world comparison of a diagnostic test in an aging population.

We first analyzed global methylation levels of cfDNA in cancer and control samples. CfDNA methylation displayed a typical bimodal distribution in all groups, with most CpG sites either fully methylated

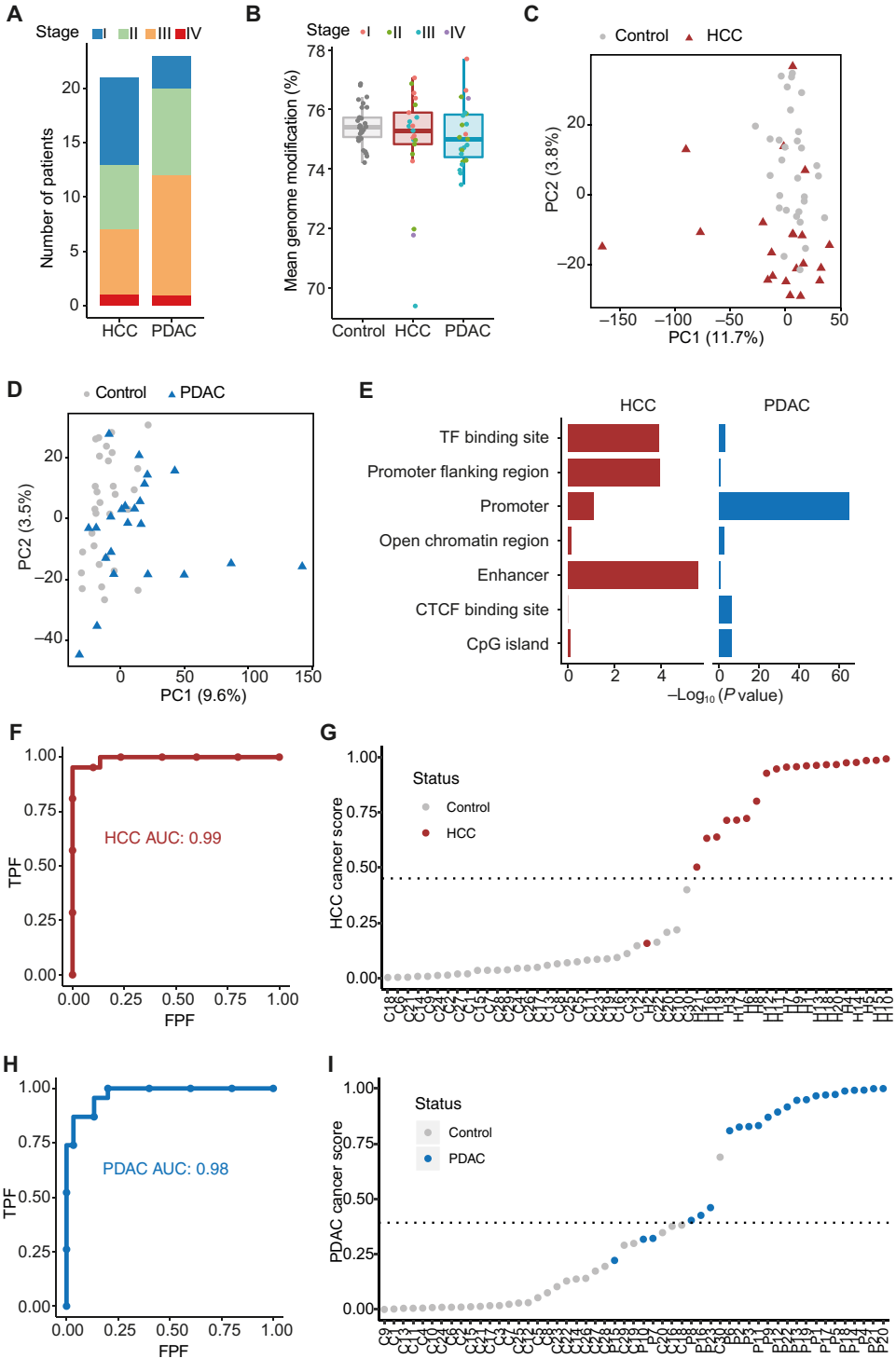

Fig. 2. cfDNA methylation in clinical samples. (**A**) Cancer stage distribution of 21 HCC and 23 PDAC patients included in the study. (**B**) Mean per CpG genome modification level in noncancer controls, HCC, and PDAC cfDNA. Each dot represents an individual sample. (**C**) PCA plot of cfDNA methylation in 1-kb genomic windows in noncancer controls and HCC. (**D**) PCA plots of cfDNA methylation in 1-kb genomic windows in noncancer controls and PDAC. (**E**) The overrepresentation analysis on the regions correlated most with PC2 for HCC and PC1 for PDAC in regulatory regions. (**F**) ROC curve of model classification performance based on differentially methylated enhancers in HCC and noncancer controls ($n = 51$, HCC = 21, noncancer controls = 30). FPF, false positive fraction; TPF, true positive fraction. (**G**) LOO cancer prediction scores for HCC and noncancer controls. Dashed line represents probability score threshold. Samples with a probability score above this threshold were predicted as HCC. (**H**) ROC curve of model classification performance based on differentially methylated promoters between PDAC and noncancer controls ($n = 53$, PDAC = 23, noncancer controls = 30). (**I**) LOO cancer prediction scores for PDAC and noncancer controls. Dashed line represents probability score threshold. Samples with a probability score above this threshold were predicted as PDAC.

or unmethylated (fig. S2B). Average CpG methylation level in control samples was 75.5% and was similar in cancer cfDNA (HCC, 74.9%; PDAC, 75.1%). Previously reported global cfDNA hypomethylation in HCC (*10*) was only observed in a few samples with late stage or large tumor size (Fig. 2B and fig. S2, C to F). By contrast, we observed a higher variance of methylation in 1-Mb genomic windows between cancer patients and controls (fig. S2, G and H).

To investigate whether whole-genome cfDNA methylation signatures have the potential to discriminate between cancer patients and noncancer controls, we first performed principal components analysis (PCA) of cfDNA methylation in 1-kb genomic windows. Both HCC (Fig. 2C) and PDAC samples (Fig. 2D) showed partial separation from controls in principal component 2 (PC2) and PC1, respectively. Note that the inflammatory patients (Crohn's disease and colitis) do not separate from the other noncancer controls (fig. S2I). We then investigated where the windows that most contributed to the cancer/control separation were enriched in the genome. We found that the top 200 windows with the highest correlation with PC2 for HCC were enriched in enhancers (see Materials and Methods; Fig. 2E). Conversely, the 200 windows most highly correlated with PC1 for PDAC were highly enriched in promoters (Fig. 2E), suggesting that different cancer types have different cfDNA methylation signals.

Differential DNA methylation from cfTAPS

Because methylation patterns in regulatory regions substantially contributed to discrimination between cancer and controls in unsupervised analysis, we investigated the predictive potential of cfDNA methylation in enhancer and promoter regions for HCC and PDAC prediction, respectively, using a supervised machine learning approach with leave-one-out (LOO) cross-validation. Briefly, in each round of LOO cross-validation, we used one sample as a validation set and the remaining samples for model training. Within each fold, we identified differentially methylated enhancers and promoters for HCC and PDAC, respectively, and used them to train a regularized generalized linear model classifier (glmnet) (*28*) to distinguish each cancer type from the control samples. This model was then evaluated on the held-out test sample for each fold (fig. S3A). Cirrhosis and pancreatitis samples were not included in model building but were used as an independent validation set to evaluate performance of our classifiers to discriminate between cancer and premalignant conditions.

We achieved excellent prediction of HCC [AUC (area under the curve) = 0.99] based on differentially methylated enhancers (see Materials and Methods; Fig. 2, F and G, and table S3). Moreover, on the basis of predicted scores, three of four cirrhosis samples could be distinguished from HCC, suggesting that our model is able to detect cancer-specific features (fig. S3B). We then performed gene ontology (GO) analysis on the differentially methylated enhancers and found significant enrichment in signaling pathways commonly affected in liver cancer, including regulation of RAC1 activity (*29*) and interleukin-8 (IL-8)– and CXCR1-mediated signaling (fig. S3C) (*30*). For example, in cfDNA of HCC patients, we observed significant hypermethylation of the enhancer that regulates expression of the *DLC1* gene, a tumor suppressor for human liver cancer involved in RAC1 and Rho signaling pathways (fig. S3D) (*31*).

We achieved accurate prediction of PDAC (AUC = 0.98) based on differentially methylated promoters (see Materials and Methods; Fig. 2, H and I, and table S4). Similarly, the classifier was able to predict six of seven pancreatitis samples as noncancer despite not being trained on any pancreatitis samples (fig. S3E). Differentially methylated promoters in PDAC cfDNA were enriched in signaling pathways affected in PDAC, including RB1 regulation (*32*) and p38 signaling pathways (fig. S3F) (*33*). For instance, we found significant hypermethylation in the *RB1* gene promoter (fig. S3G), a well-studied tumor suppressor gene. Hypermethylation of *RB1* promoter was previously found in human cancers (*34*), and down-regulation (*35*) of RB1 was reported in pancreatic cancer.

Last, we validated our HCC model on an independent dataset from a recent cfDNA WGBS study, which contains four HCC patients and four noncancer controls (*24*). We found that our models, built on differentially methylated enhancers identified from cfTAPS data, were able to correctly classify all HCC and noncancer controls from this external dataset (fig. S3H). The high sequencing depth of cfTAPS is essential for de novo differential methylation analysis from cfDNA (*36*), and the differentially methylated regions (DMRs) identified were significantly decreased when we down-sampled our data to 100 to 200 million read pairs (fig. S3I). Together, cfTAPS enables whole-genome discovery of DMRs in cfDNA, and the distinct methylation patterns in regulatory regions enable accurate prediction of HCC and PDAC.

cfTAPS informs tissue of origin

CfDNA methylation has been shown to provide tissue-of-origin information (*8*, *9*, *11*–*14*). Most approaches use 450K methylation array tissue data (*9*, *13*), which covers less than 1% of CpGs in the human genome, to infer tissue contribution from cfDNA methylation. To further use the whole-genome information from cfTAPS for cfDNA deconvolution (*11*, *14*), we collated CpG-level methylation data from 144 publicly available tissue and blood cell WGBS, stratified into 32 physiologically distinct tissue and blood cell types, including liver tumor tissue (table S5). Given the prevalence of tissue-specific DNA methylation in enhancer regions (*37*), we constructed an enhancer-aggregated reference map of tissue methylation (see Materials and Methods). The resulting methylation reference map displays good clustering of blood and immune cell types and even physiologically related solid tissues (fig. S4A).

We calculated tissue contribution in cfTAPS samples by performing nonnegative least squares regression (NNLS) (*13*, *14*). cfDNA tissue contribution was broadly similar between cancer and control groups, in agreement with previous reports (*13*, *14*), with blood and immune cells dominant, and lower proportions of solid tissues (Fig. 3A, fig. S4B, and table S6). We observed a significantly increased liver tumor contribution in HCC alone (paired *t* test, $P = 0.0016$; Fig. 3B) and a significantly increased memory T cell contribution in PDAC samples (paired *t* test, $P = 0.028$; fig. S4C). We trained a regularized generalized linear model based on tissue contribution, evaluating all samples using LOO cross-validation, and showed that it can correctly separate most samples in both cancer types (HCC versus noncancer control, AUC = 0.77; PDAC versus noncancer control, AUC = 0.81). However, these models perform worse at distinguishing pancreatitis and cirrhosis compared to methylation-based models (fig. S4, D to I). Tissue deconvolution is currently limited by the availability of public WGBS data. Nevertheless, these results indicate that cfTAPS provides valuable tissue-of-origin information for early cancer detection.

Fragmentation patterns from cfTAPS

Although the main purpose of cfTAPS is DNA methylation sequencing, it only induces base changes at modified cytosines, thus

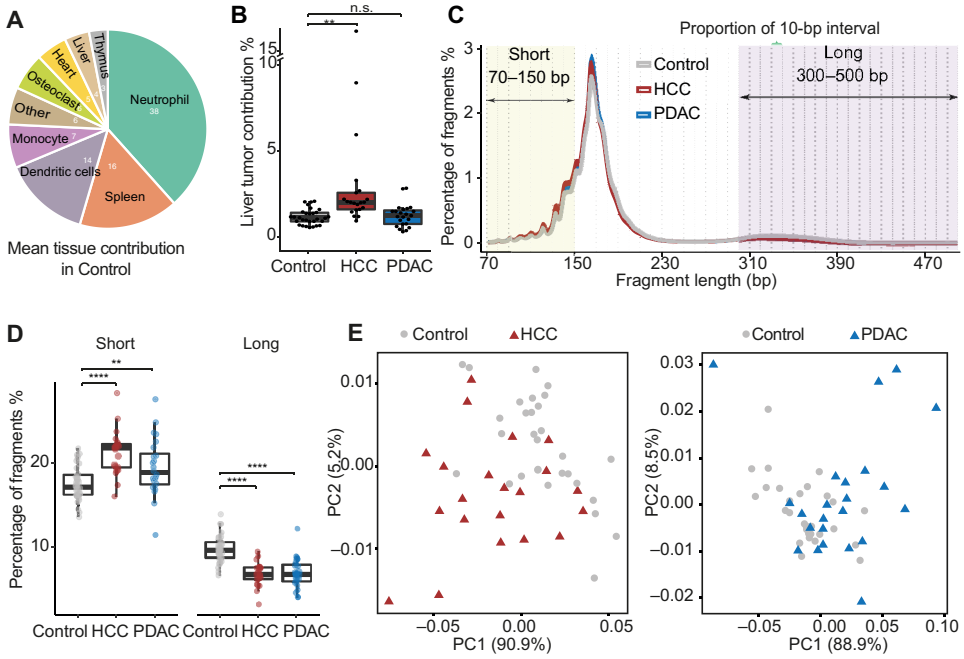

Fig. 3. cfTAPS enables analysis of tissue of origin and fragmentation patterns in cfDNA. (**A**) Mean tissue contribution in noncancer individuals estimated by NNLS. Tissue contributions less than 1.5% are aggregated as "Other." (**B**) Boxplot showing the estimated liver cancer contribution within noncancer, HCC, and PDAC groups. Statistical significance was assessed with a paired *t* test. n.s., not significant. (**C**) Length distribution of cfDNA fragments in the three groups. For each sample, proportion in 10-bp intervals of long cfDNA fragments (300 to 500 bp) was used as fragmentation features for PCA and machine learning. (**D**) Boxplot showing proportion of short (70 to 150 bp) and long (300 to 500 bp) fragments in noncancer controls, PDAC, and HCC. The Kruskal-Wallis test was performed to test differences in fragment size distribution between groups. Statistically significant differences are marked with asterisks (**$P < 0.01$, ****$P < 0.0001$). (**E**) PCA plot of cfDNA 10-bp fragment fraction in noncancer controls and HCC (left) and noncancer controls and PDAC (right).

keeping most of the DNA intact. We can therefore extract additional genetic information from cfTAPS data to further improve the sensitivity of early cancer detection. We first investigated CNVs from cfTAPS data. As expected with our nonadvanced cancer cohort, we only predicted CNVs in four HCC and three PDAC patients (fig. S5, A and B). Next, we investigated whether cfTAPS can retain reliable cfDNA fragmentation information, which has recently been shown to change substantially during cancer development and has therefore been adopted in cancer detection assays (*38*, *39*).

We first confirmed that cfDNA fragmentation patterns detected with cfTAPS are concordant with cfDNA fragmentation pattern generated by whole-genome sequencing (WGS) (*38*), with the dominant peak at 167 base pairs (bp), a secondary peak at ~320 bp, and smaller peaks below 167 bp with 10-bp periodicity, reflecting nucleosomal fragmentation patterns (Fig. 3C and table S7). By contrast, fragmentation patterns were clearly different in previously published cfDNA WGBS (*24*), as the 10-bp oscillations in the cfDNA fragmentation profile were lost presumably because of DNA damage (fig. S6A). Consistent with previous cfDNA WGS (*38*), we found that cancer patients have a higher frequency of cfDNA fragments below 150 bp (Kruskal-Wallis test: HCC, $P = 6.871 \times 10^{-6}$; PDAC, $P = 0.006731$) and a lower proportion of long fragments between 310 and 500 bp (Kruskal-Wallis test: HCC, $P = 2.627 \times 10^{-7}$; PDAC, $P = 1.263 \times 10^{-6}$) compared to noncancer controls (Fig. 3D), further confirming the faithful preservation of cfDNA fragmentation information in cfTAPS.

We then developed a new approach for characterization of cfDNA fragmentation profiles using cfTAPS. Briefly, we divided the cfDNA fragmentation distribution into 10-bp bins and calculated the proportion of fragments in each 10-bp bin (Fig. 3C). We found that cfDNA long fragment (300 to 500 bp) length proportion in 10-bp bins separated PDAC and HCC from controls in PCA (Fig. 3E). We further showed that this cfDNA fragmentation signature can be used to distinguish HCC and PDAC from noncancer controls with high accuracy (HCC, AUC = 0.92; PDAC, AUC = 0.84) (fig. S6, B, C, E, and F). However, this approach was less accurate at distinguishing cancer from cirrhosis and pancreatitis compared to methylation-based classifiers (fig. S6, D and G), suggesting that fragmentation information is less cancer specific.

Multi-cancer detection with cfTAPS

We next investigated the utility of cfTAPS for multicancer detection. We selected top five DMRs of each pairwise comparison (noncancer controls versus HCC, noncancer controls versus PDAC, and HCC versus PDAC; see Materials and Methods) as features in the multicancer differential methylation model. We trained a support vector machine (SVM) model to estimate the respective probability that the blood sample came from each group. We built similar models using tissue contribution and fragmentation profile. Using LOO cross-validation, we found that the methylation model can achieve an overall accuracy of 0.77, which outperforms the tissue contribution model and fragmentation profile model (accuracy of 0.62 and 0.46, respectively; Fig. 4A and fig. S7A).

To further enhance the multicancer predictive model, we built a multimodal classifier that combined differential methylation, tissue contribution, and fragment profile (Fig. 4B). This integrated model

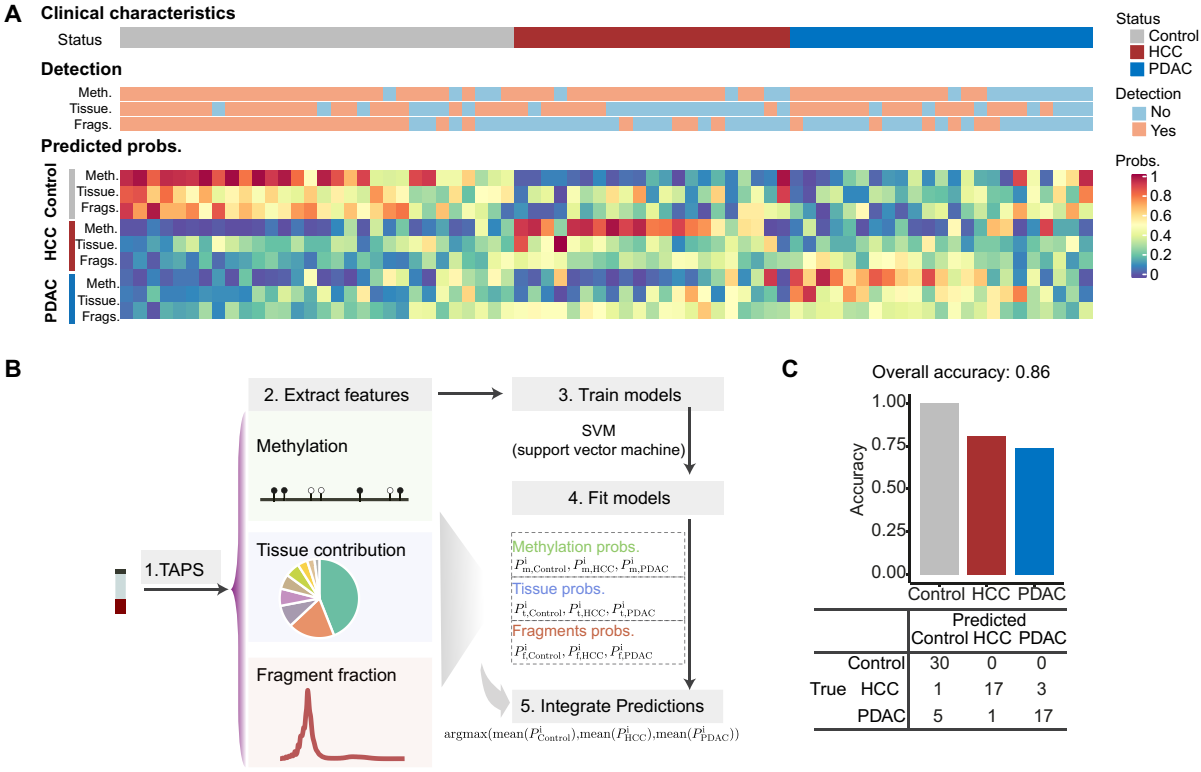

Fig. 4. Integrating multimodal features from cfTAPS enhances multicancer detection. (**A**) Heatmap showing individual model performance on multicancer prediction and the predicted probabilities for each patient. Each vertical column is a patient. Detection yes/no means that patients are correctly classified or misclassified on the basis of a particular feature. Predicted score means the probability of classifying the patients to a specific group based on a particular feature. (**B**) Schematic detailing the method of integrating multiple features (DNA methylation, tissue contribution, and fragmentation fraction) extracted from cfTAPS data for multicancer prediction. (**C**) Actual and predicted patient status calculated in LOO cross-validation.

took the averaged scores across the three modalities and used the most confident prediction for each sample. The overall accuracy of the combined model was 0.86 (64 of 74 were classified correctly), and the accuracy for distinguishing controls from any cancer type is 0.92 (Fig. 4C), which highlights the benefits of incorporating multimodal information for cancer type prediction. Last, we explored the DMRs used for multicancer prediction (fig. S7B and table S8). We found that the nearby genes of these regions were enriched in Notch and Wnt signaling (*40*) and epidermal growth factor receptor (ErbB) signaling (*41*), which provides biological support for these potential multicancer biomarkers (fig. S7C).

DISCUSSION

In this study, we successfully optimized and applied cfTAPS to characterize whole-genome base-resolution methylome in cfDNA from HCC, PDAC, and noncancer controls. Using only 10 ng of cfDNA, cfTAPS libraries demonstrated greatly improved sequencing quality and depth compared to previous cfDNA WGBS. Using less cfDNA input than previous studies, cfDNA TAPS generated the most comprehensive cell-free methylation to date. The much higher yield of informative reads allows cfTAPS to extract more information from a given amount of cfDNA and makes it a viable option for large-scale cfDNA methylation studies.

The deep sequencing achieved by cfTAPS enables detailed analysis of the cell-free methylome and whole-genome discovery of methylation biomarkers for early cancer detection. While we do not observe significant global hypomethylation, suggesting that in our cohort the fraction of cfDNA derived from tumor cells is low (as corroborated by the lack of CNVs in most cancer patients included in the study), we found that local methylation signals in regulatory regions such as enhancers and promoters contained cancer-specific information that could accurately distinguish HCC and PDAC from controls. This is particularly promising considering the inflammation-enriched real-world control group used in our patient cohort and that our HCC model can correctly identify all HCC and control patients from a cfDNA WGBS dataset as an independent validation (*24*).

Another important advantage of cfDNA methylation for early cancer detection is the ability to determine tissue-of-origin information. Using currently available public WGBS tissue databases, we performed a whole-genome tissue deconvolution of cfTAPS data and observed increased liver tumor contribution in HCC cfDNA and distinct immune signatures in cancer cfDNA. The tissue deconvolution itself can be used for cancer detection. Last, because TAPS converts modified cytosine directly, it maximally retains the underlying genetic information compared to other approaches that convert unmodified cytosines. In this study, we extracted CNVs and fragmentation information from cfTAPS, the latter of which is lost in cfDNA WGBS. We further demonstrated that an integrated approach combining differential methylation, tissue of origin, and fragmentation profiles could improve the model performance for multicancer detection.

Despite promising results, our study has marked limitations. First, our approach was tested on a relatively low number of patients. We were able to validate our methylation-based HCC classifier on an independent cfDNA HCC WGBS dataset. However, because of limited sample size and lack of available public cfDNA PDAC whole-genome methylation data, findings of this study have yet to be validated on a bigger, independent cohort, including the performance comparison to AFP and CA19-9 (AFP and CA19-9 measurements were not available for the noncancer controls in our study). Second, because of the limited amount of publicly available genome-wide tissue methylation data (for instance, no PDAC tissue data were available) and the fact that we are comparing TAPS data to a WGBS database, the current investigation was far from the full potential of tissue-of-origin information from cfTAPS (*42*). In the future, a comprehensive cell type–specific TAPS reference database could improve the resolution and sensitivity of tissue deconvolution from cfTAPS considerably. We note that emerging novel computational approaches, such as those that use read-level methylation information (*24*), will also aid future methylation deconvolution efforts. Third, our predictive pipeline uses simple, out-the-box classification models to demonstrate the signal present in cfTAPS data; as we collect more data, future work will explore more sophisticated machine learning models that allow automated feature extraction from genome-wide methylomes. cfTAPS should retain various other genetic and epigenetic information such as microbiome (*14*) and nucleosome positioning (*43*), which are interesting avenues to explore in subsequent work. Future studies with the proper design (i.e., germline controls) would also allow simultaneous detection of mutations from cfTAPS in either a whole-genome or targeted sequencing manner. These additional modalities from a single cfTAPS assay would further improve sensitivity and ability to detect outliers within highly heterogeneous disease cohorts (*18*).

In this first proof-of-concept study, we conducted deep whole-genome methylome sequencing with cfTAPS, which is necessary for unbiased methylation marker discovery at the whole-genome level. These potential methylation markers can be used in targeted cfTAPS for future large-scale validation studies to reduce the sequencing cost. On the other hand, we showed that whole-genome cfTAPS has the benefit of offering a wealth of information such as tissue-of-origin and fragmentation profiles. Recently, this type of breadth-over-depth approach has been demonstrated in whole-genome cfDNA mutational sequencing in the minimal residual disease setting (*3*, *4*). While the cost of sequencing continues to drop, it remains to be tested whether whole-genome or targeted DNA methylation sequencing is the best approach for early cancer detection.

MATERIALS AND METHODS

Experimental design

Whole-blood samples from 30 noncancer controls were obtained from John Radcliffe Hospital (ethical approvals IDs 16/YH/0247 and 18/WM/0237). Pancreatitis blood samples from eight patients were obtained from John Radcliffe Hospital. The study was approved by Oxfordshire Research Ethics Committee A (REC-A) (10/H0604/51) and is registered on the U.K. National Institute for Health Research (NIHR) portfolio as study number 10776. PDAC patients were consented for this study via the Oxford Radcliffe Biobank (09/H0606/5+5, project: 19/A177), and whole-blood samples were collected from 24 patients. Collection of plasma samples from 21 HCC and 4 cirrhosis patients was REC-approved [ethical approval 2/NE/0395, IRAS (Integrated Research Application System) project ID: 116370]. No sample size calculations were performed. Sample size was determined on the basis of availability. PDAC, HCC, pancreatitis, and cirrhosis samples were collected from subjects with clinically diagnosed disease. Noncancer control samples were collected from individuals without cancer diagnosis at the time of sample collection or previous history of cancer.

The main goal of the study was comprehensive, multidimensional characterization of cfDNA in cancer and controls by whole-genome methylation sequencing using TAPS. CfDNA TAPS libraries were constructed and paired-end 150 bp–sequenced on a NovaSeq 6000 sequencer (Illumina). Technical details are described in the sections below. Samples with 5mC conversion below 90% calculated on the basis of methylated lambda spike-in control were excluded from downstream analysis.

Collection and preparation of cfDNA samples

Blood was collected into EDTA-coated Vacutainers. Plasma was separated from collected blood samples within 4 hours from collection. Plasma was collected by centrifuging blood at 1600*g* for 10 min at 4°C and 16,000*g* for 10 min at 4°C and stored at −80°C for cfDNA purification. cfDNA from plasma was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen). cfDNA was quantified with Qubit Fluorometer (Life Technologies).

Preparation of carrier DNA and spike-in controls

Carrier DNA was prepared by PCR amplification of the pNIC28-Bsa4 plasmid (Addgene, catalog no. 26103) in a reaction containing 1 ng of DNA template, 0.5 μM primers (forward: 5′-AGGCAACTTTATGCCCATGCAA-3′, reverse: 5′-CCAAGGGGTTATGCTAGTTATTGC-3′), and 1× Phusion High-Fidelity PCR Master Mix with HF Buffer (Thermo Fisher Scientific). The CpG-methylated lambda DNA and 2-kb unmodified spike-in control DNA were prepared as described previously (*17*). CpG-methylated lambda DNA, carrier DNA, and 2-kb unmodified control were fragmented by Covaris M220 (peak incident power, 50 W; duty factor, 20%; cycles per burst, 200; time, 150 s) and size-selected on 0.9 to 1.2× AMPure XP beads to select for 150- to 250-bp fragments.

Preparation of sequencing adapters

Adapter oligos (5′-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-3′, 5′-/5Phos/GATCGGAAGAGCACACGTCT-3′) were obtained from (Integrated DNA Technologies) with high-performance liquid chromatography purification. Adapter oligos were annealed together in a 50-μl reaction containing 15 μM each oligo, 10 mM tris-Cl (pH 8.0), 0.1 mM EDTA (pH 8.0), and 50 mM NaCl with the following program: 2 min at 95°C, 140 cycles of 20 s at 95°C (decrease temperature 0.5°C every cycle), and hold at 4°C. Annealed 15 μM Illumina multiplexing adapters were then aliquoted into small single-use vials and stored at −80°C.

mTet1CD oxidation

mTet1CD was prepared as described previously (*17*). DNA was incubated in a 50-μl reaction containing 50 mM Hepes buffer (pH 8.0), 100 μM ammonium iron (II) sulfate, 1 mM α-ketoglutarate, 2 mM ascorbic acid, 2 mM dithiothreitol, 100 mM NaCl, 1.2 mM adenosine triphosphate, and 4 μM mTet1CD for 80 min at 37°C. After that, 0.8 U of proteinase K (New England Biolabs) was added to the reaction mixture and incubated for 1 hour at 50°C. The product was cleaned

up on Bio-Spin P-30 Gel Column (Bio-Rad) and 1.8× AMPure XP beads following the manufacturer's instruction.

PyBr reduction

Oxidized DNA in 35 μl of water was reduced in a 50-μl reaction containing 600 mM sodium acetate solution (pH 4.3) and 1 M PyBr (Alfa Aesar) for 16 hours at 37°C and 850 rpm in the Eppendorf ThermoMixer. The product was purified using Zymo-Spin columns.

cfDNA TAPS

cfDNA (10 ng) was spiked-in with 0.15% CpG-methylated lambda DNA and 0.015% unmodified 2-kb control and used for an end-repair and A-tailing reaction and ligated to Illumina Multiplexing adapters with a KAPA HyperPrep kit according to the manufacturer's protocol. Subsequently, 100 ng of carrier DNA was added to ligated libraries and samples were double-oxidized with mTet1CD and reduced with PyBr as described above. Converted libraries were amplified using NEBNext Multiplex Oligos for Illumina (96 Unique Dual Index Primer Pairs) with KAPA Hifi Uracil Plus Polymerase for seven cycles and cleaned up on 1× AMPure XP beads. CfDNA TAPS libraries were paired-end 150 bp–sequenced on a NovaSeq 6000 sequencer (Illumina).

TAPS mapping and preprocessing

Raw sequenced reads were processed with trim:galore (version 0.6.2; https://bioinformatics.babraham.ac.uk/projects/trim:galore/) to trim adapter and low-quality bases with the following parameters: --paired --length 35 --gzip --cores 2. Clean reads were aligned to human reference genome (GRCh38, ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/seqs_for_alignment_pipelines.ucsc_ids/GCA_000001405.15_GRCh38_no_alt_analysis_set.fna.gz.) combining spike-in sequences using bwa mem (*44*) (version 0.7.17-r1188) with the following parameters: -I 500,120,1000,20. Reads with MAPping Quality Value (MAPQ) <1 were excluded from further analysis. Picard MarkDuplicates (version 2.18.29-SNAPSHOT) was used to identify duplicate reads. MethylDackel extract (version 0.5.0; https://github.com/dpryan79/MethylDackel) was used for methylation calling using the following parameters: -q 10 -p 13 -t 4 --mergeContext --OT 10,140,75,75 --OB 10,140,75,75. CpG sites overlapped with common single-nucleotide polymorphism (SNP) (dbSNP153) (*45*), blacklisted regions (*46*), centromeres (*45*), and sex chromosomes were excluded for further analysis.

cfDNA WGBS analysis

CfDNA WGBS data were downloaded from EGAD00001004317 (*24*). Raw sequenced reads were processed with trim:galore (version 0.6.2; https://bioinformatics.babraham.ac.uk/projects/trim:galore/): We trimmed adapter and low-quality bases with the following parameters: --paired --length 35 --gzip --cores 2. Clean reads were aligned to human reference genome (GRCh38) using bismark (*47*) (Bismark version v0.22.0) with default parameters. deduplicate_bismark was used for deduplication. Samtools (*48*) was used to filter the fragments with -q 10, and only reads mapped in proper pairs were used for fragmentation analysis. Then, bismark_methylation_extractor was used to extract methylation from deduplicated bam files with default parameters.

PCA on DNA methylation and feature overrepresentation analysis

The genome was binned into 1-kb windows. Methylation level was calculated using the number of methylated CpGs divided by the number of total CpGs sequenced. Windows with mean CpG coverage (number of total CpG sequenced/total number of CpG positions) < 2 were excluded for further analysis. Dimdesc (*49*) was used with parameter proba = 0.01 to determine the regions that contribute most to each principal component obtained by the PCA function (largest eigenvalues of each eigenvector). Bedtools (*50*) fisher was used to test the number of overlaps between the top 200 contributing regions (sorted by absolute correlation value) and the selected genomic features. Selected genomic features included regulatory element (*51*) from Ensemble (ftp://ftp.ensembl.org/pub/release-97/regulation/homo_sapiens/homo_sapiens.GRCh38.Regulatory_Build.regulatory_features.20190329.gff.gz) and CpG islands from UCSC (http://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/cpgIslandExt.txt.gz).

Two-class prediction using DNA methylation signature

Two-class prediction models were trained and evaluated on the basis of a LOO approach. Briefly, one sample was held out as the testing set, while the remaining samples were used for model training. DMRs (promoters for PDAC and enhancers for HCC) were identified in the training set by *t* test ($P < 0.002$, methylation difference > 0.05). In each LOO fold, 443 to 775 differentially methylated enhancers and 160 to 318 differentially methylated promoters were identified in the HCC versus noncancer control and PDAC versus noncancer control feature selection steps, respectively. In total, 1521 enhancers and 531 promoters were selected during the cross-validation process. The predictive model was built on selected DMRs using cv.glmnet (*28*) and validated on the test sample. This procedure was repeated *N* times, where *N* = number of samples. Receiver operating characteristic (ROC) curves were prepared in R based on the predicted scores of held-out test samples from cvglm models. Cirrhosis patients and cfDNA WGBS data (*24*) were used as independent validation sets to evaluate the performance of HCC model. Pancreatitis patients were used as independent validation set to evaluate the performance of PDAC model. Aligned Binary Alignment Map (BAM) files were down-sampled from 100 million to 200 million read pairs using samtools view (*48*). For each down-sampled set, we used the method described above to detect DMRs. Ref DMRs were defined as the total unique DMR in the LOO cross-validations. The percentage of ref DMRs was computed by dividing the overlapped DMR between the down-sampled set and the ref DMR and the total ref DMR.

GO analysis of DMRs

Genes regulated by differentially methylated enhancers in HCC cfDNA were identified using the GeneHancer database (*52*). The genes closest to the differentially methylated promoters in PDAC were identified as related using following R packages: AnnotationHub (version 2.18.0), TxDb.Hsapiens.UCSC.hg38.knownGene (version 3.10.0), and org.Hs.eg.db (version 3.10.0). GO analysis was performed on these identified genes using Enrichr tool (*53*) against National Cancer Institute–Nature Pathway Interaction database.

Tissue reference map

CpG-level tissue methylation data were collated from six public sources (table S5). After filtering diseased, sex-specific, and low-coverage samples, we retained 144 healthy, adult tissue samples, grouped into 32 physiologically distinct tissue groups (table S6). One hundred thirty-three of 144 samples were already aligned to hg38; the remaining 11 samples were converted from hg19 to hg38 using the UCSC hgLiftOver tool (*54*).

We filtered 79,000 enhancers from Ensembl Regulatory Build (*51*) using a tissue-specific DMR finding algorithm similar to Moss *et al.* (*13*). Specifically, this algorithm performs pairwise one-versus-all comparisons for each tissue group in the reference atlas, selecting the regions that show the largest median methylation difference and consistent methylation across the tissue group in question. As in Moss *et al.* (*13*), we also calculated pairwise tissue group correlations and included DMRs that best separated each tissue group from the first and second most highly correlated tissue.

Tissue deconvolution by NNLS regression

Tissue deconvolution was performed using NNLS and implemented using Scipy's optimize function (*55*) in Python 3.8. Given a tissue reference matrix A and a vector of observed methylation ratios y_s in a sample s, we estimate the tissue contribution x by solving the following minimization problem

$$\min \|Ax - y_s\|^2$$

subject to $x \geq 0$

Fragmentation analysis

The length of the DNA fragments was obtained from alignment files using samtools (*48*). Fragmentation profiles were calculated as the fraction of cfDNA fragments at 10-bp length range bins. PCA and plots were generated in R.

For fragmentation-based prediction, the proportion of cfDNA fragments (300 to 500 bp) in 10-bp length range bins was calculated. Models were built and trained by LOO approach using cv.glmnet (*28*) method. ROC curves were prepared in R based on prediction scores from validation.

CNV analysis

Alignment files for each sample were down-sampled to 225 million read pairs with samtools (*48*) view. QDNAseq (*56*) package was used for CNV analysis. The bin annotation was downloaded from QDNAseq.hg38 (https://github.com/asntech/QDNAseq.hg38), and bin size 100 kb was used. Regions that were blacklisted or have mappability less than 80 were excluded for further analysis. Cutoffs 0.8 and 1.2 were used to define copy number losses and gains, respectively, in the callBins function. Patients that have copy number aberrations with length range bigger than 500 kb were classified as patients with CNV.

Three-class prediction models

Three-class prediction models were trained and evaluated on the basis of a LOO approach. For DNA methylation, we initially narrow down the candidate features to 824,320 1-kb windows encompassing mapping to regulatory regions as mentioned previously. The methylation model aims to capture the cancer type–specific methylation change by selecting DMRs based on a pairwise comparison using a *t* test. DMRs were then ranked by *P* value, and the top five DMRs in each pairwise comparison were selected for model training. The prediction model was built on DMRs selected among the training sets using an SVM model implemented in the caret package (*57*) (train method = "svmLinear2") and validated on the test sample. This procedure was repeated N times, where N = number of samples. For tissue contribution and fragmentation fraction, the raw matrices were used to build models following the same method as for DMRs. These three models were integrated by taking the average (mean) predictions across the three modalities, where the selected prediction in each case was the one with the maximum average predicted score.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at https://science.org/doi/10.1126/sciadv.abh0534

View/request a protocol for this paper from *Bio-protocol*.

REFERENCES AND NOTES

1. J. C. M. Wan, C. Massie, J. Garcia-Corbacho, F. Mouliere, J. D. Brenton, C. Caldas, S. Pacey, R. Baird, N. Rosenfeld, Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nat. Rev. Cancer* **17**, 223–238 (2017).
2. C. Abbosh, N. J. Birkbak, G. A. Wilson, M. Jamal-Hanjani, T. Constantin, R. Salari, J. Le Quesne, D. A. Moore, S. Veeriah, R. Rosenthal, T. Marafioti, E. Kirkizlar, T. B. K. Watkins, N. McGranahan, S. Ward, L. Martinson, J. Riley, F. Fraioli, M. Al Bakir, E. Grönroos, F. Zambrana, R. Endozo, W. L. Bi, F. M. Fennessy, N. Sponer, D. Johnson, J. Laycock, S. Shafi, J. Czyzewska-Khan, A. Rowan, T. Chambers, N. Matthews, S. Turajlic, C. Hiley, S. M. Lee, M. D. Forster, T. Ahmad, M. Falzon, E. Borg, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, D. Hafez, A. Naik, A. Ganguly, S. Kareht, R. Shah, L. Joseph, A. M. Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, D. Oukrif, A. U. Akarca, J. A. Hartley, H. L. Lowe, S. Lock, N. Iles, H. Bell, Y. Ngai, G. Elgar, Z. Szallasi, R. F. Schwarz, J. Herrero, A. Stewart, S. A. Quezada, K. S. Peggs, P. Van Loo, C. Dive, C. J. Lin, M. Rabinowitz, H. J. W. L. Aerts, A. Hackshaw, J. A. Shaw, B. G. Zimmermann; The TRACERx consortium; The PEACE consortium, C. Swanton, Phylogenetic ctDNA analysis depicts early-stage lung cancer evolution. *Nature* **545**, 446–451 (2017).
3. A. Zviran, R. C. Schulman, M. Shah, S. T. K. Hill, S. Deochand, C. C. Khamnei, D. Maloney, K. Patel, W. Liao, A. J. Widman, P. Wong, M. K. Callahan, G. Ha, S. Reed, D. Rotem, D. Frederick, T. Sharova, B. Miao, T. Kim, G. Gydush, J. Rhoades, K. Y. Huang, N. D. Omans, P. O. Bolan, A. H. Lipsky, C. Ang, M. Malbari, C. F. Spinelli, S. Kazancioglu, A. M. Runnels, S. Fennessey, C. Stolte, F. Gaiti, G. G. Inghirami, V. Adalsteinsson, B. Houck-Loomis, J. Ishii, J. D. Wolchok, G. Boland, N. Robine, N. K. Altorki, D. A. Landau, Genome-wide cell-free DNA mutational integration enables ultra-sensitive cancer monitoring. *Nat. Med.* **26**, 1114–1124 (2020).
4. J. C. M. Wan, K. Heider, D. Gale, S. Murphy, E. Fisher, F. Mouliere, A. Ruiz-Valdepenas, A. Santonja, J. Morris, D. Chandrananda, A. Marshall, A. B. Gill, P. Y. Chan, E. Barker, G. Young, W. N. Cooper, I. Hudecova, F. Marass, R. Mair, K. M. Brindle, G. D. Stewart, J. E. Abraham, C. Caldas, D. M. Rassl, R. C. Rintoul, C. Alifrangis, M. R. Middleton, F. A. Gallagher, C. Parkinson, A. Durrani, U. McDermott, C. G. Smith, C. Massie, P. G. Corrie, N. Rosenfeld, ctDNA monitoring using patient-specific sequencing and integration of variant reads. *Sci. Transl. Med.* **12**, eaaz8084 (2020).
5. C. Bettegowda, M. Sausen, R. J. Leary, I. Kinde, Y. Wang, N. Agrawal, B. R. Bartlett, H. Wang, B. Luber, R. M. Alani, E. S. Antonarakis, N. S. Azad, A. Bardelli, H. Brem, J. L. Cameron, C. C. Lee, L. A. Fecher, G. L. Gallia, P. Gibbs, D. Le, R. L. Giuntoli, M. Goggins, M. D. Hogarty, M. Holdhoff, S.-M. Hong, Y. Jiao, H. H. Juhl, J. J. Kim, G. Siravegna, D. A. Laheru, C. Lauricella, M. Lim, E. J. Lipson, S. K. N. Marie, G. J. Netto, K. S. Oliner, A. Olivi, L. Olsson, G. J. Riggins, A. Sartore-Bianchi, K. Schmidt, L.-M. Shih, S. M. Oba-Shinjo, S. Siena, D. Theodorescu, J. Tie, T. T. Harkins, S. Veronese, T.-L. Wang, J. D. Weingart, C. L. Wolfgang, L. D. Wood, D. Xing, R. H. Hruban, J. Wu, P. J. Allen, C. M. Schmidt, M. A. Choti, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, N. Papadopoulos, L. A. Diaz Jr., Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
6. J. D. Cohen, L. Li, Y. Wang, C. Thoburn, B. Afsari, L. Danilova, C. Douville, A. A. Javed, F. Wong, A. Mattox, R. H. Hruban, C. L. Wolfgang, M. G. Goggins, M. Dal Molin, T.-L. Wang, R. Roden, A. P. Klein, J. Ptak, L. Dobbyn, J. Schaefer, N. Silliman, M. Popoli, J. T. Vogelstein, J. D. Browne, R. E. Schoen, R. E. Brand, J. Tie, P. Gibbs, H.-L. Wong, A. S. Mansfield, J. Jen, S. M. Hanash, M. Falconi, P. J. Allen, S. Zhou, C. Bettegowda, L. A. Diaz, C. Tomasetti, K. W. Kinzler, B. Vogelstein, A. M. Lennon, N. Papadopoulos, Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* **359**, 926–930 (2018).
7. Y. van der Pol, F. Mouliere, Toward the early detection of cancer by decoding the epigenetic and environmental fingerprints of cell-free DNA. *Cancer Cell* **36**, 350–368 (2019).
8. M. C. Liu, G. R. Oxnard, E. A. Klein, C. Swanton, M. V. Seiden; CCGA Consortium, Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* **31**, 745–759 (2020).

9. S. Y. Shen, R. Singhania, G. Fehringer, A. Chakravarthy, M. H. A. A. Roehrl, D. Chadwick, P. C. Zuzarte, A. Borgida, T. T. Wang, T. Li, O. Kis, Z. Zhao, A. Spreafico, T. da S. Medina, Y. Wang, D. Roulois, I. Ettayebi, Z. Chen, S. Chow, T. Murphy, A. Arruda, G. M. O'Kane, J. Liu, M. Mansour, J. D. McPherson, C. O'Brien, N. Leighl, P. L. Bedard, N. Fleshner, G. Liu, M. D. Minden, S. Gallinger, A. Goldenberg, T. J. Pugh, M. M. Hoffman, S. V. Bratman, R. J. Hung, D. D. De Carvalho, Sensitive tumour detection and classification using plasma cell-free DNA methylomes. *Nature* **563**, 579–583 (2018).
10. K. C. Allen Chan, P. Jiang, C. W. M. Chan, K. Sun, J. Wong, E. P. Hui, S. L. Chan, W. C. Chan, D. S. C. Hui, S. S. M. Ng, H. L. Y. Chan, C. S. C. Wong, B. B. Y. Ma, A. T. C. Chan, P. B. S. Lai, H. Sun, R. W. K. Chiu, Y. M. Dennis Lo, Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 18761–18768 (2013).
11. K. Sun, P. Jiang, K. C. A. Chan, J. Wong, Y. K. Y. Cheng, R. H. S. Liang, W. K. Chan, E. S. K. Ma, S. L. Chan, S. H. Cheng, R. W. Y. Chan, Y. K. Tong, S. S. M. Ng, R. S. M. Wong, D. S. C. Hui, T. N. Leung, T. Y. Leung, P. B. S. Lai, R. W. K. Chiu, Y. M. D. Lo, Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc. Natl. Acad. Sci. U.S.A.* **112**, E5503–E5512 (2015).
12. S. Guo, D. Diep, N. Plongthongkum, H. L. Fung, K. Zhang, K. Zhang, Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635–642 (2017).
13. J. Moss, J. Magenheim, D. Neiman, H. Zemmour, N. Loyfer, A. Korach, Y. Samet, M. Maoz, H. Druid, P. Arner, K.-Y. Fu, E. Kiss, K. L. Spalding, G. Landesberg, A. Zick, A. Grinshpun, A. M. J. J. Shapiro, M. Grompe, A. D. Wittenberg, B. Glaser, R. Shemer, T. Kaplan, Y. Dor, Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).
14. A. P. Cheng, P. Burnham, J. R. Lee, M. P. Cheng, M. Suthanthiran, D. Dadhania, I. De Vlaminck, A cell-free DNA metagenomic sequencing assay that integrates the host injury response to infection. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 18738–18744 (2019).
15. X. Chen, J. Gole, A. Gore, Q. He, M. Lu, J. Min, Z. Yuan, X. Yang, Y. Jiang, T. Zhang, C. Suo, X. Li, L. Cheng, Z. Zhang, H. Niu, Z. Li, Z. Xie, H. Shi, X. Zhang, M. Fan, X. Wang, Y. Yang, J. Dang, C. McConnell, J. Zhang, J. Wang, S. Yu, W. Ye, Y. Gao, K. Zhang, R. Liu, L. Jin, Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat. Commun.* **11**, 3475 (2020).
16. J. J. Jaworski, R. D. Morgan, S. Sivakumar, Circulating cell-free tumour DNA for early detection of pancreatic cancer. *Cancers* **12**, 3704 (2020).
17. Y. Liu, P. Siejka-Zielińska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Böckler, C.-X. Song, Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).
18. Y. M. D. Lo, D. S. C. Han, P. Jiang, R. W. K. Chiu, Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* **372**, eaaw3616 (2021).
19. T. F. Greten, F. Papendorf, J. S. Bleck, T. Kirchhoff, T. Wohlberedt, S. Kubicka, J. Klempnauer, M. Galanski, M. P. Manns, Survival rate in patients with hepatocellular carcinoma: A retrospective analysis of 389 patients. *Br. J. Cancer* **92**, 1862–1868 (2005).
20. A. P. Stark, G. D. Sacks, M. M. Rochefort, T. R. Donahue, H. A. Reber, J. S. Tomlinson, D. W. Dawson, G. Eibl, O. J. Hines, Long-term survival in patients with pancreatic ductal adenocarcinoma. *Surgery* **159**, 1520–1527 (2016).
21. S. Pascual, C. Miralles, J. M. Bernabé, J. Irurzun, M. Planells, Surveillance and diagnosis of hepatocellular carcinoma: A systematic review. *World J. Clin. Cases* **7**, 2269–2286 (2019).
22. F. Trevisani, P. E. D'Intino, A. M. Morselli-Labate, G. Mazzella, E. Accogli, P. Caraceni, M. Domenicali, S. De Notariis, E. Roda, M. Bernardi, Serum α-fetoprotein for diagnosis of hepatocellular carcinoma in patients with chronic liver disease: Influence of HBsAg and anti-HCV status. *J. Hepatol.* **34**, 570–575 (2001).
23. S. Ngamruengphong, M. I. Canto, Screening for pancreatic cancer. *Surg. Clin. North Am.* **96**, 1223–1233 (2016).
24. W. Li, Q. Li, S. Kang, M. Same, Y. Zhou, C. Sun, C.-C. Liu, L. Matsuoka, L. Sher, W. H. Wong, F. Alber, X. J. Zhou, CancerDetector: Ultrasensitive and non-invasive cancer detection at the resolution of individual reads using cell-free DNA methylation sequencing data. *Nucleic Acids Res.* **46**, e89 (2018).
25. G. Ramakrishna, A. Rastogi, N. Trehanpati, B. Sen, R. Khosla, S. K. Sarin, From cirrhosis to hepatocellular carcinoma: New molecular insights on inflammation and cellular senescence. *Liver Cancer* **2**, 367–383 (2013).
26. D. Yadav, A. B. Lowenfels, The epidemiology of pancreatitis and pancreatic cancer. *Gastroenterology* **144**, 1252–1261 (2013).
27. W. Hartwig, O. Strobel, U. Hinz, S. Fritz, T. Hackert, C. Roth, M. W. Büchler, J. Werner, CA19-9 in potentially resectable pancreatic cancer: Perspective to adjust surgical and perioperative therapy. *Ann. Surg. Oncol.* **20**, 2188–2196 (2013).
28. J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
29. F. Grise, A. Bidaud, V. Moreau, Rho GTPases in hepatocellular carcinoma. *Biochim. Biophys. Acta* **1795**, 137–151 (2009).
30. H. Bi, Y. Zhang, S. Wang, W. Fang, W. He, L. Yin, Y. Xue, Z. Cheng, M. Yang, J. Shen, Interleukin-8 promotes cell migration via CXCR1 and CXCR2 in liver cancer. *Oncol. Lett.* **18**, 4176–4184 (2019).
31. C. M. Wong, J. M. F. Lee, Y. P. Ching, D. Y. Jin, O.-L. Ng, Genetic and epigenetic alterations of DLC-1 gene in hepatocellular carcinoma. *Cancer Res.* **63**, 7646–7651 (2003).
32. A. J. Gore, S. L. Deitz, L. R. Palam, K. E. Craven, M. Korc, Pancreatic cancer–associated retinoblastoma 1 dysfunction enables TGF-β to promote proliferation. *J. Clin. Invest.* **124**, 338–352 (2014).
33. M. S. Alam, M. M. Gaida, F. Bergmann, F. Lasitschka, T. Giese, N. A. Giese, T. Hackert, U. Hinz, S. P. Hussain, S. V. Kozlov, J. D. Ashwell, Selective inhibition of the p38 alternative activation pathway in infiltrating T cells inhibits pancreatic cancer progression. *Nat. Med.* **21**, 1337–1343 (2015).
34. E. A. Price, K. Kolkiewicz, R. Patel, S. Hashim, E. Karaa, I. Scheimberg, M. S. Sagoo, M. A. Reddy, Z. Onadim, Detection and reporting of RB1 promoter hypermethylation in diagnostic screening. *Ophthalmic Genet.* **39**, 526–531 (2018).
35. J. K. Park, J. C. Henry, J. Jiang, C. Esau, Y. Gusev, M. R. Lerner, R. G. Postier, D. J. Brackett, T. D. Schmittgen, MiR-132 and miR-212 are increased in pancreatic cancer and target the retinoblastoma tumor suppressor. *Biochem. Biophys. Res. Commun.* **406**, 518–523 (2011).
36. M. J. Ziller, K. D. Hansen, A. Meissner, M. J. Aryee, Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. *Nat. Methods* **12**, 230–232 (2015).
37. C. Luo, P. Hajkova, J. R. Ecker, Dynamic DNA methylation: In the right place at the right time. *Science* **361**, 1336–1340 (2018).
38. F. Mouliere, D. Chandrananda, A. M. Piskorz, E. K. Moore, J. Morris, L. B. Ahlborn, R. Mair, T. Goranova, F. Marass, K. Heider, J. C. M. M. Wan, A. Supernat, I. Hudecova, I. Gounaris, S. Ros, M. Jimenez-Linan, J. Garcia-Corbacho, K. Patel, O. Østrup, S. Murphy, M. D. Eldridge, D. Gale, G. D. Stewart, J. Burge, W. N. Cooper, M. S. Van Der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, M. S. Van Der Heijden, C. E. Massie, C. Watts, P. Corrie, S. Pacey, K. M. Brindle, R. D. Baird, M. Mau-Sørensen, C. A. Parkinson, C. G. Smith, J. D. Brenton, N. Rosenfeld, Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10**, eaat4921 (2018).
39. S. Cristiano, A. Leal, J. Phallen, J. Fiksel, V. Adleff, D. C. Bruhm, S. Ø. Jensen, J. E. Medina, C. Hruban, J. R. White, D. N. Palsgrove, N. Niknafs, V. Anagnostou, P. Forde, J. Naidoo, K. Marrone, J. Brahmer, B. D. Woodward, H. Husain, K. L. van Rooijen, M.-B. W. Ørntoft, A. H. Madsen, C. J. H. van de Velde, M. Verheij, A. Cats, C. J. A. Punt, G. R. Vink, N. C. T. van Grieken, M. Koopman, R. J. A. Fijneman, J. S. Johansen, H. J. Nielsen, G. A. Meijer, C. L. Andersen, R. B. Scharpf, V. E. Velculescu, Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* **570**, 385–389 (2019).
40. R. Wang, Q. Sun, P. Wang, M. Liu, S. Xiong, J. Luo, H. Huang, Q. Du, D. A. Geller, B. Cheng, Notch and Wnt/β-catenin signaling pathway play important roles in activating liver cancer stem cells. *Oncotarget* **7**, 5754–5768 (2016).
41. T. L. Fitzgerald, K. Lertpiriyapong, L. Cocco, A. M. Martelli, M. Libra, S. Candido, G. Montalto, M. Cervello, L. Steelman, S. L. Abrams, J. A. McCubrey, Roles of EGFR and KRAS and their downstream signaling pathways in pancreatic cancer and pancreatic cancer stem cells. *Adv. Biol. Regul.* **59**, 65–81 (2015).
42. F. Erger, D. Nörling, D. Borchert, E. Leenen, S. Habbig, M. S. Wiesener, M. P. Bartram, A. Wenzel, C. Becker, M. R. Toliat, P. Nürnberg, B. B. Beck, J. Altmüller, *cfNOMe*—A single assay for comprehensive epigenetic analyses of cell-free DNA. *Genome Med.* **12**, 54 (2020).
43. M. W. Snyder, M. Kircher, A. J. Hill, R. M. Daza, J. Shendure, Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
44. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595 (2010).
45. K. R. Rosenbloom, J. Armstrong, G. P. Barber, J. Casper, H. Clawson, M. Diekhans, T. R. Dreszer, P. A. Fujita, L. Guruvadoo, M. Haeussler, R. A. Harte, S. Heitner, G. Hickey, A. S. Hinrichs, R. Hubley, D. Karolchik, K. Learned, B. T. Lee, C. H. Li, K. H. Miga, N. Nguyen, B. Paten, B. J. Raney, A. F. A. Smit, M. L. Speir, A. S. Zweig, D. Haussler, R. M. Kuhn, W. J. Kent, The UCSC Genome Browser database: 2015 update. *Nucleic Acids Res.* **43**, D670–D681 (2015).
46. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE blacklist: Identification of problematic regions of the genome. *Sci. Rep.* **9**, 9354 (2019).
47. F. Krueger, S. R. Andrews, Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
48. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin; 1000 Genome Project Data Processing Subgroup, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
49. S. Lê, J. Josse, F. Husson, FactoMineR : An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

50. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
51. D. R. Zerbino, S. P. Wilder, N. Johnson, T. Juettemann, P. R. Flicek, The ensembl regulatory build. *Genome Biol.* **16**, 56 (2015).
52. S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, D. Cohen, GeneHancer: Genome-wide integration of enhancers and target genes in GeneCards. *Database* **2017**, bax028 (2017).
53. M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, A. Ma'ayan, Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
54. W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, A. D. Haussler, The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).
55. P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt; SciPy 1.0 Contributors, SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261–272 (2020).
56. I. Scheinin, D. Sie, H. Bengtsson, M. A. Van De Wiel, A. B. Olshen, H. F. Van Thuijl, H. F. Van Essen, P. P. Eijk, F. Rustenburg, G. A. Meijer, J. C. Reijneveld, P. Wesseling, D. Pinkel, D. G. Albertson, B. Ylstra, DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* **24**, 2022–2032 (2014).
57. M. Kuhn, Building predictive models in R using the caret package. *J. Stat. Softw.* **28**, 1–26 (2008).

Acknowledgments: We would like to acknowledge M. Youdell for sample collection and M. Muers for editing the manuscript. Computation used the Oxford Biomedical Research Computing (BMRC) facility, a joint development between the Wellcome Centre for Human Genetics and the Big Data Institute supported by Health Data Research UK and the NIHR Oxford Biomedical Research Centre. We would like to thank Drs. Xianghong Jasmine Zhou and Wenyuan Li for sharing WGBS data EGAD00001004317. **Funding:** This work was funded by the Ludwig Institute for Cancer Research (C.-X.S. and B.S.-B.), Cancer Research UK (CRUK) (C63763/A26394 and C63763/A27122 to C.-X.S.), and the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC) (C.-X.S., B.S.-B., and E.L.C.). C.-X.S. laboratory is also supported by Emerson Collective. E.B. acknowledges the Oxford NIHR BRC for support and is an NIHR Senior Investigator. E.B., C.-X.S., E.L.C., B.S.-B., H.R., and D.M. also acknowledge CRUK program grant (C30358/A29725) for HCC early cancer detection (DeLIVER) for support. H.R. and D.M. are supported by CRUK program grant C18342/A23390 and CRUK HUNTER Accelerator (C9380/A26813). F.J. is supported by a CRUK Oxford Centre Prize DPhil Studentship (C2195/A27450). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. **Author contributions:** P.S.-Z.: Conceptualization, methodology, investigation, formal analysis, data curation, writing—original draft, writing—review and editing, and visualization. J.C.: Formal analysis, data curation, writing—original draft, writing—review and editing, and visualization. F.J.: Formal analysis, data curation, writing—original draft, and writing—review and editing. Y.L.: Investigation. Z.S.: Resources. S.R.: Resources. M.S.: Resources. L.P.: Resources and data curation. M.V.-M.: Resources and data curation. E.L.C.: Resources and data curation. N.B.: Resources. B.S.-B.: Supervision and writing—review and editing. P.F.P.: Supervision and writing—review and editing. D.M.: Resources and writing—review and editing. H.R.: Resources, data curation, and writing—review and editing E.B.: Conceptualization, resources, writing—review and editing, supervision, and funding acquisition. S.S.: Conceptualization, resources, data curation, writing—review and editing, and supervision. C.-X.S.: Conceptualization, resources, writing—review and editing, supervision, and funding acquisition. **Competing interests:** C.-X.S. and Y.L. are inventors on a patent application related to this work filed by Ludwig Institute for Cancer Research (PCT/US2019/012627, filed on 8 January 2018, published on 11 July 2019), which have been licensed to Exact Sciences Innovation. C.-X.S. is a consultant to Exact Sciences Innovation, and Y.L. is an employee at Exact Sciences Innovation. S.S. has salary and research funding from Bristol Myers Squibb. The authors declare no other competing interests. **Data and materials availability:** Raw cfTAPS sequencing data have been deposited at the European Genome-phenome Archive (EGA), which is hosted at the EBI and the CRG, under accession number EGAS00001004962. The analysis scripts are available at https://gitlab.com/cf_taps/cftaps_paper/. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 12 February 2021
Accepted 14 July 2021
Published 1 September 2021
10.1126/sciadv.abh0534

Citation: P. Siejka-Zielińska, J. Cheng, F. Jackson, Y. Liu, Z. Soonawalla, S. Reddy, M. Silva, L. Puta, M. V. McCain, E. L. Culver, N. Bekkali, B. Schuster-Böckler, P. F. Palamara, D. Mann, H. Reeves, E. Barnes, S. Sivakumar, C.-X. Song, Cell-free DNA TAPS provides multimodal information for early cancer detection. *Sci. Adv.* **7**, eabh0534 (2021).

ScienceAdvances

Cell-free DNA TAPS provides multimodal information for early cancer detection

Paulina Siejka-Zieli#skaJingfei ChengFelix JacksonYibin LiuZahir SoonawallaSrikanth ReddyMichael SilvaLuminita PutaMisti Vanette McCainEmma L. CulverNoor BekkaliBenjamin Schuster-BöcklerPier Francesco PalamaraDerek MannHelen ReevesEleanor BarnesShivan SivakumarChun-Xiao Song

Sci. Adv., 7 (36), eabh0534. • DOI: 10.1126/sciadv.abh0534

Science Advances (ISSN) is published by the American Association for the Advancement of Science. 1200 New York Avenue NW, Washington, DC 20005. The title *Science Advances* is a registered trademark of AAAS.