Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing

Efrat Katsman[1,*], Shari Orlanski[1,*], Filippo Martignano[2,*], Ilana Fox-Fisher[1], Ruth Shemer[1], Yuval Dor[1], Aviad Zick[3], Amir Eden[4], Iacopo Petrini[5], Silvestro G. Conticello[2,6,#], Benjamin P. Berman[1,#]

1. Department of Developmental Biology and Cancer Research, Institute for Medical Research Israel-Canada, Faculty of Medicine, The Hebrew University of Jerusalem, Israel

2. Core Research Laboratory, ISPRO, Florence, Italy.

3. Department of Oncology, Hadassah Medical Center, Faculty of Medicine, Hebrew University of Jerusalem, Israel

4. Department of Cell and Developmental Biology, The Alexander Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Israel

5. Unit of Respiratory Medicine, Department of Critical Area and Surgical, Medical and Molecular Pathology, University Hospital of Pisa, Pisa, Italy.

6. Institute of Clinical Physiology, National Research Council, Pisa, Italy.

* Equal contribution

Jointly supervised project

Correspondence to: s.conticello@ispro.toscana.it, ben.berman@mail.huji.ac.il

Abstract

The Oxford Nanopore (ONT) platform provides portable and rapid genome sequencing, and its ability to natively profile DNA methylation without complex sample processing is attractive for clinical sequencing. We recently demonstrated ONT shallow whole-genome sequencing to detect copy number alterations (CNA) from the circulating tumor DNA (ctDNA) of cancer patients. Here, we show that cell-type and cancer-specific methylation changes can also be detected, as well as cancer-associated fragmentation signatures. This feasibility study suggests that ONT shallow WGS could be a powerful tool for liquid biopsy, especially real-time medical applications.

Introduction

Circulating cell-free DNA (cfDNA) can reveal informative features of its tissue of origin including somatic genome alterations, DNA modifications, and cell-type specific fragmentation patterns (*1*). DNA methylation is a promising cfDNA biomarker and is in widespread testing as a cancer screening tool (*2*). DNA methylation can also be used to detect turnover of damaged cells in time-sensitive conditions such as myocardial infarction, sepsis, and COVID-19 (*3*–*6*). These studies used bisulfite-based approaches to profile methylation, but alternative approaches include immunoprecipitation (*7*) and enzymatic conversion (*8*) techniques.

Accurate calling of DNA methylation from native DNA Oxford Nanopore (ONT) sequencing has matured and now produces single base-pair resolution results highly similar to bisulfite sequencing (*9*, *10*). The ONT platform is portable and has a low cost of setup, and a rapid sequencing workflow that can enable real-time medicine (*11*, *12*). Native ONT sequencing requires no complex sample processing steps and no PCR amplification, making it attractive for clinical tests. Bisulfite approaches in particular involve significant degradation and loss of input material. For these reasons, whole-genome sequencing (WGS) using the ONT platform is appealing relative to other whole-genome approaches.

Due to their low cost, targeted bisulfite PCR (including multiplexed NGS versions (*13*)) are also popular for clinical methylation sequencing, and these would not require the native modification calling capability of Nanopore. However, shallow whole-genome approaches that capture multiple genomic features could potentially be more informative, especially with regard to the course of the disease (*8*, *14*, *15*).

ONT has primarily been used for long-read sequencing, but recent work by our group and others has shown that it can be adapted for short fragments without additional processing steps (*16*–*18*). As an added benefit, the ability to capture much longer cfDNA fragments than short-read sequencing may lead to new discoveries or biomarkers, as was demonstrated recently in the case of longer fragments during pregnancy (*19*).

Here, we perform a feasibility study of ONT sequencing for circulating tumor DNA (ctDNA) detection by comparing methylation and several fragmentation features to matched Illumina samples and comparable Illumina-based datasets. While our sample size is too small to determine the limits of sensitivity and specificity, we find that both cell-type specific and cancer-specific features can be reliably detected in most of our samples. The results provide confidence for pursuing this approach in larger studies.

Results

Estimating cell type fractions from cfNano

We first performed cell type deconvolution of healthy plasma cfDNA using DNA methylation data from either published WGBS datasets or our cfNano samples. For the external WGBS datasets, we used the methylation fractions (beta values) that were provided in the published data files. For our cfNano, we performed direct modification calling using the Megalodon software provided by ONT (https://github.com/nanoporetech/megalodon). To perform deconvolution, we used 1,000-2,000 marker CpGs per cell type based on a previously published atlas of purified cell types ("MethAtlas", (*5*, *13*)), and estimated cell type fractions using Non-Negative Least Squares (NNLS) regression as described in (*5*). In order to better understand the impact of the relatively low sequencing depth of our cfNano samples (~0.2x genome coverage), we first performed deconvolution of all samples using downsampling experiments starting with full sequence depth down to 0.0001x genome coverage (Figure 1A and Supplementary Figures 1-3). Healthy plasma WGBS samples were taken from a recent study of 50-100x genomic coverage ((*13*), Figure 1A left "Fox-Fisher" samples), and another WGBS study with 0.5-1x coverage ((*20*) Figure 1A middle "Nguyen" samples). Finally, healthy cfNano samples were analyzed (Figure 1A right "this study"). From full depth down to 0.2x (about 2.5M aligned fragments), all samples were dominated by the expected cell types: monocytes, lymphocytes, megakaryocytes, neutrophils/granulocytes, and sometimes hepatocytes (*5*). Cell type proportions became significantly degraded at 0.05x coverage and below (corresponding to less than 700,000 aligned fragments). The common cell types were consistently found across the 23 healthy individuals in the Fox-Fisher dataset, the 3 healthy individuals in the Nguyen dataset, and the 7 healthy individuals in the cfNano dataset, both at full depth (Figure 1B) and when downsampled to 0.2x depth (Figure 1C). The same was found when cell type groups such as lymphocytes were broken down into the 25 individual types (Supplementary Figure 4A-C). Notably, a slight epithelial fraction was identified in some of the Fox-Fisher samples at 0.2x which did not appear at full 80x depth, suggesting a small but measurable amount of noise at the 0.2x coverage level.

The healthy cfNano individuals were divided into two groups based on source site, with one being collected and sequenced in Italy ("BC" samples) and one in Israel ("HU" samples). Despite the HU samples being lower coverage (two were between 0.10-0.15x depth), they displayed relatively similar cell type proportions (Figure 1B-C and Supplementary Figure 3).

In addition to healthy individuals, the Nguyen WGBS dataset and our cfNano dataset also contained individuals being treated for lung adenocarcinoma (LuAd), marked as "LuAd" in Figure 1B-C. In the Nguyen WGBS study, samples were collected at the time of acquired resistance to EGFR-inhibitors, and were divided into those that acquired resistance mutations in EGFR itself (labeled "on" for on-target) vs. those that acquired amplifications in alternative oncogenes MET/ERBB2 (labeled "off" for off-target). The epithelial cell fraction was much higher in the on-target patients, while the off-target patients had very low or no epithelial fraction (Figure 1B), consistent with the absence of CNAs in the off-target samples in the original study (*20*). The 6 LuAd samples in our cfNano study had similarly high epithelial fraction (Figure 1B), which was significantly higher than in the healthy patients (p=0.004). In all WGBS and cfNano samples, full depth results were highly similar to 0.2x downsampled results (Figure 1C and

Supplementary Figures 1-3). Interestingly, while the Nguyen et al. study interpreted the normal-like methylation levels of the “off-target” tumors as a difference in cancer methylation patterns, our deconvolution results strongly suggest that it is due to the relative amount of cancer DNA circulating in the blood.

The fraction of cancer cells in cfDNA (“tumor fraction”) can be estimated from somatic copy number alterations (CNAs) using the ichorCNA tool (*21*), for cancers that contain a sufficient degree of aneuploidy. We estimated tumor fraction for our cfNano samples and four matched Illumina WGS samples from LuAd patients (Figure 1D, Supplementary Table 1, and Supplementary Data File 1). While the Illumina samples were sequenced at significantly higher depth (median 1.3x), the tumor fraction estimates were highly similar between cfNano and Illumina sequencing (Figure 1E). Interestingly, the ichorCNA tumor fractions were more similar to the high-depth Illumina samples than the Illumina samples themselves were when downsampled to the same depth as the cfNano samples (Supplementary Figure 5A).

To compare ichorCNA tumor fraction estimates to methylation-based estimates, we designed a “two-component” deconvolution method based on NNLS regression that used 2,253 CpGs with differential methylation between sorted lung epithelia and healthy plasma. This was based on the same array-based MethAtlas samples from (*5*) as the full deconvolution (Figure 1F). 330-1,526 of these CpGs were covered by each cfNano sample (Supplementary Table 1), which were the CpGs used for NNLS deconvolution. These DNA methylation based estimates of lung fraction and the ichorCNA estimates of cancer cell fraction were largely in agreement (Figure 1F, bottom), with all the six LuAd samples having significantly higher lung fraction compared to the seven healthy plasma samples (p=0.003). Two LuAd cases were markedly higher in the methylation-based than the CNA-based estimate (BC09 and BC10). While we have no independent data to determine which was the more accurate estimate, we hypothesize that the discrepancy may be due to either whole-genome doubling (WGD) events that are not detected by ichorCNA (WGD occurs in 297/503 or 59% of LuAd tumors from the TCGA project (*22*)), or damage to normal lung cells surrounding the tumor which die and shed their DNA into circulation (*23*).

To verify the robustness of methylation-based deconvolution, we used a mutually exclusive set of 13,770 CpGs that could distinguish TCGA LuAd tumors from healthy plasma, but were not found in the normal lung epithelia set (Figure 1H). Before applying the NNLS regression, we corrected the methylation levels of the TCGA LuAd samples to correct for their degree of non-cancer contamination (“purity correction”), since most TCGA LuAd samples contain a significant fraction of leukocytes. After this correction, the tumor fraction estimates of our cfNano samples were highly similar to those based on normal-lung specific CpGs (Figure 1H, bottom), despite the fact that the two CpG sets were completely non-overlapping. One HU healthy sample (HU005.10) had a higher lung fraction estimate than one of the cancer samples, possibly because this was the cfNano sample with the lowest sequencing coverage (0.11x). However, the methylation-based tumor fraction was still significantly higher in the LuAd samples than healthy controls (p=0.004).

We performed all deconvolution analysis using a second, and older, base modification caller (DeepSignal, (*24*)). While Megalodon called 10-20% more CpGs, the majority of CpGs called were in common between the two methods and had identical methylation states (Supplementary

Figure 6A). Both the full cell-type deconvolution (Supplementary Figure 6B) and the two-component deconvolution (Supplementary Figure 6C) were highly similar between the two callers.

Genomic context of DNA methylation changes detected using cfNano

The deconvolution analysis above was based on unannotated differentially methylated regions. In order to investigate the genomic context of lung cancer specific DNA methylation, we analyzed one hypomethylation feature associated with cell of origin (lineage-specific transcription factor binding sites) and one associated specifically with transformation (global hypomethylation). For the TFBS analysis, we identified 5,974 predicted TFBS that were specific to lung epithelia based on a single-cell ATAC-seq atlas of open chromatin within lung and other primary human tissues (*25*). In that study, adult lung tissues from multiple donors contained a strong cluster of lung pneumocyte-specific open chromatin regions ("Pal" cluster). This cluster was most strongly enriched for the binding motif for the transcription factor NKX2-1, which is a master regulatory transcription factor in this cell type (*26*). NKX2-1 activity is also known to be highly restricted to this cell type (*27*), and NKX2-1 binding sites were also the most enriched within lung adenocarcinoma ATAC-seq sites in an independent study (*28*). Because open chromatin regions are almost universally hypomethylated, we hypothesized that the 5,974 predicted NKX2-1 TFBS in lung pneumocytes would be specifically hypomethylated in healthy lung tissues and in lung tumors. We confirmed this using WGBS data from TCGA (*29*) (Supplementary Figure 7A).

We next plotted plasma cfDNA methylation levels at these same predicted NKX2-1 sites from the published Illumina WGBS studies and our cfNano study (Figure 2A). In healthy samples from three WGBS studies and our own cfNano samples, NKX2-1 sites were fully methylated. In contrast, the LuAd samples from both the Nguyen et al. WGBS study (Figure 2A, middle) and our cfNano study (Figure 2A, right) had substantial demethylation. In both studies demethylation could only be observed in the higher tumor fraction samples ("on-target" samples in the WGBS study, and samples with ichorCNA TF>0.15 in the cfNano study). As a negative control, we selected predicted TFBS from a cell type not expected to be found either in healthy plasma or LuAd. We used the adrenal cortical cluster ("Adc" cluster) from (*25*), which was highly enriched for the KLF5 binding motif. These sites were fully methylated in plasma samples from both healthy and LuAd individuals (Supplementary Figure 7B). cfNano profiles were nearly identical using DeepSignal methylation calling (Supplementary Figure 7C).

Global DNA hypomethylation is one of the hallmarks of the cancer epigenome. It has long been proposed as a general marker for circulating tumor DNA (*30*), and this was recently verified for lung cancer using shallow plasma cfDNA WGBS (*20*). We have shown that this "global" hypomethylation is not completely global, and occurs preferentially within large domains called Partially Methylated Domains (PMDs) and specifically at CpGs with a local sequence context termed "solo-WCGW" (*29*). We replotted WGBS methylation data from TCGA normal lung and lung tumor tissues, showing a typical chromosome arm where strong hypomethylation occurs within the PMD regions identified in (*29*) in the cancer samples (Figure 2B, top). In our cfNano samples, strong hypomethylation was also found exclusively in the cancer samples in the same

PMD regions (Figure 2B, bottom). To quantify this genome-wide, we plotted the methylation change (relative to the sample-specific whole-genome average) of PMD solo-WCGW CpGs within each 10 Mbp genomic bin that overlapped a common PMD region from (*29*) (Figure 2C). As expected, five of the six cancer samples were significantly hypomethylated relative to the healthy controls ($p<0.0001$). Overall, there was significantly more hypomethylation across all cancer sample bins (mean=-0.10,SD=0.07) than across all healthy sample bins (mean=-0.05,SD=0.04), corresponding to a p-value<2.2e-16 by one-sided Wilcoxon test. In the final LuAd case (BC11), no PMD hypomethylation could be detected (Figure 2C). This is not surprising given the high degree of variability associated with global hypomethylation in cancer (*29*), a process that is not entirely understood but is affected both passively through mitotic divisions as well as actively by dysregulation of several chromatin modifiers (*31*, *32*).

Reasoning that copy number altered regions would have skewed proportions of tumor-derived DNA and thus different levels of PMD hypomethylation, we divided the PMD bins based on the copy number status of each sample from ichorCNA. In four of the five cases with significant hypomethylation overall, the amplified bins had significantly more hypomethylation than diploid regions (Figure 2D). In the one remaining case (BC09), there were not enough PMDs with CNAs for an accurate measurement. Conversely, deleted showed significantly less hypomethylation than diploid regions, although this trend only reached statistical significance in two cases. In the future, the combined analysis of CNAs and global hypomethylation may provide a stronger cancer-specific signal than each feature alone. PMD hypomethylation profiles were nearly identical using DeepSignal methylation calling (Supplementary Figure 7D-F).

cfNano preserves nucleosome positioning signal

Cell-free DNA circulates primarily as mononucleosomal fragments, and mapping the positions of these mononucleosomes can be used to identify cell-type specific positioning (reviewed in (*1*)). CTCF binding sites provide a good test of whether these signals are detectable, since they eject a central nucleosome and position 10 phased nucleosomes on either side of their binding site (*33*) (Figure 3A). Around a set of 9,780 CTCF binding sites, cfNano mononucleosome locations recapitulated this expected pattern (Figure 3B, top), which was identical to the pattern based on matched Illumina WGS of greater sequence depth (Figure 3B, bottom). These were also identical when both cfNano and Illumina libraries were downsampled to an equal number of 2M fragments (Figure 3C). The CTCF binding site also demethylates CpGs located approximately 200bp of on either side (*33*), and this DNA methylation pattern was also recapitulated in our cfNano samples (Supplementary Figure 8).

We tried the same mononucleosome mapping approach for the 5,974 lung-specific NKX2-1 TFBS from Figure 2A. While the demethylation signal was detectable (Figure 2A), we could not detect any mononucleosome positioning signal (data not shown). Lung-specific nucleosome positions would only be present on a fraction of the fragments, so the signal from these fragments may be masked by those from non-lung cell types. But given that the inherent nucleosome positioning information is present (as shown by the CTCF example), more advanced normalization and quantification techniques may reveal these cell-type specific fragments more sensitively in the future.

Cancer-associated fragmentation length features of cfNano vs. Illumina WGS

Specific fragment lengths have been associated with cancer-derived cfDNA fragments (reviewed in (*1*)) and these have been used as accurate cancer classifiers (*15*). Specifically, shorter mononucleosome fragments (<150bp) tend to be enriched for cancer-derived fragments (*34*). Density plots of fragment length showed that our cfNano cancer samples were enriched in these short mononucleosome fragments relative to healthy controls (Figure 4A). We used the definition from (*34*) and (*15*) to calculate the ratio of short mononucleosomes (100-150bp) to all mononucleosomes (100-220bp). The short mononucleosome ratio was significantly higher in the high tumor fraction cancer cases (mean=0.24, SD=0.03) than in the healthy cases (mean=0.16, SD=0.03), which corresponded to a t-test p-value of p=0.038 (Figure 4B). We compared these ratios calculated from our cfNano libraries with those calculated from the matched Illumina WGS libraries which were available for four of the six cancer samples, and the two library types were strongly correlated (Figure 4C). We hypothesized that improvements to Nanopore basecalling could improve alignment and adapter trimming, so we also compared base calling done with the Guppy basecaller at the time of sequencing ("2019" version) to the new "high accuracy calling" basecalling ("HAC") performed on all samples in 2022. The new ratios with the new basecalling were slightly more similar to the matched Illumina libraries (Figure 4C).

While they have not been studied as extensively as mononucleosomes, ref. (*34*) also showed that dinucleosomes were significantly shorter in cancer fragments than non-cancer fragments. This is clear from the density plots of our cfNano samples (Figure 4A), so we used the size range suggested by ref. (*34*) to calculate the ratio between short dinucleosomes (275-325bp) and all dinucleosomes (275-400bp). The short dinucleosome ratio showed even more separation between cancer vs. healthy cfNano samples than the mononucleosome ratio, with the high tumor fraction cancer cases (mean=0.62, SD=0.01) than in the healthy cases (mean=0.30, SD=0.04), corresponding to a t-test p-value of p=2E-7 (Figure 4D). We compared dinucleosome ratios calculated from our cfNano libraries with those calculated from the matched Illumina WGS libraries, and they were nearly perfectly correlated (Figure 4E). When we looked across all samples, the short mononucleosome ratio was highly correlated with the short dinucleosome ratio (Supplementary Figure 9D). Interestingly, this correlation held across the healthy samples as well as the cancer samples, indicating that the same underlying mechanism affects circulating cfDNA from both cancer and non-cancer cell types.

One of our cfNano samples used a different (non-barcoded) adapter design method from all other libraries, and this sample was a clear outlier in fragment length (Supplementary Figure 9A-D). This reinforces the caution that should be taken when comparing fragmentomic features across different library designs. We also investigated the effect on sequencing depth on cancer-associated features, by comparing full-depth datasets with datasets created by randomly choosing 2M fragments for each library (Supplementary Figure 9E-H). Sequence depth had almost no effect on either cfNano or Illumina samples down to 2M fragments.

Cancer-associated fragment end features of cfNano vs. Illumina WGS

The four bases immediately flanking cfDNA fragmentation sites have biased sequence composition that differs between cancer-derived and non-cancer-derived fragments (*35*, *36*). To study this in our cfNano samples, we first plotted the 25 most abundant 4-mer end motifs that were identified in an Illumina-based study of healthy plasma cfDNA (*36*), using a heatmap to indicate motif frequencies in each of our cfNano and matched Illumina samples (Figure 4F). There was broad agreement across all samples, although some differences between cfNano and Illumina libraries were clearly noticeable. When we plotted average Nanopore vs. Illumina frequencies for all 256 possible 4-mers, it appeared that the less abundant motifs had slightly higher frequencies in Nanopore, while the more abundant motifs had slightly lower frequencies in Nanopore (Figure 4G). Nevertheless, the relative frequencies were highly concordant overall (PCC=0.97). These were slightly more concordant when we used the 2022 “high accuracy” (HAC) basecalling compared to the original 2019 basecalling (PCC=0.97 vs. PCC=0.96). The degree of difference between the two batches of cfNano healthy samples (“BC” sequenced in 2019 and “HU” sequenced in 2022) showed only slight differences using the HAC basecalling (PCC=0.99).

Of particular interest is the CCCA end motif, which is typically the most abundant 4-mer in healthy plasma and its reduction was shown to be a cancer marker in several cancer types including lung cancer (*35*, *36*). CCCA indeed has the highest frequency across all our cfNano and Illumina WGS samples (Figure 4F-H), and was significantly lower in our three high tumor fraction cancer samples than the healthy samples (Figure 4I). However, there was a clear difference between the healthy samples generated in the “HU” and “ISPRO” batches, which we presume to be technical since these two batches behaved similarly with respect to fragment length and methylation features. We therefore only did a direct statistical comparison within the ISPRO batch, and indeed CCCA frequency in high TF tumors (mean=1.6,SD=0.06) was significantly lower than in ISPRO healthy samples (mean=1.9,SD=0.13), leading to a t-test p-value=0.007 (Figure 4I).

We found additional 16 motifs that were as significant as CCCA, although none survived FDR adjustment and so will have to be validated in larger studies (Supplementary Table 3). Like fragment lengths, end motif frequencies were not sensitive to downsampling to 2M fragments (Supplementary Figure 10A-D). Additionally, the relative frequencies of four cfNano cancer samples were not concordant with their matched Illumina WGS libraries (Figure 4J). We conclude from this that end motifs are particularly sensitive to changes in library strategy and sequencing platform, and caution must be taken when comparing across multiple batches. This is not surprising, given that fragment representation can be skewed by a number of variables during library construction and amplification, as well as sequencing errors and downstream bioinformatic steps such as adapter trimming (in our cfNano processing, we also exclude fragment ends that are soft-clipped). End motifs are highly susceptible to these biases, because even a single base pair difference results in a completely different motif. Recent benchmarking has highlighted how error frequencies can differ by sequence context between the Nanopore and Illumina platforms (*37*).

Discussion

While the sample size is small, our results suggest that cancer-specific features of DNA methylation, fragmentation, and CNA are broadly concordant between cfNano and Illumina-based WGS and WGBS methods. Downsampling analysis showed that the genomic coverage we targeted with cfNano (minimum of 2.5M aligned reads or 0.2x genome coverage) was sufficient to detect cancer-derived DNA in all samples based on DNA methylation. Cancer-associated fragmentomic features were not detected in all samples, but this is likely due to biological variability rather than sequence depth, based on similar studies using Illumina-based approaches (*34*), as well as our own results running the fragmentomic analysis on downsampled libraries. Notably, our results suggest that short dinucleosomes could be a more robust cancer marker than short mononucleosomes, although this will need to be validated in larger studies.

While most features agreed between cfNano and Illumina-based datasets, we identified fragment end motifs as one that was especially sensitive to sequencing platform differences. With the small sample size here, it is not possible to determine which of the two platforms provides the truer results. It is tempting to hypothesize that Nanopore provides a more accurate representation of actual fragment frequencies, since it does not include any PCR amplification. On the other hand, Nanopore error rates are higher and may lead to less accurate adapter trimming or a different spectrum of sequencing errors within the end motifs (although we did try to control for this by using the reference genome sequence rather than the read sequence, and by filtering out reads with soft-clipped ends). Despite these differences, we were able to detect the best known cancer-specific end motif (CCCA) when we constrained the analysis to samples from our main cfNano batch ("ISPRO"). We also detected slight differences in end motif frequencies between our own Illumina WGS and earlier published WGS data, and between our two cfNano batches which were sequenced two years apart. This indicates that end motif frequencies are susceptible to batch effects in general and should be analyzed cautiously between batches. In the future, proper controls should be developed to compare this feature across datasets sequenced at different times.

With the small number of metastatic cancer samples used in our cfNano study, it was not possible to define the lower limit of methylation sensitivity. However, the fact that some epithelial/lung content was found when downsampling higher coverage WGBS samples and in our cfNano healthy samples that had the lowest coverage, suggests that specificity needs to be improved at this very low coverage. We believe this could be improved significantly if whole-genome DNA methylation atlases (WGBS or similar) can be generated for individual cell types purified from human tissues. The purified cell type atlas we used here was based on the Illumina HM450k platform, which covers only about 10-15% of cell type specific methylation markers. Alternatively, cancer markers could come directly from discovery WGBS studies on cancer plasma cfDNA. Such datasets do exist in the private domain, but remain proprietary (e.g. (*2*)).

Even if deconvolution can be improved bioinformatically, we will almost certainly need higher read coverage for applications that require more sensitive detection. The cfNano samples analyzed here were all from metastatic disease and had relatively high tumor fraction (>=10%), and this does not represent the situation for cancer screening or detection of minimal residual

disease. Thus, the Nanopore platform will need to be able to consistently produce tens of millions of reads to enable those applications. Using cfNano directly to discover new cancer markers would not necessarily require greater coverage, but would require large numbers of patient samples.

Several other current limitations of Nanopore should be considered. The cost per base of Nanopore sequencing is currently several-fold higher than Illumina, although the new generation of ONT PromethION sequencers are meant to address this as well as increase throughput. Single-nucleotide and indel error rates are higher for the ONT platform, which could pose an issue for whole-genome analysis of mutational signatures (*38*, *39*), something we do not investigate here but is theoretically possible from cfNano. While Nanopore error rates have improved significantly over the past several years, this is a weak point that should be considered if mutations are a priority.

There are several areas where Nanopore methylation sequencing may provide unique strengths over other methods. Bisulfite-based sequencing leads to DNA fragmentation and degradation, and can obscure fragmentation patterns (*40*). Additionally, bisulfite sequencing can not distinguish between 5mC and other modifications such as 5hmC, 5fC, and 5CaC, whereas these are all in principle detectable by Nanopore. Longer cfDNA fragments have not been extensively studied due to limitations of short-read technology, and these could be valuable both for biomarker discovery and the basic biology of cfDNA (i.e. (*19*)).

Despite the limitations discussed above, the simplicity of native ONT sequencing and the number of features that can be extracted from a single run, combined with the low cost and portability of sequencer, make it an interesting proposition for clinical settings. Fast sample prep and sequencing times can allow a complete methylation analysis from sample preparation to computational classification in as little as 1-3 hours, enabling real-time medical applications in cancer (*11*, *12*). Because DNA methylation can differentiate non-cancer cell types as well, Nanopore liquid biopsy could be used to monitor collateral damage to adjacent tissue in cancer (*23*), or urgent conditions in other areas of medicine such as myocardial infarction, sepsis, and COVID-19 (*4–6*, *41*).

Code availability

Source code for fragmentomic analysis is available at https://github.com/Puputnik/Fragmentomics_GenomBiol. Source code for combined CNV and PMD methylation analysis is available at https://github.com/Puputnik/CNV_Methylation_Genome_Biol_2022. Source code for methylation deconvolution is available at https://github.com/methylgrammarlab/cfdna-ont. An archive of all source code is deposited in Zenodo DOI:10.5281/zenodo.6445784.

Data availability

Processed data files for the analyses described here are available at GEO accession GSE185307 and at Zenodo DOI: 10.5281/zenodo.6448476. These include “anonymized BAM files” generated by Megalodon for all cfNano samples. Anonymized BAM files contain all

fragmentomic information as well as read-specific DNA methylation calls. We also include methylation BED files for both Megalodon and DeepSignal. Raw data files (fastq) for cfNano samples are deposited in EGAD00001006888.

Competing interests

BPB, EK, SO, FM, and SGC are inventors on IP filings for Nanopore-based detection of cfDNA by Yissum Research Development Company of the Hebrew University of Jerusalem Ltd. BPB and SO receive funding from a research collaboration with Volition Belgium Rx, a company involved in cancer cfDNA detection.

Acknowledgments

We thank Joshua Moss and Yaping Liu for critical review of the manuscript, Sara Isaac and Myriam Maoz for assistance with sample processing, and Tiago Silva and Irene Unterman for helpful discussions. We thank the Dennis Lo lab for sharing healthy control cfDNA data from EGAD00001001602, and the Le Son Tran lab for sharing healthy and LuAd plasma cfDNA data from 10.6084/m9.figshare.16817941.v1. Computation was carried out on the Hebrew University Research Computing Services cluster, and we acknowledge Yaron Weitz and Ori Adam for their help and support. Ben Berman, Efrat Katsman, and Shari Orlanski received startup support from the Hebrew University, the Kamea B program of the Israel Ministry of Aliyah and Immigrant Integration, and the Israel Cancer Research Fund Project Grant (845755). Filippo Martignano was supported by the Italian Ministry of Health (SG-2019-12370279).

Author Contributions

BPB, FM, and SGC conceived the project. BPB, EK, SO, FM, and SGC designed the computational analyses. BPB, EK, SO, and FM performed computational and statistical analyses and generated figures. IP provided cancer plasma samples and associated clinical information. IFF, RS, and YD provided healthy plasma samples and associated clinical information. FM, AZ, and AE performed ONT sequencing. BPB wrote the manuscript, with contributions from SGC, FM, EK, and SO. BPB and SGC co-supervised the project. The first three authors (EK,SO,FM) are co-lead authors with an equal contribution to the work, and they have the right to list their names first in their CVs. The last two authors (SGC,BPB) are co-senior authors who jointly supervised the work, and they have the right to list their names last in their CV.

Figure legends

Figure 1: Estimating cell type fractions from cfNano. (A) Non-Negative Least Squares regression based on (*5*) was used to deconvolute cell types in healthy plasma cfDNA samples from three whole-genome DNA methylation studies. Two representative samples are shown for each study (FF8 and FF23 for the Fox-Fisher et al. study, N1 and N8 for the Nguyen et al. study, and BC03 and HU11 for our cfNano samples.) Each sample is downsampled from full read depth down to an average genome coverage of 0.001 (corresponding to approximately 13,000 fragments). All samples are shown in Supplementary Figures 1-3. (B) Deconvolution of all samples at full depth, with samples ordered within each group by epithelial cell fraction. Healthy vs. lung adenocarcinoma (LuAd) is shown as an annotation bar, as is the "on-target/off-target" status of the Nguyen et al. samples and the source site (HU Israel vs. BC Italy) for the cfNano samples. Asterisks mark the two HU samples with coverage less than 0.2x sequence depth. Statistical significance (p-value=0.004) is shown for percent epithelial in healthy cfNano samples vs. LuAd cfNano samples. (C) The same samples downsampled to 0.2x sequence depth. (D) ichorCNA CNA plots for 4 representative cfNano samples, two healthys and two LuAds. Plots for all samples are included as Additional Data File 1. (E) Tumor Fraction estimates (TF) from four LuAd samples based on ichorCNA from cfNano and matched Illumina WGS. (F) Two-component DNA methylation deconvolution of lung fraction using CpGs from MethAtlas purified lung epithelia samples, showing scatter plot of ichorCNA estimates vs. deconvolution estimates for all cfNano samples. Statistical significance is shown for DNA methylation estimate of healthy cfNano vs. LuAd cfNano samples (p-value=0.003). (G) Two-component DNA methylation deconvolution of lung cancer fraction using CpGs from TCGA LuAd tumor samples, showing scatter plot of ichorCNA estimates vs. deconvolution estimates for all cfNano samples (healthy vs. LuAd p-value=0.004). Statistical significance for panels B,C,F, and G was determined by one-tailed t-test.

Figure 2: Genomic context of DNA methylation changes detected using cfNano. (A) Plasma cfDNA methylation levels were averaged from -1kb to +1kb at 5,974 pneumocyte-specific NKX2-1 transcription factor binding sites (TFBS) taken from (*25*). All methylation values are fold change relative to the flanking region (region from 0.8kb-1kb from the TFBS). From left to right, plots show 23 healthy plasma samples from (*13*), and 32 healthy plasma samples from (*42*), 3 healthy and 18 LuAd WGBS samples from (*20*), and 7 healthy and 6 LuAd cfNano samples from this study. (B) Average DNA methylation across chr16p, comparing lung tissue WGBS (top) to plasma cfNano samples from this study (bottom). Reference Partially Methylated Domains (PMDs) are taken from (*29*). (C) Methylation delta is shown for all 10Mbp bins overlapping a reference PMD (methylation delta defined as the average methylation of the bin minus the average methylation genome-wide). Each cancer sample was compared to the group of healthy samples using a one-tailed t-test, and statistical significance is shown using asterisks. (D) 10Mbp PMD bins were stratified by copy number status for each cancer sample using ichorCNA, and statistically significant differences were calculated by performing one-tailed Wilcoxon tests within each sample.*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

Figure 3: cfNano preserves nucleosome positioning signal. (A) Alignments to 9,780 CTCF

motifs within non-promoter ChIP-seq peaks were taken from (*33*). (B) Sequence coverage of mononucleosomes (130-155bp) from cfNano samples is shown as fold-change vs. average coverage across the genome (top). Mononucleosome coverage for matched Illumina samples (bottom). (C) Same analysis, using a randomly selected downsampling of 2 million reads from each sample. Two cfNano samples with less than 2M reads total are omitted.

Figure 4: Cancer-associated fragmentation features of cfNano vs. Illumina WGS. (A) Fragment length density plot for each cfNano sample, with cancer samples divided into low tumor fraction (TF<0.15) and high tumor fraction (TF>0.15) based on ichorCNA. Short mononucleosomes are defined as 100-150bp (*15*, *34*) and short dinucleosomes are defined at 275-325bp. (B) The ratio (fraction) of short mononucleosome fragments (100-150bp) to all mononucleosome fragments (100-220bp). (C) Short mononucleosome ratios based on cfNano are compared to short mononucleosome ratios based on matched Illumina WGS libraries for four LuAd cases. cfNano samples were processed with either 2019 Oxford Nanopore Fast basecalling model (2019) or 2022 Oxford Nanopore High Accuracy model (HAC), as indicated by color. (D) The ratio (fraction) of short dinucleosome fragments (275-325bp) to all dinucleosome fragments (275-400bp). (E) Short dinucleosome ratios based on cfNano vs. Illumina WGS ratios for matched LuAd samples. (F) Frequency of 4-mer sequences occurring at fragment ends, for cfNano vs. matched Illumina samples. The 25 most frequent 4-mers are shown in ranked order based on frequencies in (*36*). (G) End motif frequencies for all 256 possible 4-mers, comparing average frequency in four cfNano samples vs four matched Illumina WGS samples. (H) End motif frequencies, comparing average frequency in four healthy HU Italy cfNano samples vs three healthy HU Israel cfNano samples. (I) Frequency of CCCA 4-mer in all cfNano samples. (J) CCCA 4-mer frequencies from cfNano samples vs. frequencies calculated from Illumina WGS for four matched LuAd samples. Statistical significance levels for panels B,D, and I were determined by two-tailed t-test.

Supplementary Figure 1: DNA methylation deconvolution for Fox-Fisher et al. samples. Each sample from (*13*) was downsampled from full depth to 0.001x coverage, and sample ordering is the same as Fig 1B-C. Short names are used, and full sample information is available in Supplementary Table 2.

Supplementary Figure 2: DNA methylation deconvolution for Nguyen et al. samples. Each sample from (*20*) was downsampled from full depth to 0.001x coverage, and sample ordering is the same as Fig 1B-C. Short names are used, and full sample information is available in Supplementary Table 2.

Supplementary Figure 3: DNA methylation deconvolution for cfNano samples. Each cfNano sample from the current study was downsampled from full depth to 0.001x coverage, and sample ordering is the same as Fig 1B-C. Short names are used, and full sample information is available in Supplementary Table 1.

Supplementary Figure 4: Full cell type assignments in deconvolution analysis. (A) Cell-type deconvolution for WGBS and cfNano datasets, using 25 cell types from (*5*). (B) 25 cell type deconvolution of all samples downsampled to 0.2x sequence coverage. (C) The four cell-type groups from Figure 1 (Lymphocyte, Granulocyte, Epithelial, and Other) and which of the 25 cell

types were collapsed into each group. All cell types not assigned to one of the four groups are shown as a singleton cell type in Figure 1.

Supplementary Figure 5: ichorCNA tumor fractions of downsampled Illumina samples. Four Illumina plasma samples from LuAd patients are shown. ichorCNA tumor fraction was computed at full sequence depth (x axis) and by randomly downsampling the Illumina samples to have the same number of fragments as the corresponding cfNano sample.

Supplementary Figure 6: Calling cfNano methylation with two different methods. (A) DeepSignal and Megalodon were used to call CpG methylation for each cfNano sample. CpGs were divided into those covered by DeepSignal only, Megalodon only, or Both. Those covered by both were divided into those that got identical methylation status vs. different methylation status. (B) Grouped cell type deconvolution is shown for all samples for Megalodon and DeepSignal processed data. Megalodon version is reproduced from Figure 1B. (C) Two-component deconvolution is shown for all samples for Megalodon and DeepSignal processed data. Megalodon versions are reproduced from Figure 1F and 1G, respectively.

Supplementary Figure 7: Genomic context of DNA methylation changes. (A) Methylation in 18 TCGA WGBS non-lung tumors (left) and 11 TCGA WGBS lung tumors and adjacent normal tissue (right) from (*29*). Plasma cfDNA methylation levels were averaged from -1kb to +1kb relative to 5,974 pneumocyte-specific NKX2-1 transcription factor binding sites (TFBS) taken from (*25*). All methylation values are shown as relative to the flanking region (from 0.8kb-1kb relative to TFBS). (B) 9,274 adrenal cortical cell specific KLF5 TFBS taken from (*25*). From left to right, plots show 23 healthy plasma samples from (*13*) and 32 healthy plasma samples from (*42*), followed by 3 healthy and 18 LuAd WGBS samples from (*20*) and 7 healthy and 6 LuAd cfNano samples from this study (C) cfNano methylation levels for lung NKX2-1 (same as Figure 2A), using DeepSignal methylation calling. (D) cfNano methylation levels for adrenal cortical KLF5 (same as this Figure, panel B), using DeepSignal methylation calling. (E) IGV analysis (same as Figure 2B) using DeepSignal methylation calling. (F-G) Genome-wide PMD bin analysis (same as Figure 2C-D) using DeepSignal methylation calling.

Supplementary Figure 8: cfNano preserves fragmentomic and DNA methylation markers of nucleosome positioning. Alignments to CTCF motifs within 9,780 distal ChIP-seq peaks from (*33*). (A, top) cfDNA fragment coverage shown as fold-change vs. average coverage depth across the genome. The plot includes only fragments of length 130-155bp to maximize resolution. (A, bottom) Matched Illumina samples of higher sequencing depth (median 17.0M fragments in Illumina vs. 6.4M in ONT samples). (B) CTCF DNA methylation of Nanopore samples from this study at CTCF sites. (C) DNA methylation from seven lung tissue WGBS samples from TCGA (*29*).

Supplementary Figure 9: Effects of downsampling on fragment length of cfNano and Illumina WGS. (A-C) Data from Figure 4A,B,D are reproduced with the addition of sample 19_326 (which used a different, non-barcoded, cfNano adapter design), as well as matched Illumina samples. (D) Short mononucleosome ratios (x axis) plotted against short dinucleotide ratios (y axis). Panels (E-H) show the same plots as panels A-D, but with each sample randomly downsampled to 2M fragments. Statistical significance levels for panels B,C,F, and G were

determined by two-tailed t-test.

Supplementary Figure 10: Effects of downsampling on fragment end features of cfNano and Illumina WGS. (A-B) are reproduced from main Figure 4F and 4I, with the addition of sample 19_326 (which used a different, non-barcoded, cfNano adapter design), as well as matched Illumina samples. Panels (C-D) show the same plots, but with each sample randomly downsampled to 2M fragments.Statistical significance levels for panels B and D were determined by two-tailed t-test.

Supplementary Table 1: cfNano sample information and statistics

Supplementary Table 2: External WGBS sample information

Supplementary Table 3: Tumor vs. normal differences for 4-mer end motifs

Supplementary Data File 1: ichorCNA plots for all cfNano and matched Illumina WGS samples.

Methods

ISPRO Plasma cfDNA samples, library construction, and sequencing. ISPRO Samples, library construction and sequencing were described in our initial publication of these sequences (*17*). The original sample names from that study are listed in Supplementary Table 1. Notably, one sample (S1/19_326) was produced using a different library kit (SQK-LSK109 vs. NBD-EXP104+SQK-LSK109 for all other samples). This is the singleplex library kit, which results in shorter adapter-ligated templates overall (due to the lack of barcodes) and thus responds differently to the equivalent clean up bead concentration. Also adapter trimming is performed differently in 19_326 due to the library kit differences. For these reasons, fragmentomic properties are not directly comparable between 19_326 and other samples. We thus analyzed 19_326 separately for all fragmentomic analyses (included in Supplementary Figures 9-10), but included it in all figures when analyzing methylation and copy number alterations, where small differences in fragment length are not expected to make a difference. Standard MinKNOW runtime control was used without modification (S1 using distribution version 18, and all others using version 19).

HU Plasma cfDNA samples, library construction, and sequencing. HU healthy samples are cfDNA extracted from 4mL plasma as described in (*13*). The original sample names from that study are listed in Supplementary Table 1. Barcoded libraries were created using the NBD-EXP104 and SQK-LSK109 kits as for ISPRO samples. They were sequenced on a single flow cell, using standard MinKNOW runtime control (distribution v.21.11.7) without modification.

2019 Fast model basecalling and alignment of Nanpore fast5 files. Basecalling was done using MinKNOW distribution v.19.06.9 Guppy version 3.0.6+9999d81. For singleplex sample (S1/19_326), adapters were trimmed with Porechop (https://github.com/rrwick/Porechop) using: "--discard_middle --extra_end_trim 0". For both HU and ISPRO samples, alignments were performed to GCF_000001405.39_GRCh38.p13 with minimap2 (Version 2.13-r850), using the parameters "-ax map-ont --MD".

2022 High Accuracy Calling (HAC) basecalling and alignment of Nanopore fast5 files. HU and ISPRO Fast5 files were demultiplexed with guppy_barcoder (Version 5.0.16+b9fcd7b5b) using "--trim_barcodes --barcode_kits EXP-NBD104". Individual barcode Fast5 files were processed into fastq and bam files using Megalodon v. 2.4.2 with the following command-line parameters "--edge-buffer 0 --mod-min-prob 0 --guppy-params '-d /usr/local/hurcs/guppy/6.0.1/data --barcode_kits EXP-NBD104 --trim_barcodes' --remora-modified-bases dna_r9.4.1_e8 hac 0.0.0 5mc CG 0 --guppy-config dna_r9.4.1_450bps_hac.cfg". Internally, Megalodon used Guppy server version 6.0.1+652ffd1, and basecalling model r9.4.1_450bps_hac. By default, Megalodon filters out multi-mapping (supplementary) reads and uses the minimap2 "map-ont" mode to filter low quality mappings.

Individual tile Fast5 files were run individually, and mod_mapping.bam files were merged using samtools merge (v1.14). Samtools/HTSlib versions before v.1.14 do not handle the Mm/Ml modification tages. Fastq files were merged by concatenating (cat) all individual basecalls.fastq files. Merged mod_mapping.bam files have been deposited in GEO GSE185307 and at Zenodo DOI: 10.5281/zenodo.6448476. Because Megalodon reports only the *reference* sequence in the BAM records, and does not report any base substitutions, these are anonymous BAM files which do not contain any SNP information, and thus contain no personally identifiable information. These are the primary files used for both fragmentomic and methylation analysis, described in more detail below, and are available from GEO GSE185307 and at Zenodo DOI: 10.5281/zenodo.6448476.

DeepSignal methylation calling and processing. We used DeepSignal Version 0.1.8 (4), with model "model.CpG.R9.4_1D.human_hx1.bn17.sn360.v0.1.7+/bn_17.sn_360.epoch_9.ckpt", which was downloaded from the DeepSignal Google Drive (https://drive.google.com/open?id=1zkK8Q1gyfviWWnXUBMcIwEDw3SocJg7P). We used the DeepSignal call_mods (modification_call) output tsv file, extracting the (strand-specific) methylation calls for each CpG from column 9 (called_label field), and calculated a methylation beta value by taking the number of methylated reads (value 1) divided by the total number of reads (value 0 or value 1). These were collapsed into a bedgraph file with a value between 0-1 for every CpG covered. These are available as file "grouped-beta-value_bedgraph.zip" in GEO accession GSE185307 and at Zenodo DOI: 10.5281/zenodo.6448476. All genomic coordinates are in GRCh38 and are zero-based.

Megalodon methylation calling and processing. Basic running of Megalodon is described above. To extract (stranded) methylation information from the mod_mapping.bam files, we used modbam2bed (https://github.com/epi2me-labs/modbam2bed) v.0.4.5, specifying a minimum probability threshold of 0.667, and filtering out positions with 0 confident reads using awk. The full command line was "modbam2bed --cpg -t 4 -a 0.333 -b 0.667 | awk '($5>0){print} > out.bed".

All coordinates are in GRCh38 and are 0-based. These files are named "*.5mC.cut0.667.hg38.bed.gz". Column 11 corresponds to the percent of reads methylated. Modbam2bed does not provide a column for the actual number of reads that this percentage is based on, but it can be calculated from the other columns. `readCount=(col5*col10)/1000`. We also profide a simple bedgrpah with just the methylation fraction (beta) values in files named "*cut0.667.hg38.sorted.bedgraph.gz". These can be loaded into any genome browser. Both file types are available in GEO accession GSE185307 and at Zenodo DOI: 10.5281/zenodo.6448476

Mapping of cfNano methylation data to HM450k probes. Using the zero-based stranded bed files from modbam2bed ("5mC.cut0.667.hg38.bed.gz" files), we mapped each CpG covering either the forward or reverse strand of each CpG on the Infinium 450k array. For each modbam2bed stranded column, we first got the readCount as `(col5*col10)/1000`. We then multiplied the methylation percentage by the read count for each strand to get the number of methylated reads. Then we divided the sum of the methylated read counts by the sum of the total read counts to get the unstranded percent methylation (beta value).

Methylation calling from external WGBS datasets. For Fisher-Fox et al., methylation "beta.gz" files were obtained from GSE186888, and processed as recommended using wgbs_tools (https://github.com/nloyfer/wgbs_tools) beta2bed function to obtain fraction methylated and read count for each CpG. For Nguyen et al., bed files with methylation fractions and read counts were obtained from Figshare 10.6084/m9.figshare.16817941.v1.

For Sun et al., we obtained fastq files from EGAD00001001602 and aligned using Biscuit (https://github.com/huishenlab/biscuit) v.0.3.15.20200318 using the command line "biscuit align -t 16 hg38.fa CTR153.fq.gz -b 1" piped into samblaster (*43*) v.0.1.24 to mark and remove duplicates with the command line "samblaster -i stdin -o stdout -M --excludeDups --addMateTags --ignoreUnmated -d CTR153.hg38_discordant.sam -s CTR153.hg38_split.sam --maxSplitCount 2 --maxUnmappedBases 50 --minIndelSize 50 --minNonOverlap 20 -u CTR153.hg38_.fastq --minClipSize 20".

Methylation coverage downsampling. To downsample methylation coverage from bed files with read count and fraction methylated columns, we used a custom Perl script in the https://github.com/methylgrammarlab/cfdna-ont repository called downsampleMethylBed.pl. This script treats each read at each CpG as an independent observation, and then randomly samples from these until it has enough observations to reach the average genomic coverage requested. To obtain the coverage levels shown in Figure 1, it was run with the command line "downsampleMethylBed.pl --coverageLevels 1E-3,2E-3,5E-3,1E-2,2E-2,5E-2,1E-1,2E-1,5E-1,1E0,2E0,5E0,1E1,2E1,5E1,8E1 --fracTotalFieldsFrom0 -3,4 --ncpgsGenome 28217005".

Full cell type methylation deconvolution. For the full cell type deconvolution in Figures 1A-C, we used the non-negative least squares regression (NNLS) method from (*5*). Specifically, we used the code from https://github.com/nloyfer/meth_atlas/blob/master/deconvolve.py. For Megalodon methylation analysis, we used the top 1,000 hypermethylated and 1,000 hypomethylated CpGs for each of the 25 cell types provided. Full results were plotted using a modified version of the original deconvolve.py which we have deposited here:

https://github.com/methylgrammarlab/cfdna-ont/blob/main/deconvolution_code/deconvolution_moss/plot_deconv.py. These are shown in Supplementary Figure 4A-B.

For Figure 1A-C, we collapsed cell types into 8 groups using the file https://github.com/methylgrammarlab/cfdna-ont/blob/main/deconvolution_code/deconvolution_moss/group_file_for_plot_green_epithilial.csv (shown visually in Supplementary Figure 4C). We plotted results using code in https://github.com/methylgrammarlab/cfdna-ont/blob/main/deconvolution_code/deconvolution_moss/deconvolution_plot.R. For DeepSignal methylation data, the procedure was the same except we used the top 2,000 hypermethylated and top 2,000 hypomethylated CpGs, to account for the significantly smaller number of CpGs called in the DeepSignal data (shown in Supplemental Figure 6A).

ichorCNA analysis. Samtools (Version 1.9) was used to filter BAM alignments, unmapped reads, secondary and supplementary reads, reads with mapping quality less than 20 as in (*44*), and reads longer than 700bp. For Illumina alignments we trimmed all 'N' nucleotides from the 3' ends of fastq data, alignments were performed to GCF_000001405.39_GRCh38.p13 with BWA mem (43), duplicates were marked using picard MarkDuplicates and removed with samtools;read pairs without the properly-paired flag were removed. Pipelines used for preprocessing and filtering of both Nanopore and Illumina data are available at https://github.com/Puputnik/Fragmentomics_GenomBiol. Somatic copy number analysis was performed using the ichorCNA package v.0.3.2 (*21*). We used default settings to determine copy number alterations and tumor fraction for each cancer sample. If the percentage of genome covered by CN alterations was less than 15%, then the tumor fraction was determined to be unstable and set to 0. ichorCNA tumor fraction estimates are available in Supplementary Table 1, and genomic CNA plots for all samples are available as Additional Data File 1.

Two-component cell type methylation deconvolution using healthy lung epithelia. To determine lung fraction specifically from different datasets, we devised a "two-component" version of the NNLS regression model described above. To compose the atlas of differentially-methylated probes in 25 human tissues and cell types, we used the data collected and tissue-specific feature selection method from the MethAtlas package (https://github.com/nloyfer/meth_atlas)(*5*). The script feature_selection.m was used to select Lung_cell epithelial specific CpGs. For the Megalodon version, the cutoff was set to select the top 1,000 hypermethylated and the top 1,000 hypomethylated probes, for the three Lung_cell epithelia samples vs. the four healthy plasma cfDNA samples from (*5*). For DeepSignal methylation data, the procedure was the same except we used the top 2,000 hypermethylated and top 2,000 hypomethylated CpGs, to account for the significantly smaller number of CpGs called in the DeepSignal data (shown in Supplemental Figure 6A). We removed any probe that did not have valid (non-NA) values for 2 or more of the Lung_cell samples and 2 or more of the healthy plasma samples.

For each probe, the 450k beta values were averaged to produce a single Lung-specific beta value $\underline{X_1}$. The same was done for the four plasma cfDNA samples from to yield a healthy cfDNA beta value $\underline{X_2}$. We used the Lawson-Hanson algorithm for non-negative least squares (NNLS)

(https://cran.r-project.org/web/packages/nnls) to perform non-negative least squares regression as in (*5*). Specifically, we identified non-negative coefficients $\beta_1$and β_2, representing the fraction of Lung cells and normal blood cells in the Nanopore cfDNA mixture, respectively, subject to the constraints $argmin_{\beta}\left|\left|X\beta - Y\right|\right|_2$ and $\beta \geq 0$. Then a single Lung fraction β was determined by having $\beta_1$and β_2 sum to 1, with the equation $\beta = \frac{\beta_1}{(\beta_1+\beta_2)}$.

Two-component cell type methylation deconvolution using TCGA lung tumors. We downloaded the Infinium 450k beta value files for TCGA Lung Adenocarcinoma (LUAD) tumors using the ELMER packaged in Bioconductor (*45*). We removed any probe that did not have valid (non-NA) values for 2 or more of the LUAD samples and 2 or more of the healthy plasma samples. In order to make this analysis completely independent from the healthy lung epithelia deconvolution analysis, we excluded 488 that were in the 2,000 probe set for the Megalodon healthy lung analysis, and an additional 396 that were in the 4,000 probe set for the DeepSignal healthy lung analysis (described above). We then performed a t-test to compare the methylation beta values of these LUAD specific probes to the four plasma cfDNA samples from the MethAtlas paper (*5*), requiring a Benjamini-Hochberg corrected FDR of <0.001 and an absolute beta value difference of 0.3 or greater.

Correcting TCGA methylation model for cancer cell purity. NNLS was performed as above for the TCGA lung tumor deconvolution, with the following adaptation. The deconvolution assumes that each of the reference cell types is a representation of the purified cell type, but this is not the case for bulk TCGA tumors which have a median of luekocyte fraction of 30% (*22*). For each probe in each TCGA cancer sample, we corrected for this by solving for the equation $M_m = M_c\beta + M_l(1-\beta)$, where M_m is the methylation of the mixture, M_c is (unknown) methylation of the cancer cells, M_l is the (known) methylation of the leukocytes, and β is the (known) percentage of cancer cells in the mixture. M_l was taken as the average of white blood cell samples from the MethAtlas (*5*), and β was taken as the "tumor purity" estimate based on somatic copy number alterations from the TCGA PanCan Atlas project using the ABSOLUTE program (*22*), downloaded from the PanCan Atlas website (TCGA_mastercalls.abs_tables_JSedit.fixed.txt, URL https://gdc.cancer.gov/about-data/publications/pancanatlas). We used the pure cancer cell estimates M_c, and performed NNLS regression as described above.

DNA methylation in 10 Mbp PMD bins. To generate Figures 2D-E, GRCh19 segmentation results from our previous CNV analysis (*17*) were divided into non-overlapping 10Mb bins. Copy number status of each bin was determined by log2ratio segment mean > 0.10 and < -0.10 for Gain and Loss respectively. For the healthy samples, 10Mb bins were generated from the whole genome. GRCh38 Methylation files were converted to GRCh37 using liftover R package. We selected only the bins overlapping one or more common Partially Methylated Domains (PMDs) from (*29*). Within these PMD bins, we took the average of all "solo-WCGW" CpGs overlapping a PMDs, with "solo-WCGW" annotation also from (*29*). We calculated the bin average from these CpGs as sum(frac_methylation_each_CpG)/CpG_count. We then subtracted this bin average and subtracted it from the average of all CpGs in the genome, to get the Methylation Delta shown in Figure 2C-D. Common PMD and solo-WCGW annotations were taken from file

https://zwdzwd.s3.amazonaws.com/pmd/solo_WCGW_inCommonPMDs_hg38.bed.gz. Statistical significance between each cancer sample's bins and all pooled healthy sample bins in Figure 2D was calculated by one-sided Wilcoxon test (because we decided a priori to look only for hypomethylation in the cancer samples). For copy number analysis in Figure 2D, each pair of copy number groups was compared using a one-sided Wilcoxon test, to test the hypotheses that diploid should have higher methylation than amplified regions, and deleted regions should have the highest methylation. The files and pipeline used for this analysis (including segmentation results from Martignano et al. (1)) are available at https://github.com/Puputnik/CNV_Methylation_Genome_Biol_2022.

Transcription factor binding site (TFBS) analysis. First, we used HOMER to identify predicted NKX2-1 binding sites (using the HOMER built in matrix "nkx2.1.motif") across the GRCh38 genome, and removed any site within the ENCODE blacklist. For normal lung cell analysis, we intersected this list with 6,754 ATAC-seq peaks identified in the pneumocyte (PAL) cluster 13 CREs from (*46*) (downloaded from supplemental table 6 of that paper "Table_S6_Union_set_of_cCREs.xlsx"). We then selected 5,974 peaks that overlapped a predicted NKX2-1 TFBS, and centered each on the predicted NXK2-1 TFBS. If multiple TFBS were present in the peak, we took the motif with the highest HOMER log-odds match score. This TFBS set is available as file "nkx2.1.incluster13_distalPeaks_PAL.bed.highestScoreMotifs.hg38.bed" in GEO accession GSE185307 and at Zenodo DOI: 10.5281/zenodo.6448476. To calculate relative methylation levels, raw methylation levels in each bin were divided by the mean methylation within all bins from -1000 to -800 and +800 to +1000 across all NKX2-1 sites. For Supplementary Figure 7A, we used all WGBS cancer types that were represented by normal tissues in the scATAC-seq atlas, as this was the atlas used to define pneumocyte specific (PAL) peaks. For TCGA lung and non-lung samples in Supplementary Figure 7A, we downloaded TCGA WGBS bedgraph files from https://zwdzwd.github.io/pmd (*29*). We used all WGBS cancer types that were represented by normal tissues in the scATAC-seq atlas, as this was the atlas used to define pneumocyte specific (PAL) peaks. These TGCA types included LUAD and LUSC (Lung tissue from atlas), CRC (Transverse colon tissue from atlas), BRCA (Breast tissue from atlas), STAD (Stomach tissue from atlas), and UCEC (Uterus tissue from atlas).

KLF5 transcription factor binding site (TFBS) analysis (Supplementary Figure 7B). As with *NKX.2* above, we used HOMER to identify predicted KLF5 binding sites (using the HOMER built in matrix "klf5.motif") across the GRCh38 genome, and removed any site within the ENCODE blacklist. As a control we intersected this list with 9,274 ATAC-seq peaks identified in the cluster 43 CREs from (*46*) (downloaded from supplemental table 6 of that paper "Table_S6_Union_set_of_cCREs.xlsx"). We then selected 1,762 peaks that overlapped a predicted KLF5 TFBS, and centered each on the predicted KLF5 TFBS. If multiple TFBS were present in the peak, we took the motif with the highest HOMER log-odds match score. This TFBS set is available as file "klf5.incluster43Distal.txt.highestScoreMotifs.bed" in GEO accession GSE185307.

CTCF nucleosome positioning analysis. We used 9,780 evolutionarily conserved CTCF motifs occurring in distal ChIP-seq peaks, which were taken from (*33*). Nanopore or Illumina fragments

within the size range of 130-155bp were used for fragment coverage analysis, with reads being extracted from BAMs as described above. These shorter mononucleosomal fragments showed similar nucleosomal patterns but gave higher spatial resolution than 156-180 bp fragments. Deeptools (Version 3.5.0) bamCoverage was used with the parameters "--ignoreDuplicates --binSize -bl ENCODE_blacklist -of bedgraph --effectiveGenomeSize 2913022398 --normalizeUsing RPGC". For Illumina WGS, we used the additional parameter "--extendedReads 145". The bedgraph was converted to a bigwig file using bigWigToBedGraph downloaded from UCSC Genome Browser. This bigwig file was passed to Deeptools computeMatrix with the command line parameters "reference-point --referencePoint center -out table.out", and the table.out file was imported into R to create fragment coverage heatmap.

Fragment length analysis. Samtools (Version 1.9) was used to filter BAM alignments, unmapped reads, secondary and supplementary reads, reads with mapping quality less than 20 as in (*44*), and reads longer than 700bp. For Illumina alignments we trimmed all 'N' nucleotides from the 3' ends of fastq data, alignments were performed to GCF_000001405.39_GRCh38.p13 with BWA mem (43),duplicates were marked using picard MarkDuplicates and removed with samtools. Pipelines used for preprocessing and filtering of both Nanopore and Illumina data, and analyzed data are available at https://github.com/Puputnik/Fragmentomics_GenomBiol. In addition, only reads with barcodes at both ends (obtained using the --require_barcodes_both_ends flag while demultiplexing) were used for fragment length analysis of the multiplexed samples (all except 19_326) . Reads with soft clipping at either the 5' or 3' ends were removed. Fragment length was calculated from the Minimap2 BAM CIGAR column by summing all counts. Short mononucleosome ratio was calculated as $\frac{numfrags_{100-150bp}}{numfrags_{100-220bp}}$ (150bp is the same cutoff for short fragments used in (*34*)). Short dinucleosome ratio was calculated as $\frac{numfrags_{275-3250bp}}{numfrags_{275-400bp}}$ (this was determined visually from Figure 2D of publication (*34*)).

End motif analysis. Fragments and reads were processed and filtered as in fragment length analysis. We only used end 1 of both cfNano and Illumina reads, to avoid problems with adapter trimming at end 2. To avoid biases that would affect end motif analysis, we also removed reads with any soft clipping at end 1. The first 4 bases of each fragment were extracted and used for 4-mer analysis. To avoid errors in Nanopore base calling, these 4 bases were extracted from the reference genome. Motif frequency was calculated as $\frac{numfrags_{4mer}}{numfrags_{total}}$. For top 25 motifs and ranking order in Figures 4 and Supplementary Figure 10, we used (*36*). Files and pipelines used for Fragment length and end motif analyses are available at https://github.com/Puputnik/Fragmentomics_GenomBiol

Statistical tests. Student's t-test for all sample comparisons where at least one test group had less than five samples, otherwise Wilcoxon test was used.

References

1. 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** (2021), doi:10.1126/science.aaw3616.

2. E. A. Klein, D. Richards, A. Cohn, M. Tummala, R. Lapham, D. Cosgrove, G. Chung, J. Clement, J. Gao, N. Hunkapiller, A. Jamshidi, K. N. Kurtzman, M. V. Seiden, C. Swanton, M. C. Liu, Clinical validation of a targeted methylation-based multi-cancer early detection test using an independent validation set. *Ann. Oncol.* **32**, 1167–1177 (2021).

3. P. Y. Dor, P. H. Cedar, Principles of DNA methylation and their implications for biology and medicine. *Lancet*. **392**, 777–786.

4. H. Zemmour, D. Planer, J. Magenheim, J. Moss, D. Neiman, D. Gilon, A. Korach, B. Glaser, R. Shemer, G. Landesberg, Y. Dor, Non-invasive detection of human cardiomyocyte death using methylation patterns of circulating DNA. *Nat. Commun.* **9**, 1443 (2018).

5. 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. 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).

6. T. E. Andargie, N. Tsuji, F. Seifuddin, M. K. Jang, P. S. Yuen, H. Kong, I. Tunc, K. Singh, A. Charya, K. Wilkins, S. Nathan, A. Cox, M. Pirooznia, R. A. Star, S. Agbor-Enoh, Cell-free DNA maps COVID-19 tissue injury and risk of death and can cause tissue injury. *JCI Insight*. **6** (2021), doi:10.1172/jci.insight.147610.

7. S. Y. Shen, R. Singhania, G. Fehringer, A. Chakravarthy, M. H. 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).

8. 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).

9. Z. W.-S. Yuen, A. Srivastava, R. Daniel, D. McNevin, C. Jack, E. Eyras, Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat. Commun.* **12**, 3438 (2021).

10. Y. Liu, W. Rosikiewicz, Z. Pan, N. Jillette, P. Wang, A. Taghbalout, J. Foox, C. Mason, M. Carroll, A. Cheng, S. Li, DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, 295 (2021).

11. H. Dogan, A. Patel, C. Herold-Mende, S. Pfister, W. Wick, M. Loose, A. von Deimling, M. Sill, D. Jones, F. Sahm, P07.04 Rapid-CNS2: Rapid comprehensive adaptive nanopore-

sequencing of CNS tumors, a proof of concept study. *Neuro. Oncol.* **23**, ii25–ii26 (2021).

12. L. Djirackor, S. Halldorsson, P. Niehusmann, H. Leske, D. Capper, L. P. Kuschel, J. Pahnke, B. J. Due-Tønnessen, I. A. Langmoen, C. J. Sandberg, P. Euskirchen, E. O. Vik-Mo, Intraoperative DNA methylation classification of brain tumors impacts neurosurgical strategy. *Neurooncol Adv*. **3**, vdab149 (2021).

13. I. Fox-Fisher, S. Piyanzin, B. L. Ochana, A. Klochendler, J. Magenheim, A. Peretz, N. Loyfer, J. Moss, D. Cohen, Y. Drori, N. Friedman, M. Mandelboim, M. E. Rothenberg, J. M. Caldwell, M. Rochman, A. Jamshidi, G. Cann, D. Lavi, T. Kaplan, B. Glaser, R. Shemer, Y. Dor, Remote immune processes revealed by immune-derived circulating cell-free DNA. *Elife*. **10** (2021), doi:10.7554/eLife.70520.

14. D. Mathios, J. S. Johansen, S. Cristiano, J. E. Medina, J. Phallen, K. R. Larsen, D. C. Bruhm, N. Niknafs, L. Ferreira, V. Adleff, J. Y. Chiao, A. Leal, M. Noe, J. R. White, A. S. Arun, C. Hruban, A. V. Annapragada, S. Ø. Jensen, M.-B. W. Ørntoft, A. H. Madsen, B. Carvalho, M. de Wit, J. Carey, N. C. Dracopoli, T. Maddala, K. C. Fang, A.-R. Hartman, P. M. Forde, V. Anagnostou, J. R. Brahmer, R. J. A. Fijneman, H. J. Nielsen, G. A. Meijer, C. L. Andersen, A. Mellemgaard, S. E. Bojesen, R. B. Scharpf, V. E. Velculescu, Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nat. Commun.* **12**, 1–14 (2021).

15. 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).

16. S. H. Cheng, P. Jiang, K. Sun, Y. K. Y. Cheng, K. C. A. Chan, T. Y. Leung, R. W. K. Chiu, Y. M. D. Lo, Noninvasive prenatal testing by nanopore sequencing of maternal plasma DNA: feasibility assessment. *Clin. Chem.* **61**, 1305–1306 (2015).

17. F. Martignano, U. Munagala, S. Crucitta, A. Mingrino, R. Semeraro, M. Del Re, I. Petrini, A. Magi, S. G. Conticello, Nanopore sequencing from liquid biopsy: analysis of copy number variations from cell-free DNA of lung cancer patients. *Mol. Cancer*. **20**, 32 (2021).

18. T. Baslan, S. Kovaka, F. J. Sedlazeck, Y. Zhang, R. Wappel, S. Tian, S. W. Lowe, S. Goodwin, M. C. Schatz, High resolution copy number inference in cancer using short-molecule nanopore sequencing. *Nucleic Acids Res.* **49**, e124 (2021).

19. S. C. Y. Yu, P. Jiang, W. Peng, S. H. Cheng, Y. T. T. Cheung, O. Y. O. Tse, H. Shang, L. C. Poon, T. Y. Leung, K. C. A. Chan, R. W. K. Chiu, Y. M. D. Lo, Single-molecule sequencing reveals a large population of long cell-free DNA molecules in maternal plasma. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021), doi:10.1073/pnas.2114937118.

20. H.-N. Nguyen, N.-P. T. Cao, T.-C. Van Nguyen, K. N. D. Le, D. T. Nguyen, Q.-T. T. Nguyen, T.-H. T. Nguyen, C. Van Nguyen, H. T. Le, M.-L. T. Nguyen, T. V. Nguyen, V. U. Tran, B. A. Luong, L. G. H. Le, Q. C. Ho, H.-A. T. Pham, B. T. Vo, L. T. Nguyen, A.-T. H. Dang, S. D. Nguyen, D. M. Do, T.-T. T. Do, A. V. Hoang, K. T. Dinh, M.-D. Phan, H. Giang, L. S. Tran, Liquid biopsy uncovers distinct patterns of DNA methylation and copy number

changes in NSCLC patients with different EGFR-TKI resistant mutations. *Sci. Rep.* **11**, 1–12 (2021).

21. V. A. Adalsteinsson, G. Ha, S. S. Freeman, A. D. Choudhury, D. G. Stover, H. A. Parsons, G. Gydush, S. C. Reed, D. Rotem, J. Rhoades, D. Loginov, D. Livitz, D. Rosebrock, I. Leshchiner, J. Kim, C. Stewart, M. Rosenberg, J. M. Francis, C.-Z. Zhang, O. Cohen, C. Oh, H. Ding, P. Polak, M. Lloyd, S. Mahmud, K. Helvie, M. S. Merrill, R. A. Santiago, E. P. O'Connor, S. H. Jeong, R. Leeson, R. M. Barry, J. F. Kramkowski, Z. Zhang, L. Polacek, J. G. Lohr, M. Schleicher, E. Lipscomb, A. Saltzman, N. M. Oliver, L. Marini, A. G. Waks, L. C. Harshman, S. M. Tolaney, E. M. Van Allen, E. P. Winer, N. U. Lin, M. Nakabayashi, M.-E. Taplin, C. M. Johannessen, L. A. Garraway, T. R. Golub, J. S. Boehm, N. Wagle, G. Getz, J. C. Love, M. Meyerson, Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).

22. A. M. Taylor, J. Shih, G. Ha, G. F. Gao, X. Zhang, A. C. Berger, S. E. Schumacher, C. Wang, H. Hu, J. Liu, A. J. Lazar, S. J. Caesar-Johnson, J. A. Demchok, I. Felau, M. Kasapi, M. L. Ferguson, C. M. Hutter, H. J. Sofia, R. Tarnuzzer, Z. Wang, L. Yang, J. C. Zenklusen, J. (julia) Zhang, S. Chudamani, J. Liu, L. Lolla, R. Naresh, T. Pihl, Q. Sun, Y. Wan, Y. Wu, J. Cho, T. DeFreitas, S. Frazer, N. Gehlenborg, G. Getz, D. I. Heiman, J. Kim, M. S. Lawrence, P. Lin, S. Meier, M. S. Noble, G. Saksena, D. Voet, H. Zhang, B. Bernard, N. Chambwe, V. Dhankani, T. Knijnenburg, R. Kramer, K. Leinonen, Y. Liu, M. Miller, S. Reynolds, I. Shmulevich, V. Thorsson, W. Zhang, R. Akbani, B. M. Broom, A. M. Hegde, Z. Ju, R. S. Kanchi, A. Korkut, J. Li, H. Liang, S. Ling, W. Liu, Y. Lu, G. B. Mills, K.-S. Ng, A. Rao, M. Ryan, J. Wang, J. N. Weinstein, J. Zhang, A. Abeshouse, J. Armenia, D. Chakravarty, W. K. Chatila, I. de Bruijn, J. Gao, B. E. Gross, Z. J. Heins, R. Kundra, K. La, M. Ladanyi, A. Luna, M. G. Nissan, A. Ochoa, S. M. Phillips, E. Reznik, F. Sanchez-Vega, C. Sander, N. Schultz, R. Sheridan, S. O. Sumer, Y. Sun, B. S. Taylor, J. Wang, H. Zhang, P. Anur, M. Peto, P. Spellman, C. Benz, J. M. Stuart, C. K. Wong, C. Yau, D. N. Hayes, J. S. Parker, M. D. Wilkerson, A. Ally, M. Balasundaram, R. Bowlby, D. Brooks, R. Carlsen, E. Chuah, N. Dhalla, R. Holt, S. J. M. Jones, K. Kasaian, D. Lee, Y. Ma, M. A. Marra, M. Mayo, R. A. Moore, A. J. Mungall, K. Mungall, A. G. Robertson, S. Sadeghi, J. E. Schein, P. Sipahimalani, A. Tam, N. Thiessen, K. Tse, T. Wong, A. C. Berger, R. Beroukhim, A. D. Cherniack, C. Cibulskis, S. B. Gabriel, G. F. Gao, G. Ha, M. Meyerson, S. E. Schumacher, J. Shih, M. H. Kucherlapati, R. S. Kucherlapati, S. Baylin, L. Cope, L. Danilova, M. S. Bootwalla, P. H. Lai, D. T. Maglinte, D. J. Van Den Berg, D. J. Weisenberger, J. T. Auman, S. Balu, T. Bodenheimer, C. Fan, K. A. Hoadley, A. P. Hoyle, S. R. Jefferys, C. D. Jones, S. Meng, P. A. Mieczkowski, L. E. Mose, A. H. Perou, C. M. Perou, J. Roach, Y. Shi, J. V. Simons, T. Skelly, M. G. Soloway, D. Tan, U. Veluvolu, H. Fan, T. Hinoue, P. W. Laird, H. Shen, W. Zhou, M. Bellair, K. Chang, K. Covington, C. J. Creighton, H. Dinh, H. Doddapaneni, L. A. Donehower, J. Drummond, R. A. Gibbs, R. Glenn, W. Hale, Y. Han, J. Hu, V. Korchina, S. Lee, L. Lewis, W. Li, X. Liu, M. Morgan, D. Morton, D. Muzny, J. Santibanez, M. Sheth, E. Shinbrot, L. Wang, M. Wang, D. A. Wheeler, L. Xi, F. Zhao, J. Hess, E. L. Appelbaum, M. Bailey, M. G. Cordes, L. Ding, C. C. Fronick, L. A. Fulton, R. S. Fulton, C. Kandoth, E. R. Mardis, M. D. McLellan, C. A. Miller, H. K. Schmidt, R. K. Wilson, D. Crain, E. Curley, J. Gardner, K. Lau, D. Mallery, S. Morris, J. Paulauskis, R. Penny, C. Shelton, T. Shelton, M. Sherman, E. Thompson, P. Yena, J. Bowen, J. M. Gastier-Foster, M. Gerken, K. M. Leraas, T. M. Lichtenberg, N. C. Ramirez, L. Wise, E. Zmuda, N. Corcoran, T. Costello, C. Hovens, A. L. Carvalho, A. C. de Carvalho, J. H. Fregnani, A. Longatto-Filho, R. M. Reis, C. Scapulatempo-Neto, H. C. S. Silveira, D. O. Vidal, A. Burnette, J. Eschbacher, B. Hermes, A. Noss, R. Singh, M. L. Anderson, P. D. Castro, M. Ittmann, D. Huntsman, B. Kohl, X. Le, R. Thorp, C. Andry, E. R. Duffy, V. Lyadov, O.

Paklina, G. Setdikova, A. Shabunin, M. Tavobilov, C. McPherson, R. Warnick, R. Berkowitz, D. Cramer, C. Feltmate, N. Horowitz, A. Kibel, M. Muto, C. P. Raut, A. Malykh, J. S. Barnholtz-Sloan, W. Barrett, K. Devine, J. Fulop, Q. T. Ostrom, K. Shimmel, Y. Wolinsky, A. E. Sloan, A. De Rose, F. Giuliante, M. Goodman, B. Y. Karlan, C. H. Hagedorn, J. Eckman, J. Harr, J. Myers, K. Tucker, L. A. Zach, B. Deyarmin, H. Hu, L. Kvecher, C. Larson, R. J. Mural, S. Somiari, A. Vicha, T. Zelinka, J. Bennett, M. Iacocca, B. Rabeno, P. Swanson, M. Latour, L. Lacombe, B. Têtu, A. Bergeron, M. McGraw, S. M. Staugaitis, J. Chabot, H. Hibshoosh, A. Sepulveda, T. Su, T. Wang, O. Potapova, O. Voronina, L. Desjardins, O. Mariani, S. Roman-Roman, X. Sastre, M.-H. Stern, F. Cheng, S. Signoretti, A. Berchuck, D. Bigner, E. Lipp, J. Marks, S. McCall, R. McLendon, A. Secord, A. Sharp, M. Behera, D. J. Brat, A. Chen, K. Delman, S. Force, F. Khuri, K. Magliocca, S. Maithel, J. J. Olson, T. Owonikoko, A. Pickens, S. Ramalingam, D. M. Shin, G. Sica, E. G. Van Meir, H. Zhang, W. Eijckenboom, A. Gillis, E. Korpershoek, L. Looijenga, W. Oosterhuis, H. Stoop, K. E. van Kessel, E. C. Zwarthoff, C. Calatozzolo, L. Cuppini, S. Cuzzubbo, F. DiMeco, G. Finocchiaro, L. Mattei, A. Perin, B. Pollo, C. Chen, J. Houck, P. Lohavanichbutr, A. Hartmann, C. Stoehr, R. Stoehr, H. Taubert, S. Wach, B. Wullich, W. Kycler, D. Murawa, M. Wiznerowicz, K. Chung, W. J. Edenfield, J. Martin, E. Baudin, G. Bubley, R. Bueno, A. De Rienzo, W. G. Richards, S. Kalkanis, T. Mikkelsen, H. Noushmehr, L. Scarpace, N. Girard, M. Aymerich, E. Campo, E. Giné, A. L. Guillermo, N. Van Bang, P. T. Hanh, B. D. Phu, Y. Tang, H. Colman, K. Evason, P. R. Dottino, J. A. Martignetti, H. Gabra, H. Juhl, T. Akeredolu, S. Stepa, D. Hoon, K. Ahn, K. J. Kang, F. Beuschlein, A. Breggia, M. Birrer, D. Bell, M. Borad, A. H. Bryce, E. Castle, V. Chandan, J. Cheville, J. A. Copland, M. Farnell, T. Flotte, N. Giama, T. Ho, M. Kendrick, J.-P. Kocher, K. Kopp, C. Moser, D. Nagorney, D. O'Brien, B. P. O'Neill, T. Patel, G. Petersen, F. Que, M. Rivera, L. Roberts, R. Smallridge, T. Smyrk, M. Stanton, R. H. Thompson, M. Torbenson, J. D. Yang, L. Zhang, F. Brimo, J. A. Ajani, A. M. Angulo Gonzalez, C. Behrens, J. Bondaruk, R. Broaddus, B. Czerniak, B. Esmaeli, J. Fujimoto, J. Gershenwald, C. Guo, A. J. Lazar, C. Logothetis, F. Meric-Bernstam, C. Moran, L. Ramondetta, D. Rice, A. Sood, P. Tamboli, T. Thompson, P. Troncoso, A. Tsao, I. Wistuba, C. Carter, L. Haydu, P. Hersey, V. Jakrot, H. Kakavand, R. Kefford, K. Lee, G. Long, G. Mann, M. Quinn, R. Saw, R. Scolyer, K. Shannon, A. Spillane, J. Stretch, M. Synott, J. Thompson, J. Wilmott, H. Al-Ahmadie, T. A. Chan, R. Ghossein, A. Gopalan, D. A. Levine, V. Reuter, S. Singer, B. Singh, N. V. Tien, T. Broudy, C. Mirsaidi, P. Nair, P. Drwiega, J. Miller, J. Smith, H. Zaren, J.-W. Park, N. P. Hung, E. Kebebew, W. M. Linehan, A. R. Metwalli, K. Pacak, P. A. Pinto, M. Schiffman, L. S. Schmidt, C. D. Vocke, N. Wentzensen, R. Worrell, H. Yang, M. Moncrieff, C. Goparaju, J. Melamed, H. Pass, N. Botnariuc, I. Caraman, M. Cernat, I. Chemencedji, A. Clipca, S. Doruc, G. Gorincioi, S. Mura, M. Pirtac, I. Stancul, D. Tcaciuc, M. Albert, I. Alexopoulou, A. Arnaout, J. Bartlett, J. Engel, S. Gilbert, J. Parfitt, H. Sekhon, G. Thomas, D. M. Rassl, R. C. Rintoul, C. Bifulco, R. Tamakawa, W. Urba, N. Hayward, H. Timmers, A. Antenucci, F. Facciolo, G. Grazi, M. Marino, R. Merola, R. de Krijger, A.-P. Gimenez-Roqueplo, A. Piché, S. Chevalier, G. McKercher, K. Birsoy, G. Barnett, C. Brewer, C. Farver, T. Naska, N. A. Pennell, D. Raymond, C. Schilero, K. Smolenski, F. Williams, C. Morrison, J. A. Borgia, M. J. Liptay, M. Pool, C. W. Seder, K. Junker, L. Omberg, M. Dinkin, G. Manikhas, D. Alvaro, M. C. Bragazzi, V. Cardinale, G. Carpino, E. Gaudio, D. Chesla, S. Cottingham, M. Dubina, F. Moiseenko, R. Dhanasekaran, K.-F. Becker, K.-P. Janssen, J. Slotta-Huspenina, M. H. Abdel-Rahman, D. Aziz, S. Bell, C. M. Cebulla, A. Davis, R. Duell, J. B. Elder, J. Hilty, B. Kumar, J. Lang, N. L. Lehman, R. Mandt, P. Nguyen, R. Pilarski, K. Rai, L. Schoenfield, K. Senecal, P. Wakely, P. Hansen, R. Lechan, J. Powers, A. Tischler, W. E. Grizzle, K. C. Sexton, A. Kastl, J. Henderson, S. Porten, J. Waldmann, M. Fassnacht, S. L. Asa, D. Schadendorf, M. Couce, M. Graefen, H. Huland, G. Sauter, T. Schlomm, R. Simon, P. Tennstedt, O. Olabode, M. Nelson, O. Bathe, P. R. Carroll, J. M. Chan, P. Disaia, P. Glenn,

R. K. Kelley, C. N. Landen, J. Phillips, M. Prados, J. Simko, K. Smith-McCune, S. VandenBerg, K. Roggin, A. Fehrenbach, A. Kendler, S. Sifri, R. Steele, A. Jimeno, F. Carey, I. Forgie, M. Mannelli, M. Carney, B. Hernandez, B. Campos, C. Herold-Mende, C. Jungk, A. Unterberg, A. von Deimling, A. Bossler, J. Galbraith, L. Jacobus, M. Knudson, T. Knutson, D. Ma, M. Milhem, R. Sigmund, A. K. Godwin, R. Madan, H. G. Rosenthal, C. Adebamowo, S. N. Adebamowo, A. Boussioutas, D. Beer, T. Giordano, A.-M. Mes-Masson, F. Saad, T. Bocklage, L. Landrum, R. Mannel, K. Moore, K. Moxley, R. Postier, J. Walker, R. Zuna, M. Feldman, F. Valdivieso, R. Dhir, J. Luketich, E. M. Mora Pinero, M. Quintero-Aguilo, C. G. Carlotti, J. S. Dos Santos, R. Kemp, A. Sankarankuty, D. Tirapelli, J. Catto, K. Agnew, E. Swisher, J. Creaney, B. Robinson, C. S. Shelley, E. M. Godwin, S. Kendall, C. Shipman, C. Bradford, T. Carey, A. Haddad, J. Moyer, L. Peterson, M. Prince, L. Rozek, G. Wolf, R. Bowman, K. M. Fong, I. Yang, R. Korst, W. K. Rathmell, J. L. Fantacone-Campbell, J. A. Hooke, A. J. Kovatich, C. D. Shriver, J. DiPersio, B. Drake, R. Govindan, S. Heath, T. Ley, B. Van Tine, P. Westervelt, M. A. Rubin, J. I. Lee, N. D. Aredes, A. Mariamidze, A. D. Cherniack, R. Beroukhim, M. Meyerson, Genomic and Functional Approaches to Understanding Cancer Aneuploidy. *Cancer Cell*. **33**, 676–689.e3 (2018).

23. A. Lubotzky, H. Zemmour, D. Neiman, M. Gotkine, N. Loyfer, S. Piyanzin, B.-L. Ochana, R. Lehmann-Werman, D. Cohen, J. Moss, J. Magenheim, M. F. Loftus, L. Brais, K. Ng, R. Mostoslavsky, B. M. Wolpin, A. Zick, M. Maoz, A. Grinshpun, A. Kustanovich, C. Makranz, J. E. Cohen, T. Peretz, A. Hubert, M. Temper, A. Salah, S. Avniel-Polak, S. Grozinsky-Glasberg, K. L. Spalding, A. Rokach, T. Kaplan, B. Glaser, R. Shemer, Y. Dor, Liquid biopsy reveals collateral tissue damage in cancer. *JCI Insight*. **7** (2022), doi:10.1172/jci.insight.153559.

24. P. Ni, N. Huang, Z. Zhang, D.-P. Wang, F. Liang, Y. Miao, C.-L. Xiao, F. Luo, J. Wang, DeepSignal: detecting DNA methylation state from Nanopore sequencing reads using deep-learning. *Bioinformatics*. **35**, 4586–4595 (2019).

25. K. Zhang, J. D. Hocker, M. Miller, X. Hou, J. Chiou, O. B. Poirion, Y. Qiu, Y. E. Li, K. J. Gaulton, A. Wang, S. Preissl, B. Ren, A single-cell atlas of chromatin accessibility in the human genome. *Cell*. **184**, 5985–6001.e19 (2021).

26. Y. Maeda, V. Davé, J. A. Whitsett, Transcriptional control of lung morphogenesis. *Physiol. Rev.* **87**, 219–244 (2007).

27. G. Eraslan, E. Drokhlyansky, S. Anand, A. Subramanian, E. Fiskin, M. Slyper, J. Wang, N. Van Wittenberghe, J. M. Rouhana, J. Waldman, O. Ashenberg, D. Dionne, T. S. Win, M. S. Cuoco, O. Kuksenko, P. A. Branton, J. L. Marshall, A. Greka, G. Getz, A. V. Segrè, F. Aguet, O. Rozenblatt-Rosen, K. G. Ardlie, A. Regev, Single-nucleus cross-tissue molecular reference maps to decipher disease gene function. *bioRxiv* (2021), p. 2021.07.19.452954.

28. M. R. Corces, J. M. Granja, S. Shams, B. H. Louie, J. A. Seoane, W. Zhou, T. C. Silva, C. Groeneveld, C. K. Wong, S. W. Cho, A. T. Satpathy, M. R. Mumbach, K. A. Hoadley, A. G. Robertson, N. C. Sheffield, I. Felau, M. A. A. Castro, B. P. Berman, L. M. Staudt, J. C. Zenklusen, P. W. Laird, C. Curtis, C. G. A. A. Network, W. J. Greenleaf, H. Y. Chang, The chromatin accessibility landscape of primary human cancers. *Science*. **362** (2018), doi:10.1126/science.aav1898.

29. W. Zhou, H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird, B. P. Berman, DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* **50**, 591–602 (2018).

30. K. C. A. 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. D. 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).

31. I. F. López-Moyado, A. Tsagaratou, H. Yuita, H. Seo, B. Delatte, S. Heinz, C. Benner, A. Rao, Paradoxical association of TET loss of function with genome-wide DNA hypomethylation. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 16933–16942 (2019).

32. N. Farhangdoost, C. Horth, B. Hu, E. Bareke, X. Chen, Y. Li, M. Coradin, B. A. Garcia, C. Lu, J. Majewski, Chromatin dysregulation associated with NSD1 mutation in head and neck squamous cell carcinoma. *Cell Rep.* **34**, 108769 (2021).

33. T. K. Kelly, Y. Liu, F. D. Lay, G. Liang, B. P. Berman, P. A. Jones, Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506 (2012).

34. 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. 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, Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci. Transl. Med.* **10** (2018), doi:10.1126/scitranslmed.aat4921.

35. P. Jiang, K. Sun, W. Peng, S. H. Cheng, M. Ni, P. C. Yeung, M. M. S. Heung, T. Xie, H. Shang, Z. Zhou, R. W. Y. Chan, J. Wong, V. W. S. Wong, L. C. Poon, T. Y. Leung, W. K. Jacky Lam, J. Y. K. Chan, H. L. Y. Chan, K. C. Allen Chan, R. W. K. Chiu, Y. M. Dennis Lo, Plasma DNA End-Motif Profiling as a Fragmentomic Marker in Cancer, Pregnancy, and Transplantation. *Cancer Discov.* **10**, 664–673 (2020).

36. R. W. Y. Chan, L. Serpas, M. Ni, S. Volpi, L. T. Hiraki, L.-S. Tam, A. Rashidfarrokhi, P. C. H. Wong, L. H. P. Tam, Y. Wang, P. Jiang, A. S. H. Cheng, W. Peng, D. S. C. Han, P. P. P. Tse, P. K. Lau, W.-S. Lee, A. Magnasco, E. Buti, V. Sisirak, N. AlMutairi, K. C. A. Chan, R. W. K. Chiu, B. Reizis, Y. M. D. Lo, Plasma DNA Profile Associated with DNASE1L3 Gene Mutations: Clinical Observations, Relationships to Nuclease Substrate Preference, and In Vivo Correction. *Am. J. Hum. Genet.* **107**, 882–894 (2020).

37. J. Foox, S. W. Tighe, C. M. Nicolet, J. M. Zook, M. Byrska-Bishop, W. E. Clarke, M. M. Khayat, M. Mahmoud, P. K. Laaguiby, Z. T. Herbert, D. Warner, G. S. Grills, J. Jen, S. Levy, J. Xiang, A. Alonso, X. Zhao, W. Zhang, F. Teng, Y. Zhao, H. Lu, G. P. Schroth, G. Narzisi, W. Farmerie, F. J. Sedlazeck, D. A. Baldwin, C. E. Mason, Performance assessment of DNA sequencing platforms in the ABRF Next-Generation Sequencing Study. *Nat. Biotechnol.* **39**, 1129–1140 (2021).

38. 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).

39. A. J. Widman, M. Shah, N. Øgaard, C. C. Khamnei, A. Frydendahl, A. Deshpande, A. Arora, M. Zhang, D. Halmos, J. Bass, T. Langanay, S. Rajagopalan, Z. Steinsnyder, W. Liao, M. H. Rasmussen, S. Ø. Jensen, J. Nors, C. Therkildsen, J. Sotelo, R. Brand, R. H. Shah, A. P. Cheng, C. Maher, L. Spain, K. Krause, D. T. Frederick, M. S. Malbari, M. Marton, D. Manaa, L. Winterkorn, M. K. Callahan, G. Boland, J. D. Wolchok, A. Saxena, S. Turajlic, M. Imielinski, M. F. Berger, N. K. Altorki, M. A. Postow, N. Robine, C. L. Andersen, D. A. Landau, Machine learning guided signal enrichment for ultrasensitive plasma tumor burden monitoring. *bioRxiv* (2022), p. 2022.01.17.476508.

40. 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**, 1–14 (2020).

41. R. R. C. Cuadrat, A. Kratzer, H. G. Arnal, K. Wreczycka, A. Blume, V. Ebenal, T. Mauno, B. Osberg, M. Moobed, J. Hartung, C. Seppelt, D. Meteva, A. Haghikia, D. Leistner, U. Landmesser, A. Akalin, Cardiovascular disease biomarkers derived from circulating cell-free DNA methylation. *medRxiv*, 2021.11.05.21265870 (2021).

42. 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**, E5503–12 (2015).

43. G. G. Faust, I. M. Hall, SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*. **30**, 2503–2505 (2014).

44. T. Baslan, S. Kovaka, F. J. Sedlazeck, Y. Zhang, R. Wappel, S. W. Lowe, S. Goodwin, M. C. Schatz, High resolution copy number inference in cancer using short-molecule nanopore sequencing. *bioRxiv* (2020), p. 2020.12.28.424602.

45. T. C. Silva, S. G. Coetzee, N. Gull, L. Yao, D. J. Hazelett, H. Noushmehr, D.-C. Lin, B. P. Berman, ELMER v.2: an R/Bioconductor package to reconstruct gene regulatory networks from DNA methylation and transcriptome profiles. *Bioinformatics*. **35**, 1974–1977 (2019).

46. K. Zhang, J. D. Hocker, M. Miller, X. Hou, J. Chiou, O. B. Poirion, Y. Qiu, Y. E. Li, K. J. Gaulton, A. Wang, S. Preissl, B. Ren, A cell atlas of chromatin accessibility across 25 adult human tissues. *bioRxiv* (2021), doi:10.1101/2021.02.17.431699.

Figure 1

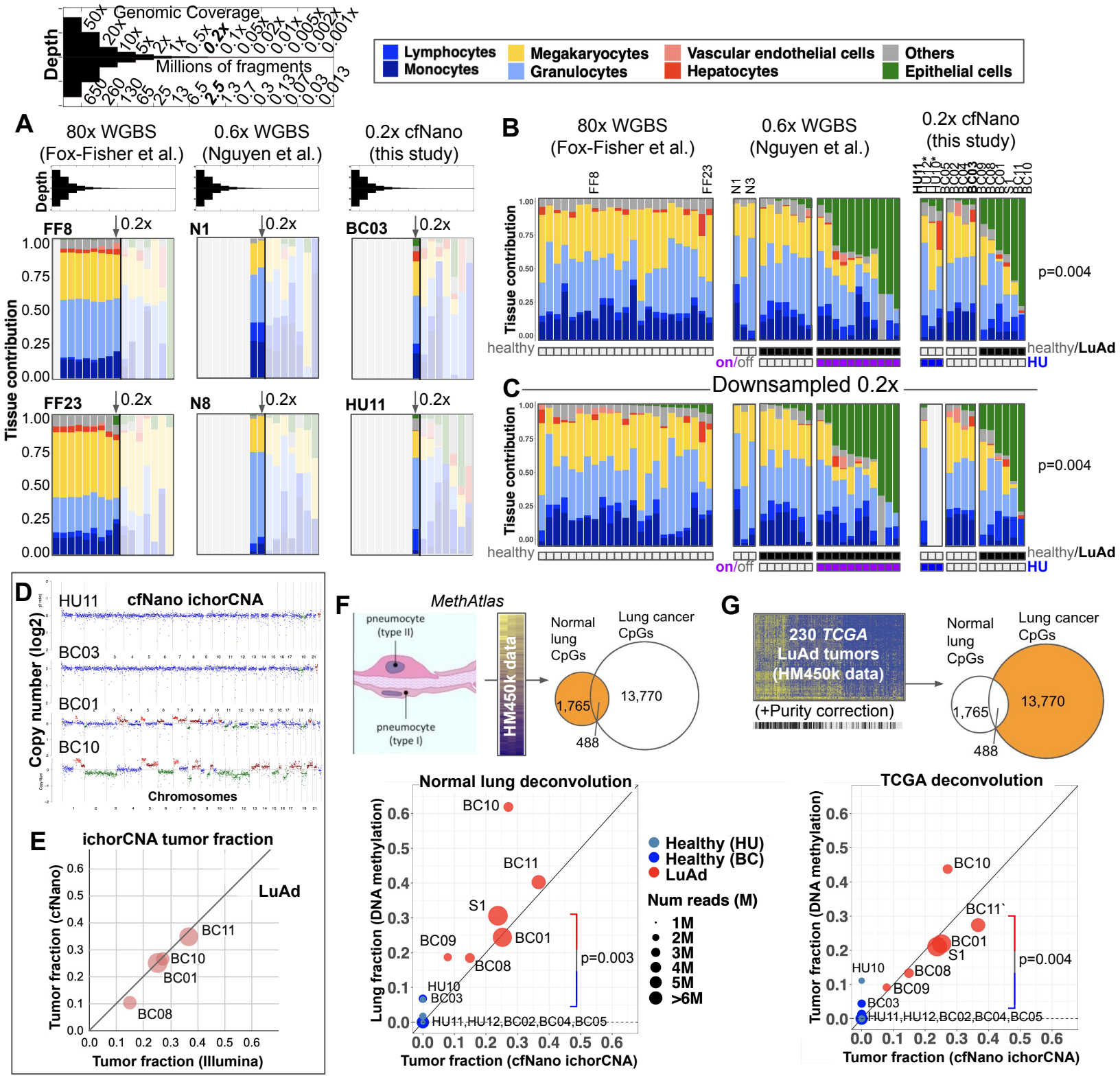

Genomic Coverage
Depth
Millions of fragments
Lymphocytes
Monocytes
Megakaryocytes
Granulocytes
Vascular endothelial cells
Hepatocytes
Others
Epithelial cells
A
80x WGBS (Fox-Fisher et al.)
0.6x WGBS (Nguyen et al.)
0.2x cfNano (this study)
FF8
N1
BC03
FF23
N8
HU11
0.2x
Tissue contribution
B
C
Downsampled 0.2x
p=0.004
healthy
on/off
healthy/LuAd
HU
D
cfNano ichorCNA
Copy number (log2)
Chromosomes
E
ichorCNA tumor fraction
LuAd
Tumor fraction (cfNano)
Tumor fraction (Illumina)
F
MethAtlas
pneumocyte (type II)
pneumocyte (type I)
HM450k data
Normal lung CpGs
Lung cancer CpGs
1,765
13,770
488
G
230 TCGA LuAd tumors (HM450k data)
(+Purity correction)
Normal lung deconvolution
Lung fraction (DNA methylation)
Tumor fraction (cfNano ichorCNA)
p=0.003
Healthy (HU)
Healthy (BC)
LuAd
Num reads (M)
1M
2M
3M
4M
5M
>6M
TCGA deconvolution
p=0.004

Figure 2

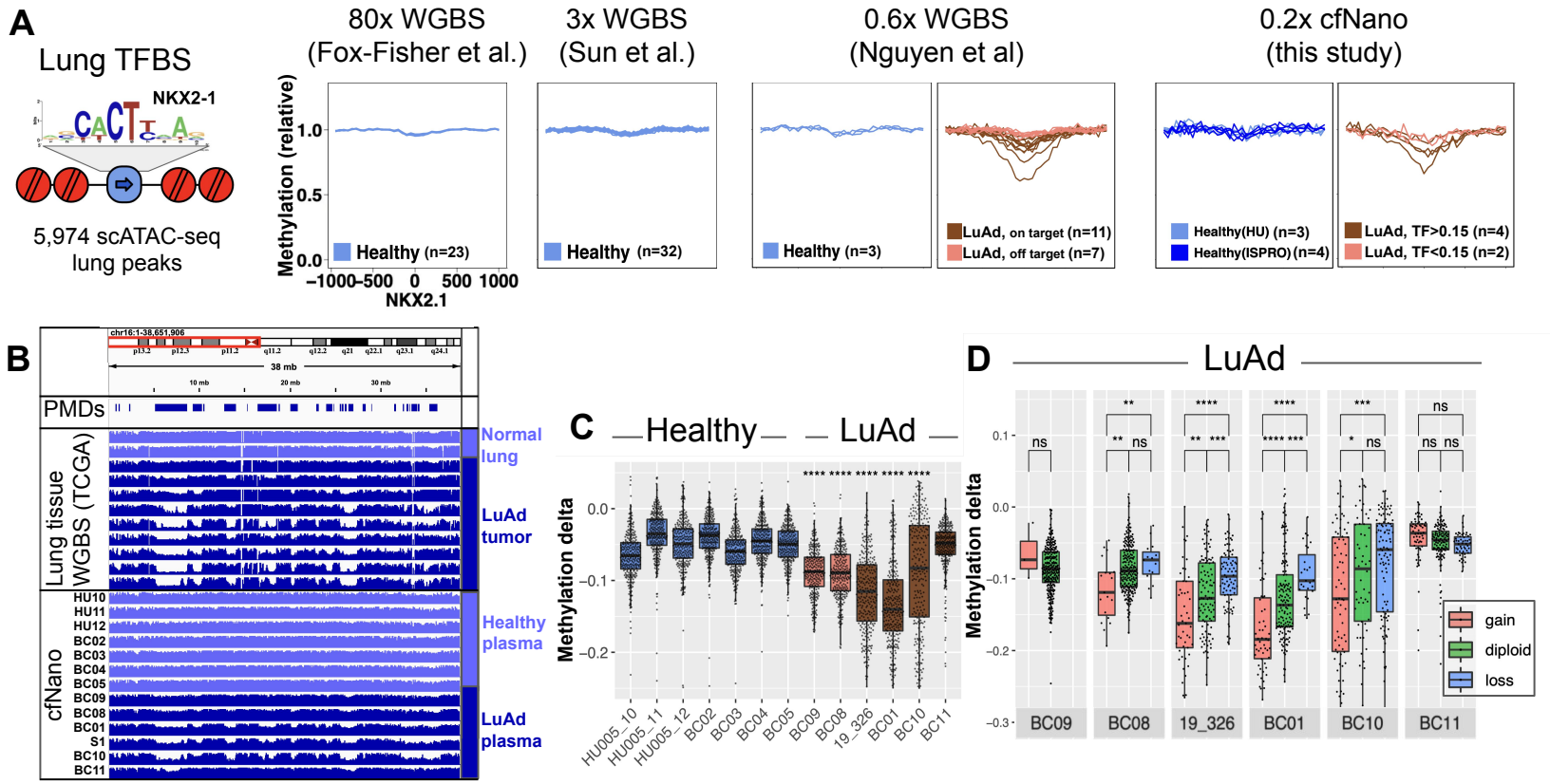

A
Lung TFBS
NKX2-1
5,974 scATAC-seq lung peaks
80x WGBS (Fox-Fisher et al.)
3x WGBS (Sun et al.)
0.6x WGBS (Nguyen et al)
0.2x cfNano (this study)
Methylation (relative)
Healthy (n=23)
Healthy (n=32)
Healthy (n=3)
LuAd, on target (n=11)
LuAd, off target (n=7)
Healthy(HU) (n=3)
Healthy(ISPRO) (n=4)
LuAd, TF>0.15 (n=4)
LuAd, TF<0.15 (n=2)
NKX2.1
B
PMDs
Lung tissue WGBS (TCGA)
cfNano
Normal lung
LuAd tumor
Healthy plasma
LuAd plasma
C
Healthy
LuAd
Methylation delta
D
LuAd
gain
diploid
loss
BC09
BC08
19_326
BC01
BC10
BC11

Figure 3

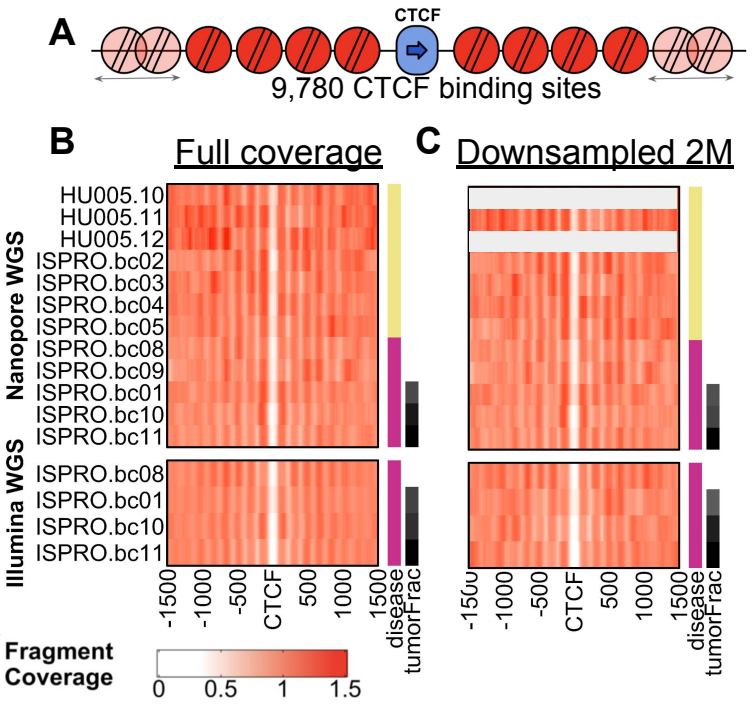
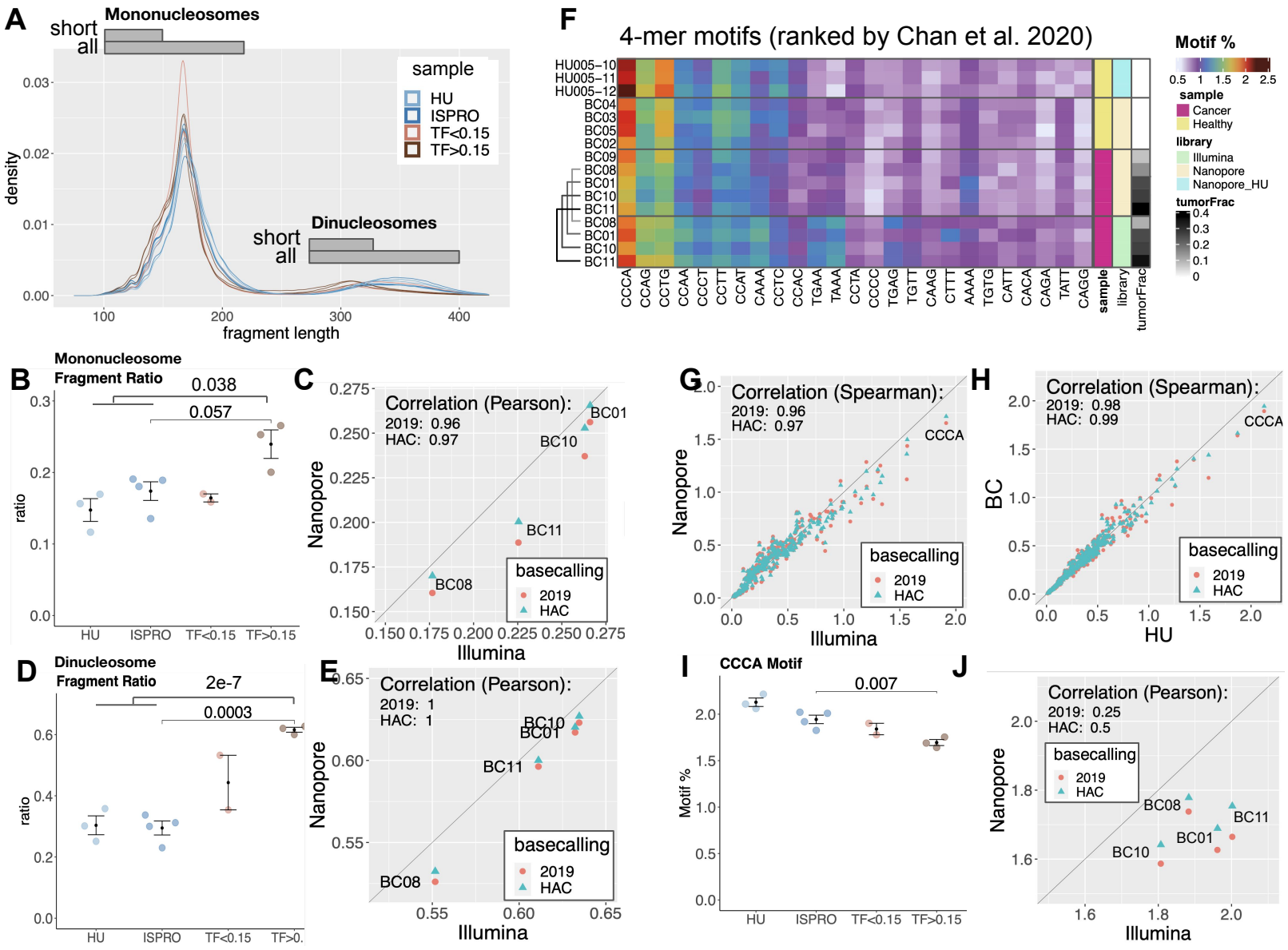

A
CTCF
9,780 CTCF binding sites
B
Full coverage
C
Downsampled 2M
Nanopore WGS
HU005.10
HU005.11
HU005.12
ISPRO.bc02
ISPRO.bc03
ISPRO.bc04
ISPRO.bc05
ISPRO.bc08
ISPRO.bc09
ISPRO.bc01
ISPRO.bc10
ISPRO.bc11
Illumina WGS
ISPRO.bc08
ISPRO.bc01
ISPRO.bc10
ISPRO.bc11
-1500
-1000
-500
CTCF
500
1000
1500
disease
tumorFrac
Fragment Coverage
0
0.5
1
1.5

Figure 4


A
Mononucleosomes
short
all
sample
HU
ISPRO
TF<0.15
TF>0.15
Dinucleosomes
short
all
density
fragment length
F
4-mer motifs (ranked by Chan et al. 2020)
Motif %
sample
Cancer
Healthy
library
Illumina
Nanopore
Nanopore_HU
tumorFrac
HU005-10
HU005-11
HU005-12
BC04
BC03
BC05
BC02
BC09
BC08
BC01
BC10
BC11
CCCA
CCAG
CCTG
CCAA
CCCT
CCTT
CCAT
CAAA
CCTC
CCAC
TGAA
TAAA
CCTA
CCCC
TGAG
TGTT
CAAG
CTTT
AAAA
TGTG
CATT
CACA
CAGA
TATT
CAGG
B
Mononucleosome
Fragment Ratio
0.038
0.057
ratio
HU
ISPRO
TF<0.15
TF>0.15
C
Correlation (Pearson):
2019: 0.96
HAC: 0.97
BC01
BC10
BC11
BC08
basecalling
2019
HAC
Nanopore
Illumina
G
Correlation (Spearman):
2019: 0.96
HAC: 0.97
CCCA
basecalling
2019
HAC
Nanopore
Illumina
H
Correlation (Spearman):
2019: 0.98
HAC: 0.99
CCCA
basecalling
2019
HAC
BC
HU
D
Dinucleosome
Fragment Ratio
2e-7
0.0003
ratio
HU
ISPRO
TF<0.15
TF>0.
E
Correlation (Pearson):
2019: 1
HAC: 1
BC10
BC01
BC11
BC08
basecalling
2019
HAC
Nanopore
Illumina
I
CCCA Motif
0.007
Motif %
HU
ISPRO
TF<0.15
TF>0.15
J
Correlation (Pearson):
2019: 0.25
HAC: 0.5
basecalling
2019
HAC
BC08
BC11
BC01
BC10
Nanopore
Illumina