Spatial co-fragmentation pattern of cell-free DNA recapitulates *in vivo* chromatin organization and identifies tissues-of-origin

Yaping Liu [1 * #] , Tzu-Yu Liu [1], David E. Weinberg[1], Brandon W. White[2], Chris J. De La Torre[3], Catherine L. Tan[3], Anthony D. Schmitt[3], Siddarth Selvaraj[3], Vy Tran[4], Louise C. Laurent[4], Luc Cabel[5], François-Clément Bidard[5], Girish Putcha [1,#], Imran S. Haque [2, #]

1. Freenome Inc., South San Francisco, CA, USA
2. Former employee of Freenome Inc., South San Francisco, CA, USA
3. Arima Genomics Inc., San Diego, CA, USA
4. Department of Reproductive Medicine, University of California San Diego, San Diego, CA, USA
5. Institut Curie, Paris, France

*Current affiliations: Division of Human Genetics, Cincinnati Children's Hospital Medical Center & Department of Pediatrics, University of Cincinnati College of Medicine, Cincinnati, OH

Corresponding authors

Contact information:

Email: authors@freenome.com

Keyword: Cell-free DNA, chromosome conformation changes, *in vivo*, tissues-of-origin, whole genome sequencing

ABSTRACT

Three-dimensional chromatin organization varies across cell types and is essential for gene regulation. However, current technologies are unable to assess *in vivo* genome-wide chromatin organization non-invasively. Here we show that distant correlations in the fragment length of cell-free DNA (cfDNA) recapitulate three-dimensional chromatin organization. The inferred organization is highly concordant with that measured by Hi-C in white blood cells from healthy donors, and is not explained by technical bias or sequence composition. Furthermore, the inferred organization reflects different genomic organization in the various cell types contributing to cfDNA, allowing identification and quantification of tissues of origin. This approach is concordant with previous methods, but with more complete representations of cfDNA. Our results, demonstrated in cfDNA from healthy individuals and cancer patients, may enable noninvasive monitoring of *in vivo* genome organization and accurate quantification of cell death in different clinical conditions.

INTRODUCTION

Functional genomic elements that reside kilobases to megabases apart in linear distance can be brought into spatial proximity by chromatin folding(Sexton and Cavalli 2015; Sexton et al. 2007; Cremer and Cremer 2001; Bickmore 2013; Misteli 2007). In fixed cells, fluorescence *in situ* hybridization combined with super-resolution microscopy and chromosome conformation capture followed by next-generation sequencing (Hi-C) are able to capture three-dimensional chromatin organization between pairs of loci(Lieberman-Aiden et al. 2009; Rao et al. 2014; Dixon et al. 2012), which provides a static snapshot of genome compaction and organization in different cellular states(Ma et al. 2018). However, *in vivo* dynamic genome-wide chromatin organization in patients is poorly understood due to the limitations of current technologies. Recently, cell-free DNA (cfDNA) in blood has been shown to be a promising biomarker of

longitudinal *in vivo* information on the genetic and local epigenetic changes within tissues(Snyder et al. 2016; Christina Fan et al. 2008; Ivanov et al. 2015; Ulz et al. 2016; Wan et al. 2017; Heitzer et al. 2018); hence we enquired whether chromatin conformation data could also be obtained from the same material.

The fragment lengths of cfDNA are not uniform across the genome and are influenced by the local epigenetic environment and different physiological conditions(Ivanov et al. 2015; Snyder et al. 2016). High-resolution DNA fragment length profiling of cfDNA has revealed a predominant 167-base pair (bp) peak and 10-bp periodicity pattern in cfDNA fragments, which is highly correlated with local nucleosomal structure(Snyder et al. 2016). The size of cfDNA fragments differ according to their tissues-of-origin, including differences between fetal and maternal DNA, and between tumor and non-tumor derived DNA(Sun et al. 2018; Chan et al. 2004; Sun et al. 2015). However, long-range co-variation of fragment lengths from linearly distant genomic regions in spatial proximity is not yet well understood.

RESULTS

Co-fragmentation pattern of cfDNA recapitulates the 3D genome

Here, we hypothesized that the three-dimensional structure of the genome may be reflected in long-range correlated fragment lengths of cfDNA which could be derived from the co-fragmentation of genomic DNA in spatial proximity (Figure 1a-b). To test this hypothesis, we performed paired-end whole-genome sequencing (WGS) of cfDNA from 100 healthy individuals. We obtained an average of 395 million paired-end reads for each sample (~12.8X coverage), which after quality control yielded 310 million high-quality paired-end reads for each sample on average (~10X coverage) (see Methods, Supplemental Table 1). Since most cfDNA molecules in healthy individuals originate from white blood cells (WBCs)(Lui et al. 2002), we also generated Arima-HiC data of WBCs isolated from the whole blood of two healthy individuals for comparison. We then used the previously described HiCCUPS method to call 3,763 autosomal

chromatin loops from the Hi-C data (see Methods)(Rao et al. 2014; Durand et al. 2016). As hypothesized, we observed that the median cfDNA fragment lengths were highly correlated between the two anchors of chromatin loops compared to random pairs of regions with matched genomic distance (Figure 1c, Supplemental Figure 1).

To quantify the long-range correlated regions of cfDNA fragment lengths across the whole genome in an unbiased manner, we divided the autosomes into 500-kilobase (kb) non-overlapping bins and used fragment lengths to calculate a "normalized fragmentation score" at each bin for each individual (see Methods). We then calculated the Pearson correlation coefficient between the fragmentation scores in each pair of bins across all individuals (as shown in Figure 1c). We found similar patterns between the correlation matrix derived from cfDNA fragmentation and the correlation matrix derived from WBC Hi-C data (Figure 2a-b, Supplemental Figure 3). To quantify the degree of similarity, we calculated the Pearson correlation at the "pixel level" (see Methods) between the two matrices, which confirmed that there is a significant similarity (Pearson correlation $r = 0.73$, $p < 2.2e^{-16}$, Supplemental Figure 3), suggesting that higher-order chromatin organization features measured directly from Hi-C are indeed captured by cfDNA co-fragmentation patterns. As a continuation of our analyses, we defined A/B compartments from Hi-C data using previously described methods(Lieberman-Aiden et al. 2009), and also applied this method on the co-fragment correlation matrix inferred from cfDNA co-fragmentation. We observed even higher concordance in A/B compartmentalization ("compartment level") between Hi-C and the inferred co-fragment correlation matrix from cfDNA (Figure 2b-c, Pearson correlation $r = 0.87$, $p < 2.2e^{-16}$) but not between Hi-C and cfDNA fragment lengths alone (Figure 2b). We refer to our approach of recapitulating Hi-C genome conformation from cfDNA WGS as FREE-C (**FR**agmentation **E**valuation of **E**pigenetics from **C**fDNA sequencing).

The multi-sample FREE-C method described above requires cfDNA WGS datasets from multiple individuals to quantify spatial co-fragmentation. To expand the application of FREE-C to individual samples, we developed an alternative approach to quantify spatial co-fragmentation within a single sample. For each sample, we performed a two-sample Kolmogorov–Smirnov test (K-S test) to measure the similarity of fragment length distributions between each pair of 500-kb bins (Supplemental Figure 2a). We again observed a high similarity between the fragment length distribution at two anchors for chromatin loops identified from the WBC Hi-C data compared to random paired regions with matched genomic distance (Supplemental Figure 2b-c, $p < 2.2e^{-16}$ in Bartlett's test for test of equal variance in Supplemental Figure 2c). Furthermore, for each sample, we observed a high correlation between WBC Hi-C and single-sample FREE-C at both the pixel and compartment level (Figure 2, Supplemental Figure 4).

FREE-C is robust to technical bias and sequence composition

To rule out possible technical bias caused by sequence composition, we first normalized fragment length in each bin with its mean GC content (G+C%). After regressing out GC content, we still observed high similarity between WBC Hi-C and multi-sample FREE-C (mean Pearson correlation r = 0.58, Figures 3a-b). As a negative control, we repeated the same analyses on publically available WGS data generated from randomly sheared genomic DNA (gDNA) of WBCs. Before regressing out GC content, we initially observed a relatively high similarity between Hi-C contact maps and gDNA fragmentation correlation maps (mean Pearson correlation r = 0.51, Figure 3a-b). However, after normalizing by GC content in gDNA, we only observed a low residual similarity between these maps (mean Pearson correlation r = 0.12, Figure 3a-b) and did not find a Hi-C-like structure in gDNA anymore. To further understand the non-linear interaction effect of GC content and mappability between paired bins, we applied a Gradient Boosting Machine (GBM) regression tree on FREE-C to regress out GC content and mappability in two-dimensional space (see Methods). After successfully regressing out the GC

content and mappability, we still observed significant residual similarity between Hi-C and both multi-sample (Pearson correlation r = 0.28, $p < 2.2e^{-16}$, n_estimator = 500, Figure 3c) and single-sample FREE-C (Pearson correlation r = 0.31, $p < 2.2e^{-16}$, n_estimator = 500, Figure 3c). In contrast, there was minimal residual similarity between Hi-C and both multi-sample (Pearson correlation r = 0.06, Figure 3c) and single-sample (Pearson correlation r = -0.02, Figure 3c) gDNA in the same model complexity. Further, we performed another negative control experiment: for each paired bin in FREE-C, we substituted one of the bins with a random bin from another chromosome that has the same GC content and mappability, and then we recalculated the fragmentation correlation map. By using the same GBM regression tree approach on this substituted FREE-C matrix, we observed a much lower residual similarity with Hi-C in the same model complexity (Pearson correlation r = 0.12, $p < 2.2e^{-16}$, Figure 3c). As a positive control, we applied the same GBM regression tree approach to WBC Hi-C from another individual and still observed the high similarity with the replicate (Pearson correlation r = 0.53, $p < 2.2e^{-16}$, Figure 3c). To explore the effect of model complexity on our analyses, we repeated the regression tree with a range of model complexities (n_estimator). We found that even with high model complexity the residual correlations with Hi-C are difficult to remove for multi-sample FREE-C, single-sample FREE-C, and Hi-C (rep1, another replicate), which is not true of negative control samples (i.e., multi-sample gDNA, single-sample gDNA, and FREE-C with substituted bins) (Figure 3c). This suggests that the co-fragmentation pattern observed by FREE-C is not due to sequence composition bias.

The co-fragmentation pattern observed in multi-sample FREE-C may also be introduced by the systematic variations of fragment length across different batches (for example, during the library preparation or sequencing steps). To mitigate this possibility, we generated a multi-sample FREE-C matrix from 18 samples taken from the same processing batch and observed a similar Pearson correlation with Hi-C at the pixel level ($r = 0.72$, $p < 2.2e^{-16}$, Supplemental

Figure 6b, d) when compared to that of 18 samples taken from different batches ($r = 0.75$, $p < 2.2e^{-16}$, Supplemental Figure 6c-d). We also randomly shuffled one of the paired bins across individuals in FREE-C and, as expected, no longer observed pixel level correlation with Hi-C (Pearson correlation $r = 0.001$, $p = 0.85$, Supplemental Figure 6a,d).

To evaluate the generalizability of our findings, we repeated our analyses on a publically available deep cfDNA WGS dataset that was generated using a different library preparation kit and sequencing platform (BH01)(Snyder et al. 2016). We still observed Hi-C-like patterns in this healthy cfDNA sample (Figure 2b-c, Supplemental Figure 3).

To understand the robustness of our multi-sample FREE-C approach to sample size, we randomly sub-sampled the data. With 10 samples, we still achieved an average correlation coefficient with WBC Hi-C of 0.58 at the pixel level and 0.78 at the compartment level. With additional samples, the correlation increased until it saturated with a sample size of ~60 (Supplemental Figure 7). We also varied bin sizes and consistently demonstrated high concordance with Hi-C across different resolutions (Supplemental Figure 8). Finally, to understand the effect of sequencing depth in single-sample FREE-C, we performed fragment downsampling experiments. Even with ~1X coverage, we still achieved an average correlation coefficient with WBC Hi-C of 0.48 at the pixel level and 0.72 at the compartment level (Supplemental Figure 9).

The signal of multi-sample FREE-C is mainly from chromatin organization

We next investigated the biological underpinnings of the co-fragmentation patterns we observed. If the long-range, spatially correlated fragmentation patterns we observed in cfDNA are mainly affected by the epigenetic landscape, we should find similar two-dimensional Hi-C-like patterns in related epigenetic signals. Previous studies showed that the Hi-C-like patterns

were observed at different epigenetic signals across multiple samples(Buenrostro et al. 2015; Fortin and Hansen 2015; Dileep and Gilbert 2018). To test this hypothesis at the single-sample level, we applied the two-sample KS-test approach similar to FREE-C to evaluate the similarity between paired bins using different epigenetic signals from the well-characterized cell line GM12878 (see Methods). We observed high concordances with Hi-C (GM12878) at both the pixel and compartment levels for DNA accessibility from DNase-seq, DNA methylation from whole-genome bisulfite sequencing (WGBS), H3K4me1 and H3K4me2 histone modifications from ChIP-seq (Supplemental Figure 10). These results suggested the opportunity to use "virtual chromatin organization" inferred by a variety of epigenetic marks to understand the relative contribution of different epigenetic signals on variations in FREE-C signal. We first applied GBM regression tree to regress out Hi-C signal, together with "virtual chromatin organization" inferred from the other four epigenetic signals identified above, from the signal at each pixel of the FREE-C matrix. We found that Hi-C signal is the major contributor to the signal from the multi-sample FREE-C matrix, as evidenced by relative feature importance in the regression tree across many model complexities (Figure 3d, Supplemental Figure 11b-g). To further understand the directionality of each feature's effect on the FREE-C signal, we also performed multivariate analysis and confirmed that the Hi-C signal is the major contributor to multi-sample FREE-C (Supplemental Figure 11a). For single-sample FREE-C, Hi-C and "virtual chromatin organization" inferred from H3K4me1 and WGBS were the top three contributors (Figure 3d).

FREE-C is able to quantify the tissues-of-origin in cfDNA

Identification of the tissues-of-origin for cfDNA may help detect and monitor certain disorders, such as cancer and sepsis, independent of mutation detection(Moss et al. 2018; Snyder et al. 2016). In oncology, the ability to detect subtle differences in the relative proportions of different

leukocyte subtypes may be useful for predicting or assessing therapeutic response(Lalani et al. 2018; Capone et al. 2018).

Therefore, we next asked whether we could infer the relative contributions from different tissues-of-origin to cfDNA in healthy individuals and patients with cancer by deconvolving the FREE-C signals in the cfDNA using the A/B compartments inferred from Hi-C and "virtual compartments" inferred from epigenetic markers (H3K4me1 and DNA methylation) in different reference cell types.

To answer this question, we first explored whether the FREE-C signal we observed varies between the healthy state and different types of cancer. We generated additional WGS data of cfDNA from 34 colorectal cancer, 48 lung cancer, and 19 melanoma patients (Supplemental Table 1). We applied principal component analysis (PCA) on FREE-C data from all of the healthy samples and to selected cancer samples with high tumor fractions (≥30% as estimated by ichorCNA). Strikingly, even at 500-kb resolution we observed marked separation between the healthy and cancer samples (Supplemental Figure 12a). By further applying a supervised dimensionality reduction method, Canonical Correlation Analysis (CCA), on all samples (including low tumor fraction cancer samples), we could observe a clear separation between the healthy and cancer samples (Supplemental Figure 12b-f). These results indicate that cfDNA co-fragmentation patterns differ between cancer and healthy samples, and differ among different types of cancer.

We next tested whether the eigenvector amplitude we observed in the Hi-C data is correlated with the amplitude of open or closed status seen in chromatin based on DNase-seq data (see Methods). We observed a high correlation between the signal strengths of DNase-seq and Hi-C eigenvectors at 500-kb resolution in GM12878 (Pearson correlation $r = 0.8$, $p < 2.2e^{-16}$,

Supplemental Figure 13, Supplemental Table 3), which suggested that we could use a compartment-level eigenvector to quantify the openness of chromatin (rather than just serving as a binary A/B compartment classifier). To generate a reference compartment and "virtual compartment" panel for cfDNA tissues-of-origin, we collected 65 published datasets (Hi-C, H3K4me1 and WGBS) from 16 cell/tissue/cancer types representing different normal and cancer cell types (Supplemental Table 2-3)(Zhou et al. 2018; Stunnenberg et al. 2016; Roadmap Epigenomics Consortium et al. 2015; ENCODE Project Consortium 2012). According to previous studies, neutrophils contribute most of the cfDNA in healthy individuals(Moss et al. 2018); however, no Hi-C data from human neutrophils was available. Therefore we generated our own Arima-HiC dataset from purified neutrophils, comprising ~2 billion paired-end reads and >1 billion high-quality contacts. By using quantile-normalized compartment level eigenvectors calculated from the reference panel, we inferred from FREE-C that 95% of cfDNA in healthy samples originated from different types of hematopoietic cells and almost none from cancer cells (Figure 4a-c, Supplemental Figure 14). This result is consistent with the recent finding that hematopoietic cell death is the major contributor to cfDNA (Figure 4c)(Snyder et al. 2016; Sun et al. 2015; Lui et al. 2002; Moss et al. 2018). In contrast, FREE-C analysis in samples from patients with colorectal cancer, lung cancer, and melanoma demonstrated increased contributions to cfDNA from the expected cell types (Figure 4a-b). To exclude the possibility that this result was an artifact of either our specific samples or our experimental methods, we also analyzed publically available cfDNA WGS data from healthy, colorectal cancer, and lung cancer patients(Snyder et al. 2016) and achieved similar results (Figure 4a-b). To quantify the accuracy of our approach, we compared the tumor fraction (TF) estimated by FREE-C with that estimated by ichorCNA(Adalsteinsson et al. 2017), which uses an orthogonal method based on copy number variations (CNVs) to estimate TF from cfDNA WGS data. We observed that both methods were highly concordant, whether in healthy samples (two-sided Mann–Whitney U test $p = 0.41$, Figure 4c) or in high TF cancer samples (Figure 4d). To avoid

confounding by CNVs in high TF cancer samples, we excluded the genomic regions with any significant CNV signals from the tissues-of-origin analysis and the results were essentially unchanged (Supplemental Figure 15).

DISCUSSION

In summary, we have demonstrated that cfDNA co-fragmentation patterns can be used to generate *in vivo* maps of higher-order chromatin organization. More specifically, FREE-C analysis of cfDNA fragments generated by physiological and pathological cell death may diminish the need for invasive, tissue-based analyses for mapping chromatin conformation, thus enabling non-invasive longitudinal assessment of chromatin organization across the genome.

In this study, we still observed some discrepancies at the pixel level between multi-sample FREE-C and Hi-C from WBCs, even with saturated sample sizes. This discrepancy may be due to differences in the composition and more importantly the turnover rates of hematopoietic cell types, which will skew their contributions to cfDNA.(Borghans and de Boer 2007; von Vietinghoff and Ley 2008; Summers et al. 2010). It could also be due to other factors that affect cfDNA co-fragmentation patterns , such as genetic background, cfDNA secretion by viable cells as cargo in or on extracellular vesicles(Eagen et al. 2017; Kalluri and LeBleu 2016), and multi-points 3D chromatin proximity affected by polycomb group proteins that may be challenging to detect by current Hi-C technologies(Eagen et al. 2017). The fact that cfDNA is highly heterogeneous, originating from multiple cell types undergoing different forms of cell death(Lui et al. 2002; Sun et al. 2015; Moss et al. 2018), is simultaneously a limitation and an advantage of FREE-C, which generates a comprehensive epigenetic snapshot of different cell types that are dying at a given point in time. We believe that deconvolution of the heterogeneous signals from FREE-C using appropriate computational methods combined with a comprehensive multi-omics reference cell atlas will facilitate our understanding of the cross-talk between high-order

chromatin organization, primary sequence variation, and other epigenetic factors within contributing cell types.

Furthermore, the ability to not just identify but also quantify contributions from different tissues-of-origin using epigenetic features in cfDNA may enable clinical applications in a vast array of physiological and pathological conditions beyond just those relying on the detection of genetic alterations from specific cell types of interest. Recently, several studies have developed novel methods for identifying tissues-of-origin of cfDNA, based on inferred nucleosome positioning signals from WGS(Snyder et al. 2016) and DNA methylation from WGBS(Sun et al. 2015; Lehmann-Werman et al. 2016; Guo et al. 2017; Moss et al. 2018). Library construction for bisulfite sequencing is still challenging for low DNA inputs. More importantly, during the bisulfite library preparation, the non-uniform sequence-dependent loss of the majority of DNA fragments skews the representations of cfDNA molecule, and ultimately compromises the ability to accurately capture subtle changes in the contributions of different tissues to the pool of cfDNA over time. Another limitation of previous studies(Sun et al. 2015; Guo et al. 2017; Moss et al. 2018) is the use of normal tissues as the reference panel, with or without pre-selecting a single cancer type, which limits the ability to capture the heterogeneity among different cancer types. Moreover, lack of reference nucleosome maps and reliance on an indirect correlation approach based on gene expression data(Snyder et al. 2016) do not allow for quantitative measurement of contributions to cfDNA from different tissues. In contrast, FREE-C enables quantitation of tissues-of-origin by using less biased reference maps together with a straightforward paired-end WGS library preparation without significant loss of cfDNA fragments. The accuracy of tissues-of-origin by FREE-C should improve with greater depth and breadth of high-quality reference panels, better algorithms to model heterogeneity, and better experimental approaches to study cfDNA patterns, enabling an ever broader range of physiological and pathological applications,

including (for example) the detection of early-stage, low tumor fraction cancers by non-invasive means.

Figures

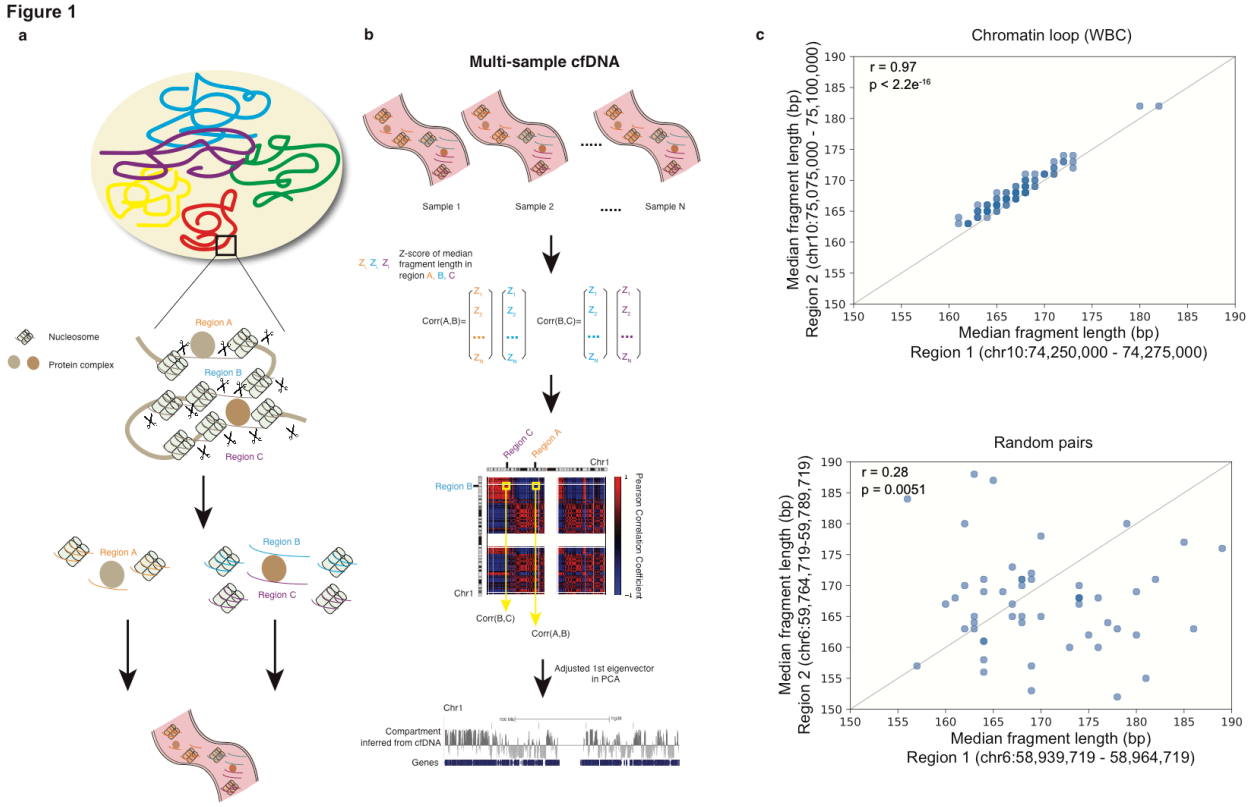


Figure 1. **Spatial co-fragmentation patterns of cfDNA in human plasma**

a. Schematic depicting how co-fragmentation is hypothesized to occur at chromatin regions in spatial proximity

b. Schematic on how co-fragmentation patterns across multiple samples are used to calculate the inferred compartment from cfDNA

c. Scatter plot of the coordinated fragment lengths across multiple samples at a spatially proximal paired region (identified by HiCCUPS loops called from Arima-HiC) in WBC (upper) and random pair of regions with the same linear distance as the loop (lower).

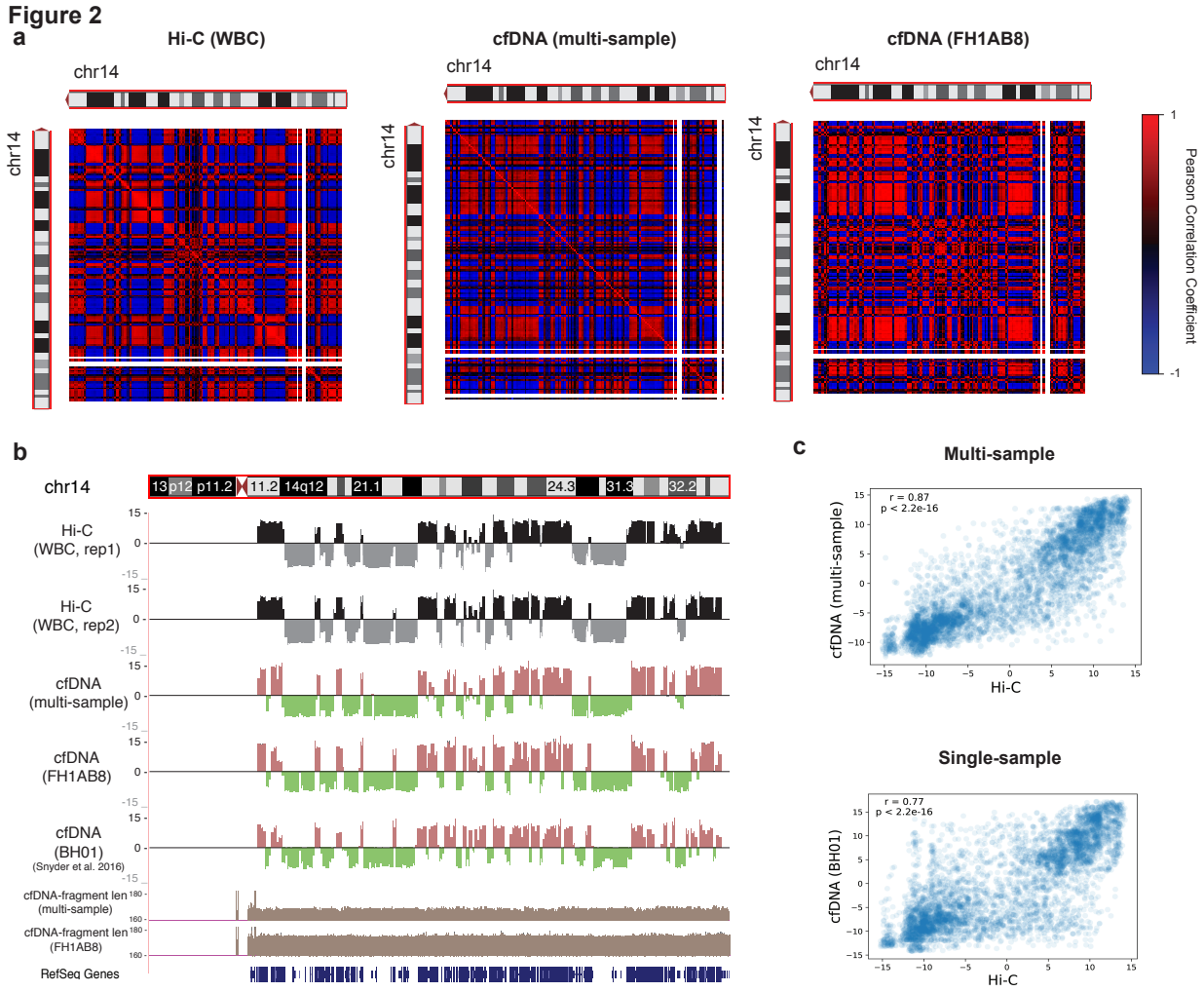


Figure 2. **Chromatin organization is recapitulated by the spatial co-fragmentation pattern in cfDNA**

a. Comparison of correlation map from Hi-C (WBC), spatially correlated fragment length from multiple cfDNA samples, and spatially correlated fragment length distribution from a single cfDNA sample. All comparisons are from chromosome 14 (chr14).

b. Genome browser tracks of compartment A/B from Hi-C (WBC), multi-sample cfDNA, single-sample cfDNA, median fragment length in each genomic bin from multi-sample cfDNA and single-sample cfDNA across chr14. Y-axis is the first eigenvector after applying PCA on the correlation matrix (Hi-C, multi-sample cfDNA and single-sample cfDNA) or median fragment length (multi-sample cfDNA and single-sample cfDNA).

c. Scatter plot of the genome-wide concordance in compartment level between Hi-C (WBC) and multi-sample cfDNA (top) or single-sample cfDNA (bottom). Values plotted on the x and y axes are the first eigenvector after applying PCA on the correlation matrix.

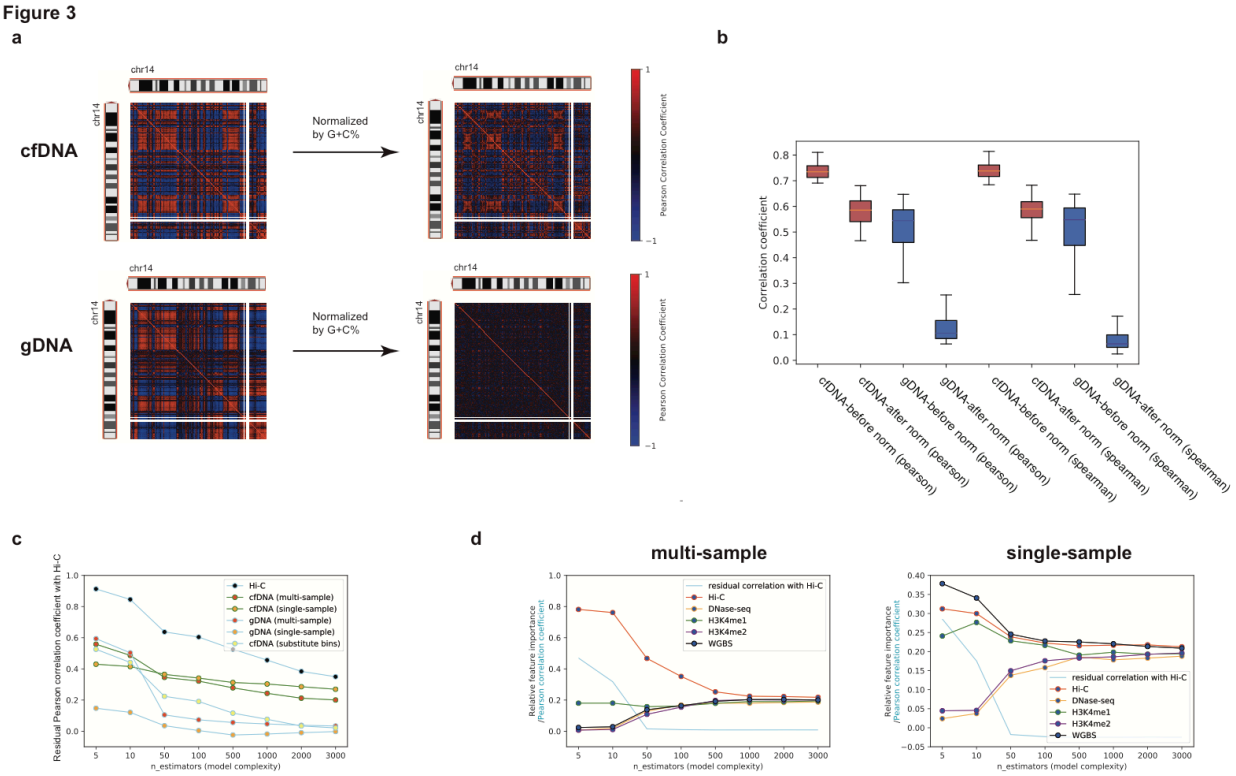


Figure 3. The spatial co-fragmentation pattern in cfDNA reflects 3D chromatin organization, not sequence composition

a. Heatmap of fragment length correlations in cfDNA (top) and gDNA (bottom), before (left) and after (right) G+C% is regressed out by Locally Weighted Scatterplot Smoothing (LOWESS) in each bin on chr14.
b. Boxplot of the Pearson and Spearman correlation coefficients with WBC Hi-C before and after G+C% is regressed out by LOWESS in each bin across autosomes.
c. Regression tree is performed to regress out G+C% and mappability at each pair of bins from the signal on chr1 from Hi-C, multi-sample FREE-C, single-sample FREE-C, multi-sample gDNA, single-sample gDNA, and randomized multi-sample FREE-C (negative control). Residual correlation with WBC Hi-C is plotted for each case at different model complexities.
d. Regression tree is performed to regress out Hi-C, DNase-seq, H3K4me1, H3K4me2, and WGBS signal together from multi-sample (left) and single-sample (right) FREE-C. Relative feature importance (calculated from GBM regression tree) and residual correlation with WBC Hi-C (blue) is plotted for each case at different model complexities.

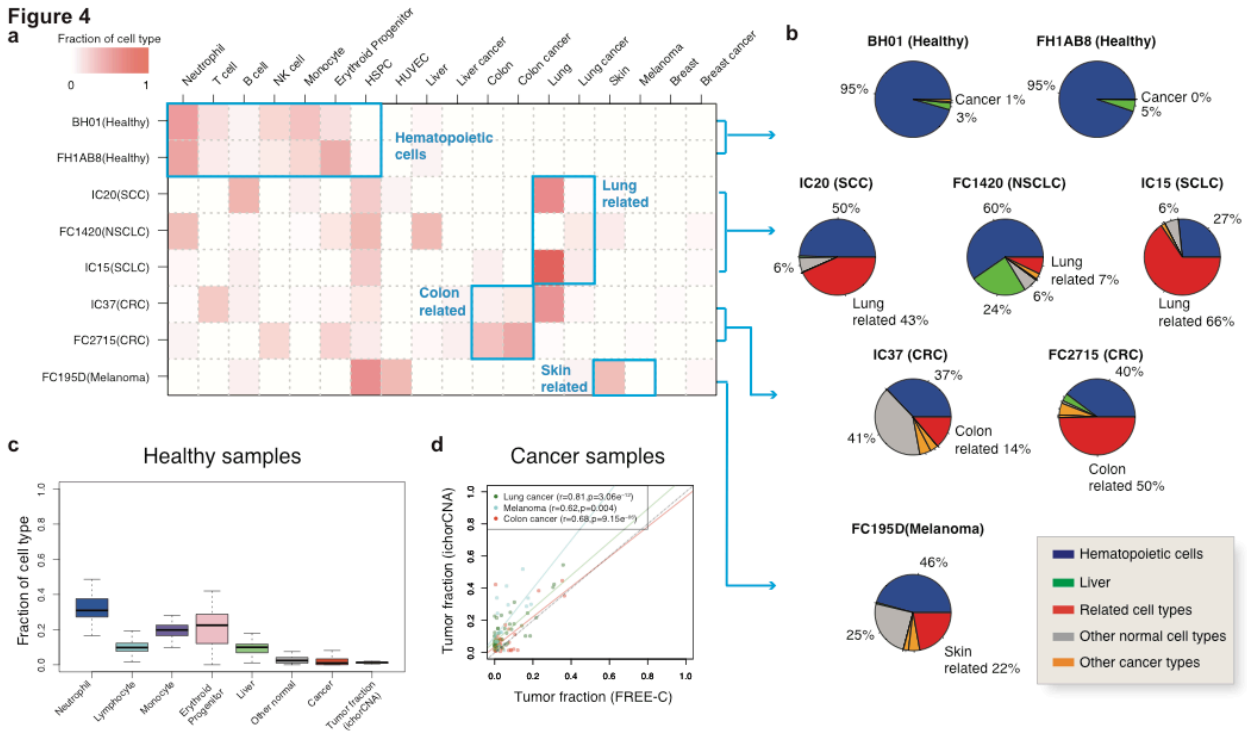


Figure 4. **cfDNA tissues-of-origin can be inferred from FREE-C**

a. Heatmap of cell type contributions to cfDNA as inferred from single-sample FREE-C of healthy, lung cancer, colorectal cancer, and melanoma samples.

b. Pie chart of cell compositions inferred from single-sample FREE-C of healthy, lung cancer, colorectal cancer, and melanoma samples.

c. Distributions of normal cell fractions and tumor fractions of cfDNA as inferred from single-sample FREE-C and ichorCNA on 100 healthy samples.

d. Scatter plot of tumor fractions estimated from FREE-C (x axis) and ichorCNA (y axis) in lung cancer (green), colorectal cancer (red), and melanoma (cyan) samples. Each colored dashed line is a line fit by linear regression between tumor fractions estimated from FREE-C and ichorCNA.

METHODS

Sample collection

Retrospective human plasma samples (>0.25 mL) were acquired from commercial biobanks for 34 patients diagnosed with colorectal cancer, 49 diagnosed with lung cancer, and 19 diagnosed with melanoma. Healthy controls comprised 100 samples from patients without a current cancer diagnosis. All plasma samples were de-identified prior to receipt, with no key available to re-identify. Plasma was stored at -80°C and thawed prior to use.

Purified human neutrophils (10 million cells) were purchased from Astarte Biologics Inc.

Whole-genome sequencing of cfDNA

Cell-free DNA was extracted from 250 µl plasma using the MagMAX Cell-Free DNA Isolation Kit (Applied Biosystems), per manufacturer instructions. Paired-end sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit (New England Biolabs) and sequenced on the Illumina NovaSeq 6000 Sequencing System with unique dual indexes across multiple flowcells at 2x51 base pairs.

WGS data processing

Reads were de-multiplexed and aligned to the human genome (GRCh38 with decoys, alt contigs, and HLA contigs) using BWA-MEM 0.7.15(Li and Durbin 2010). PCR-duplicate fragments were removed based on unique molecular identifiers (comprising the first five nucleotides of each read). Only high-quality autosomal reads (both ends uniquely mapped, end 1 with mapping quality score of 60, properly paired, and not a PCR duplicate) were used for all of the downstream analysis. Details of sequencing coverage are provided in Supplemental Table 1.

Arima-HiC library preparation and sequencing

Hi-C library preparation on WBCs and neutrophils was performed using the Arima-HiC kit (Arima Genomics, Inc.; Cat # A510008), according to the manufacturer's protocols. Briefly, WBCs (freshly isolated from whole blood) or purified neutrophils (obtained frozen from a commercial source and then thawed) were crosslinked using formaldehyde. Crosslinked cells were then subject to the Arima-HiC protocol, which utilizes multiple restriction enzymes to digest chromatin. Arima-HiC sequencing libraries were prepared by first shearing purified proximally-ligated DNA and then size-selecting 200-600 bp DNA fragments using SPRI beads. The size-selected fragments were then enriched using Enrichment Beads (provided in Arima-HiC kit), and then converted into Illumina-compatible sequencing libraries with TruSeq adapters using the KAPA Hyper Prep kit. The purified, PCR-amplified DNA underwent standard QC (qPCR and Bioanalyzer) and was sequenced with unique dual indexes on the Illumina HiSeq X (WBC) or NovaSeq 6000 (neutrophils) Sequencing System at 2x151 base pairs. The sequencing summary statistics and related metadata information are shown in Supplemental Table 2.

Hi-C data processing

Raw fastq files from various Hi-C experiments were uniformly processed using Juicer command line tools v1.5.6(Durand et al. 2016). Only results after filtering reads with mapping quality score < 30 were used to generate a Pearson correlation matrix and compartment A/B eigenvector. PCA was performed using the PCA function in scikit-learn 0.19.1 in Python 3.6. The first principal component was used to segment the compartments. For each chromosome, compartments were split into two groups based on their signs. After excluding regions with mappability less than 0.75, the group of compartments with the lower mean gene density (based on Ensembl v84 annotations) was defined as compartment B; the other group was defined as compartment A. HiCCUPS loops were called by hiccups command line in Juicer with

parameters "-k KR -m 512 -r 5000,10000,25000 -f .1,.1,.1 -p 4,2,1 -i 7,5,3 -t 0.02,1.5,1.75,2 -d 20000,20000,50000".

Multi-sample FREE-C

Each 500-kb bin (excluding those bins with mappability less than 0.75) was first divided into 50-kb sub-bins. The median of fragment length in each sub-bin was summed up in 500-kb bins first and then converted to a z-score using the mean and standard deviation in each chromosome and each sample. Pearson correlation matrix was calculated between each paired bin across all individuals. The "pixel level" similarity with WBC Hi-C is calculated by vectorizing the correlation matrices of both FREE-C and Hi-C and then computing the Pearson correlation, across all autosomes. The "compartment" level similarity is calculated by the correlation of eigenvector between FREE-C and Hi-C level across all genomic bins. The implementation details are described within the iPython notebook.

Single-sample FREE-C

The distribution of lengths of all high-quality fragments within each 500-kb bin (excluding those with mappability less than 0.75) was first obtained. The similarity of fragment-length distribution within each pair of 500-kb bins was then calculated using a two-sample K-S test (ks_2samp function implemented in scipy 1.1.0 in Python 3.6). The Pearson correlation for a particular paired bin was then calculated between log10 p-values in all the bins at that row and log10 p-values in all the bins at that column(Durand et al. 2016; Lieberman-Aiden et al. 2009). The details are implemented in "bamTo1DKsMatAndEigen.py" with default parameters.

Sequence composition and mappability bias analysis

Mappability scores were generated by GEM v2.5.2.1(Derrien et al. 2012) for read length 45bp. G+C% was taken from the gc5base track in UCSC Genome Browser. For each pair (bin1 and

bin2) of 500-kb bins, we obtained the G+C% and mappability scores for bin1 and bin2. We then applied a GBM regression tree (GradientBoostingRegressor function implemented in scikit-learn 0.19.1 in Python 3.6) to regress out G+C% and mappability on each pixel of correlation coefficient score from the matrix in FREE-C, gDNA, and Hi-C data. N_estimators were varied at different model complexities with depth=5. Residual values after the regression were then used to calculate the correlation with WBC Hi-C data at the pixel level. r^2 was calculated to measure the goodness-of-fit of the model. The implementation details are documented within the iPython notebook.

"Virtual chromatin organization and compartment A/B" inferred from DNase-seq, WGBS, and ChIP-seq and their contributions to the signal in FREE-C

A similar approach was performed as that in single-sample FREE-C. Instead of fragment length, signal strength was used (mean DNA methylation level for WGBS or mean -log10 p-value calculated from MACS2 for DNase-seq/ChIP-seq in each 500-kb bin). The implementation details are described within the iPython notebook.

For all of the histone modification ChIP-seq, DNA accessibility and DNA methylation signals available in encode (hg38) for GM12878, we generated the inferred "virtual chromatin organization" (Pearson correlation matrix for each type of signal). We then vectorized the correlation matrices for each type of signal, Hi-C and FREE-C matrix. Then we applied GBM regression tree to regress out the combination of all signals (or separately for each of the single signal) from the FREE-C matrix and then measured the relative feature importance of each feature (provided by GradientBoostingRegressor function implemented in scikit-learn 0.19.1). An ordinary least squares model was applied for the multivariate analysis to measure the directionality of contribution from each feature to the FREE-C matrix (statsmodel package 0.9.0 in Python 3.6). The implementation details are described within the iPython notebook.

Tissues-of-origin in FREE-C

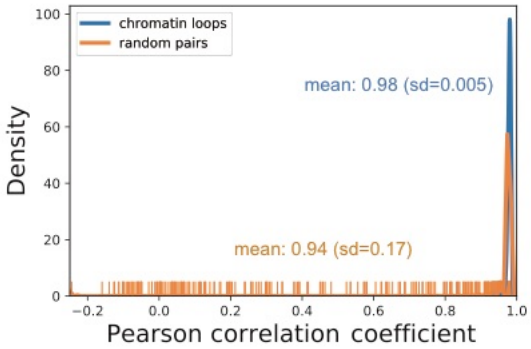
To infer tissues-of-origin from FREE-C data, we modeled the compartment of FREE-C data (first principal component on correlation matrix in FREE-C) as a linear combination of the compartments in each of the reference data (first principal component on correlation matrix in Hi-C, H3K4me1, or WGBS). The eigenvector was re-evaluated to ensure compartment A always has a positive sign. Genomic bins with mappability less than 0.75 were filtered. Eigenvectors across FREE-C and reference panels were first transformed by quantile normalization (preprocessCore package in R). We constrained the weights of the coefficients to sum to one so that the weights can be interpreted as fractional tissue contributions to cfDNA. Quadratic programming was utilized to solve the constrained optimization problem (quadprog package in R). For each tissue/cell type, we summed the fractions from all assay that belong to the tissue/cell type. To define tumor fraction, we summed the fractional tissue contributions from cancer cells. The detailed fractional tissue contributions are available for each sample in Supplementary Table 4. The implementation details are described within the R code of README file (tissue_of_origin function and tissue_of_origin_filter_cna function).

ichorCNA analysis

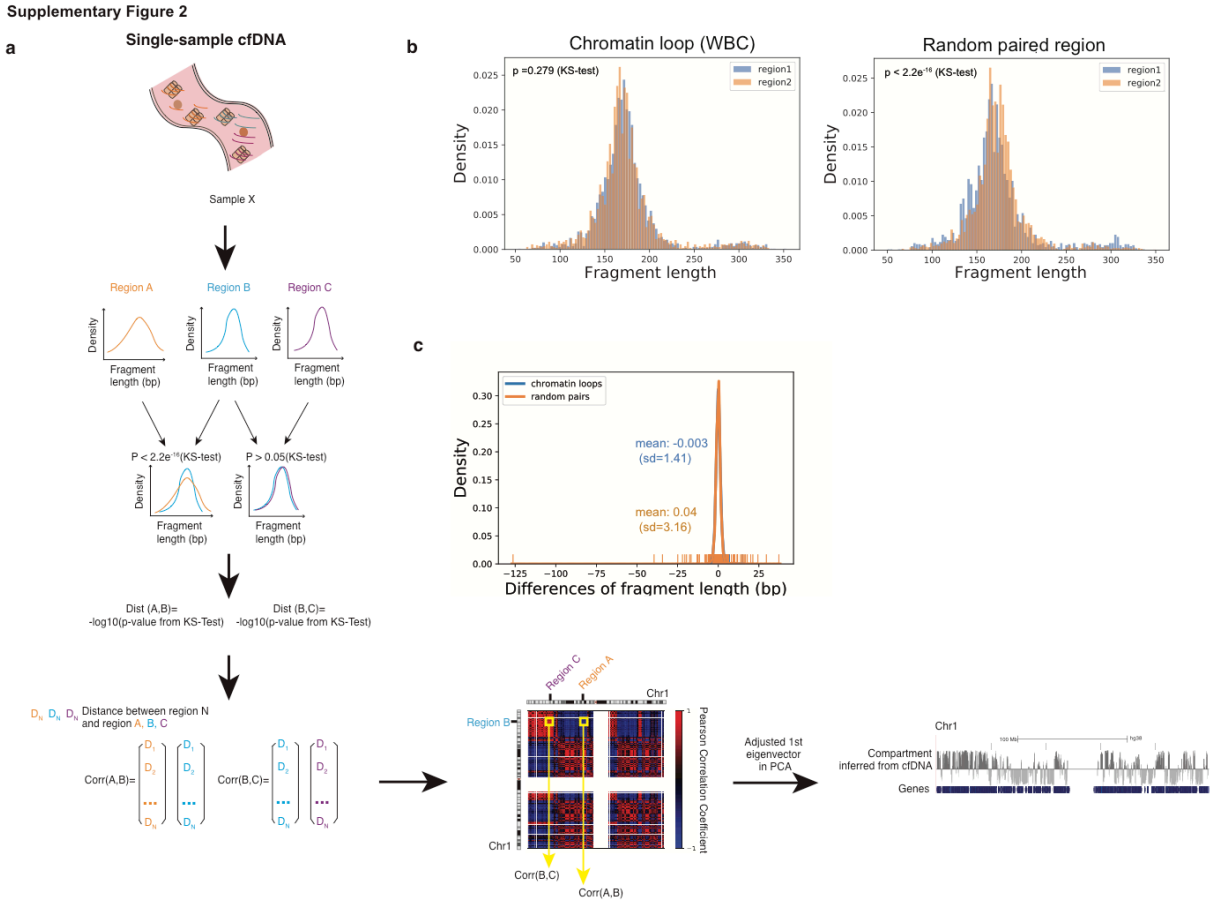
ichorCNA v0.1.0(Adalsteinsson et al. 2017) was run with the following parameters: --normal "c(0.95, 0.99, 0.995, 0.999)" --ploidy "c(2)" --maxCN 3 --estimateScPrevalence FALSE --scStates "c()" --chrs "c(1:22)" --chrTrain "c(1:22)". It is used to estimate the tumor fraction in each cfDNA sample after normalizing to the group of our internal healthy samples. The detailed tumor fractions are available for each sample in Supplementary Table 4. The detailed ichorCNA segmentation results for each sample are available at GEO.

Supplementary materials

Supplementary Figure 1



Supplementary Figure 1. **Spatial co-fragmentation patterns of cfDNA in multiple human plasma samples.** Summary density plot of Pearson correlation coefficient between median fragment length in paired regions identified by HiCCUPS loops called from Arima-HiC in WBC (blue color) and between random pairs of regions with the same distance as the loop (orange color). Additional rug plot on x-axis is plotted to show the long tails in random pairs.



Supplementary Figure 2. **Spatial co-fragmentation patterns of cfDNA in a single human plasma sample**

a. Schematic on how co-fragmentation patterns in single samples is used to infer chromatin organization.

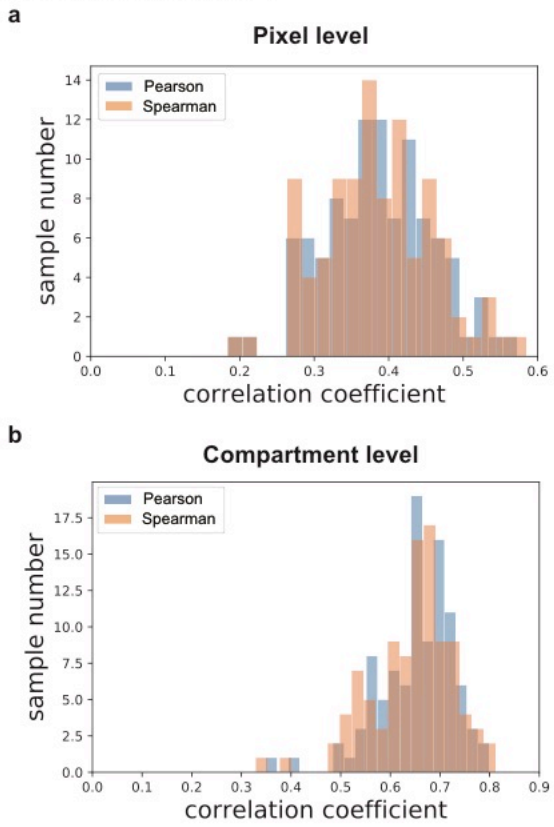
b. Example scatter plot of the coordinated fragment length in a single sample at a spatially proximal paired region identified by HiCCUPS loops called from Arima-HiC in WBC (left) and a random pair of regions with the same distance as the loop (right).

c. Summary histogram of fragment length differences between paired regions identified by HiCCUPS loops called from Arima-HiC in WBC (blue color) and a random pair of regions with the same distance as the loop (orange color).

Supplementary Figure 3

Supplementary Figure 3: 500-kb resolution correlation map from Hi-C (WBC), multi-sample FREE-C, single-sample FREE-C, and external single-sample FREE-C (BH01)[10]. Pixel level correlation coefficient shown in Hi-C is calculated by replicates from two different healthy individuals. Pixel level correlation coefficient shown in cfDNA (multi-sample and single-sample) is calculated by correlation with WBC individual 2.

Supplementary Figure 4

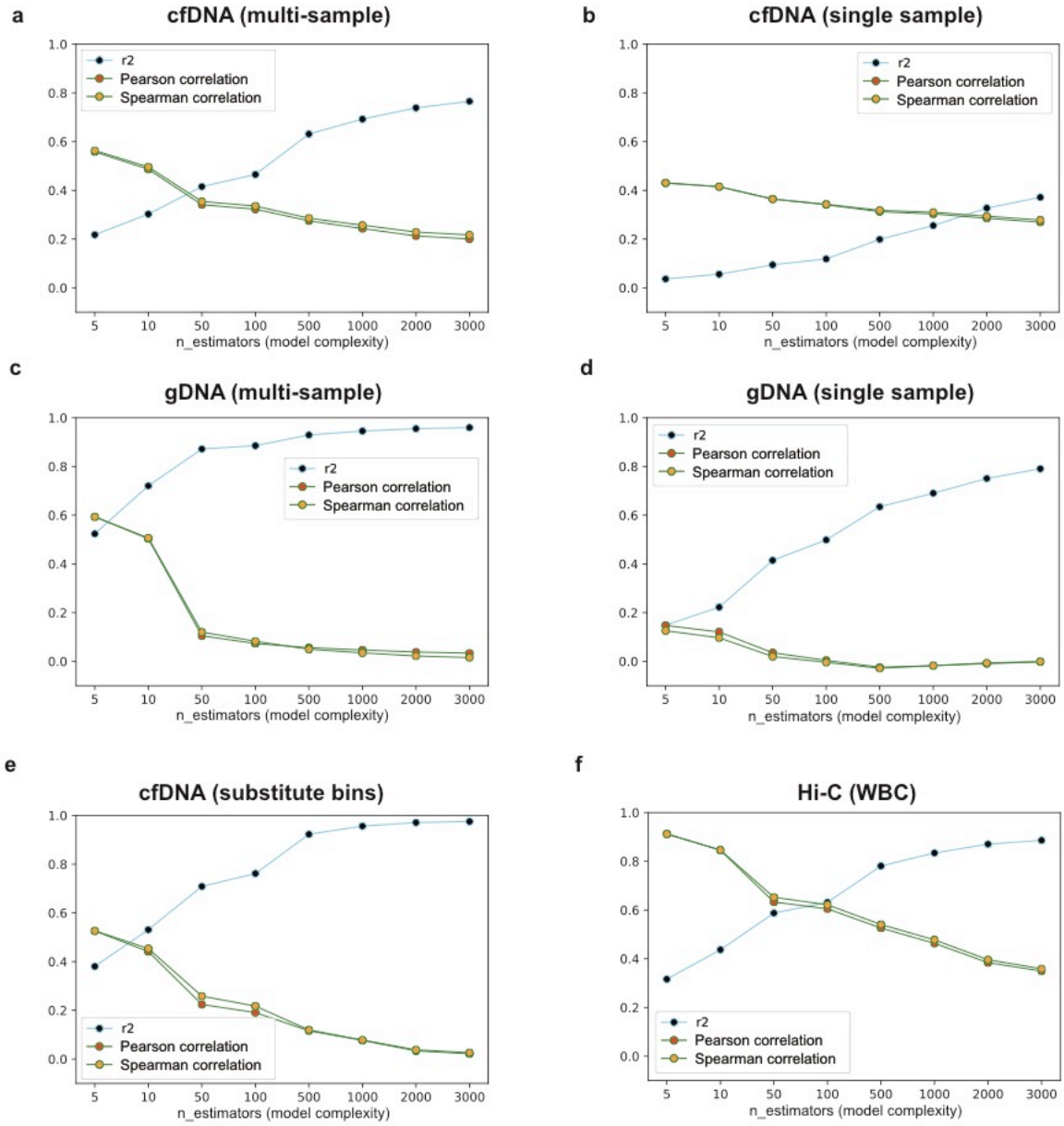


Supplementary Figure 4: The correlation between Hi-C (WBC) and FREE-C at the single sample.

a. The correlation between Hi-C (WBC) and FREE-C at the pixel level (500-kb bin).

b. The correlation between Hi-C (WBC) and FREE-C at the compartment level (500-kb bin).

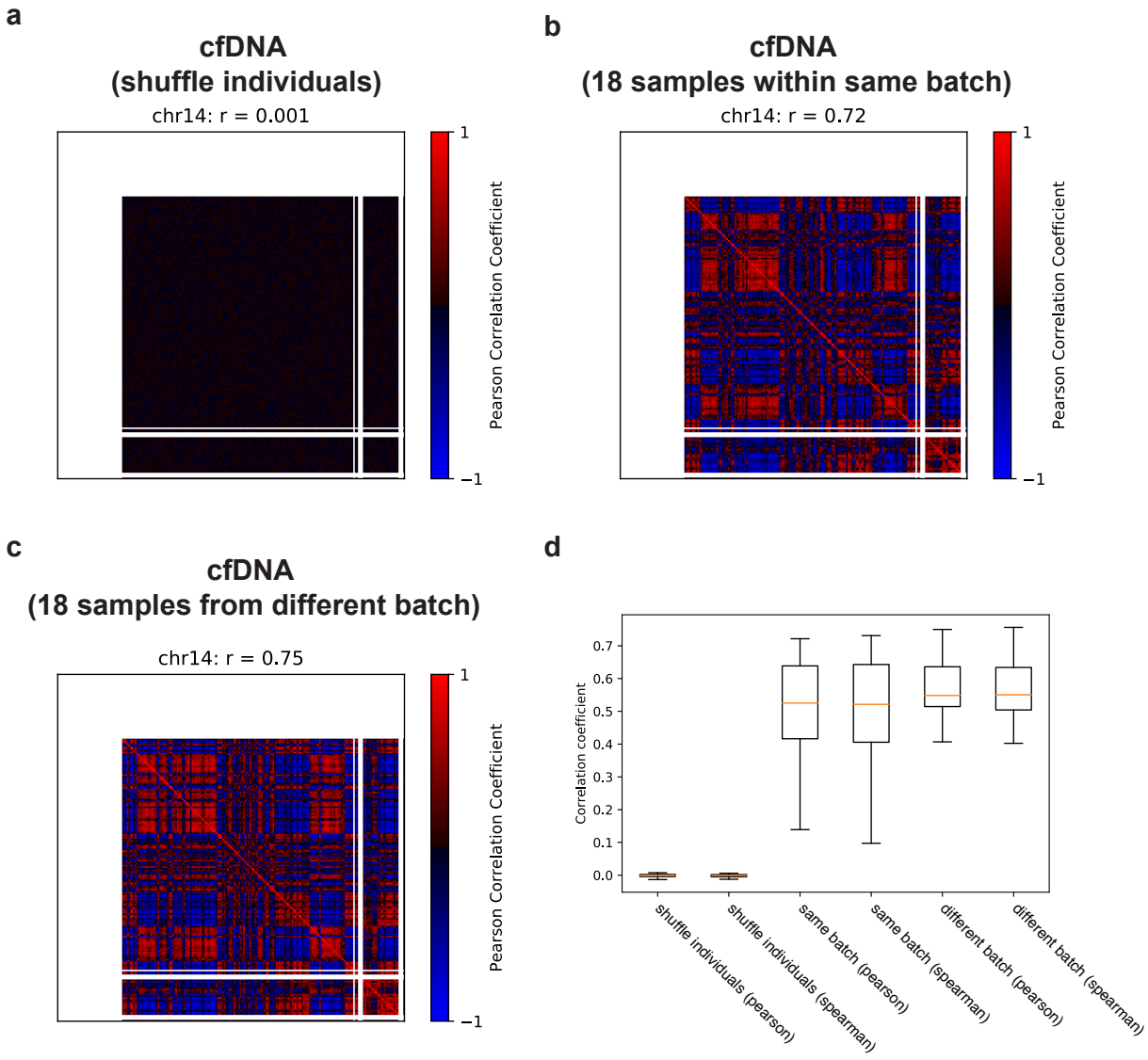
Supplementary Figure 5



Supplementary Figure 5: G+C% and mappability bias analysis. Regression tree was performed to regress out G+C% and mappability at each pair of bins from the signal on chr1 from (a) multi-sample FREE-C; (b) single-sample FREE-C; (c) multi-sample gDNA; (d) single-sample gDNA; (e) multi-sample FREE-C in which one of the paired bins is substituted by a random bin from any other chromosome with the same G+C% and mappability; and (f) Hi-C

(WBC). Residual correlation with WBC Hi-C (green line) and r^2 of the model goodness-of-fit (blue line) is plotted for each case at different model complexities.

Supplementary Figure 6



Supplementary Figure 6: Batch effect analysis.

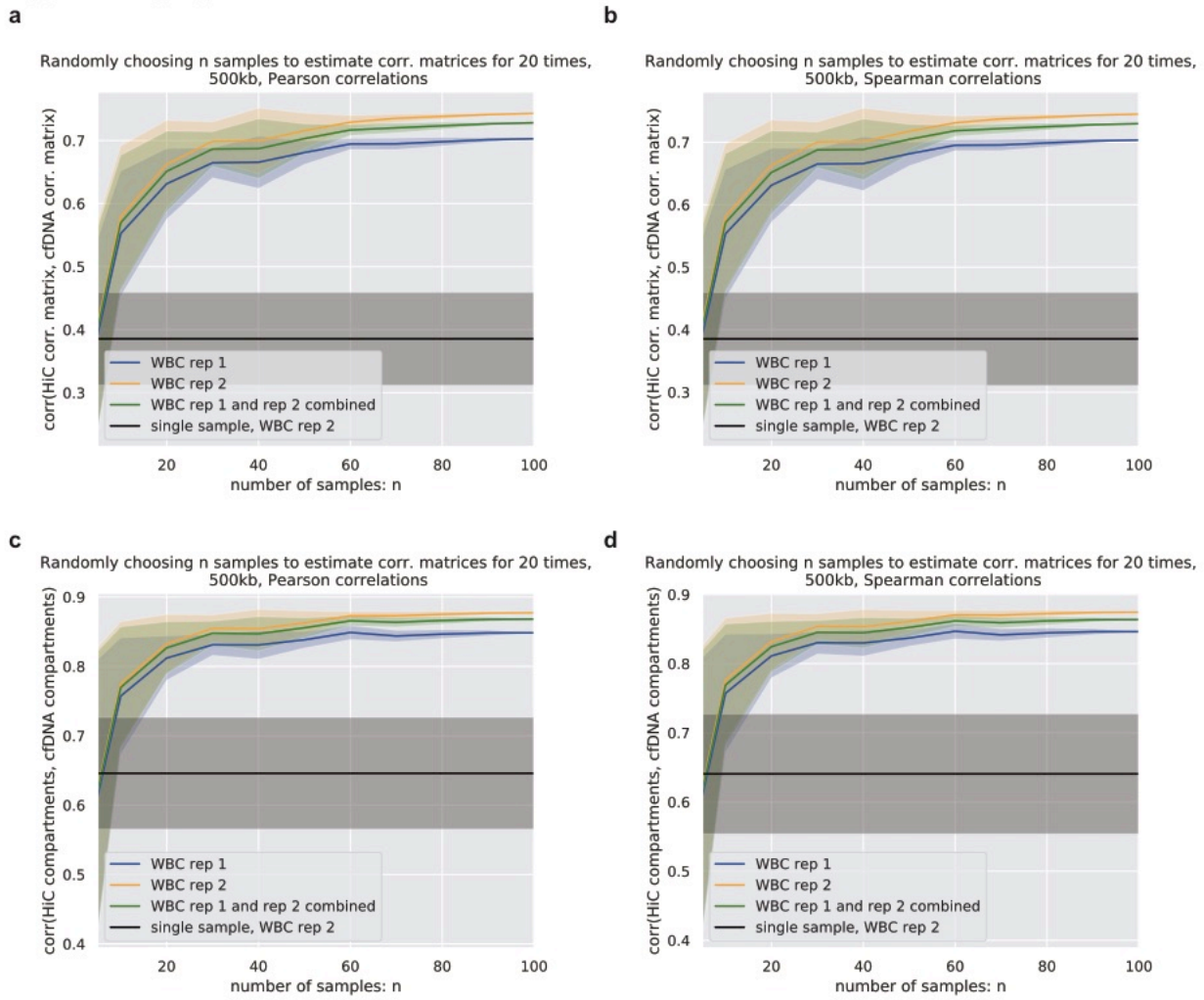
a. Heatmap of multi-sample FREE-C when one of the paired bins is randomly shuffled from any other individual (chr14).

b. Heatmap of multi-sample FREE-C (chr14) on samples (n = 18) from the same batch.

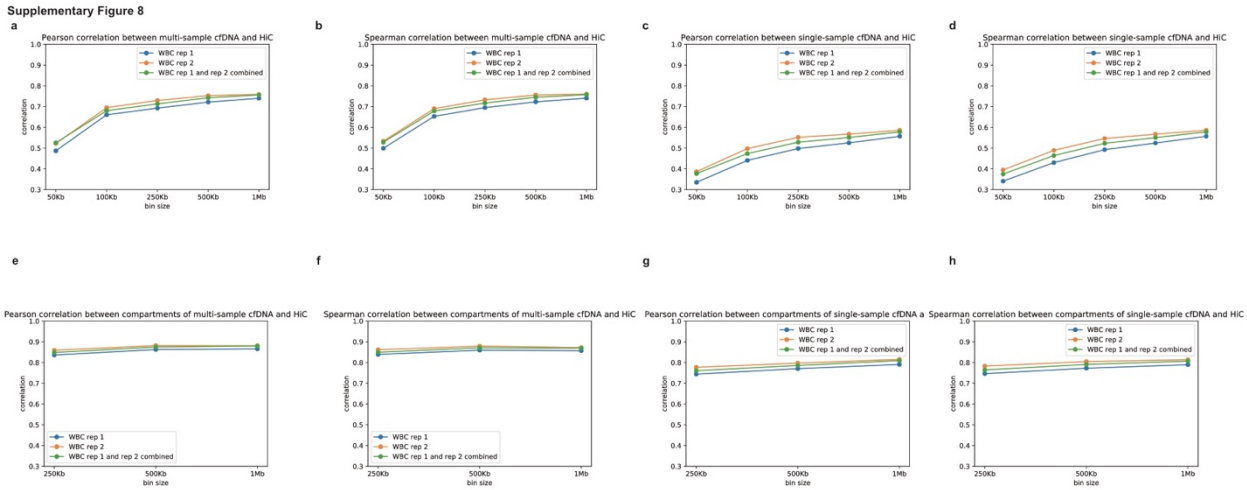
c. Heatmap of multi-sample FREE-C (chr14) on samples (n = 18) from different batches.

d. Boxplot of pixel level correlation with Hi-C (WBC, rep2) across all of the autosomes represented in a-c.

Supplementary Figure 7



Supplementary Figure 7: Correlation between Hi-C and FREE-C at different sample sizes. Pearson (a) and Spearman (b) correlation between Hi-C (WBC) and multi-sample FREE-C at pixel level in different sample sizes. Pearson (c) and Spearman (d) correlation between Hi-C (WBC) and multi-sample FREE-C at compartment level in different sample sizes. Shaded regions represent the standard deviation of the correlation coefficient. Correlation coefficient with Hi-C (WBC, rep2) in single-sample FREE-C is also plotted as a comparison.



Supplementary Figure 8: Correlation between Hi-C and FREE-C at different bin sizes.

a. Pearson correlation at pixel level between Hi-C (WBC) and multi-sample FREE-C at different bin sizes.

b. Spearman correlation at pixel level between Hi-C (WBC) and multi-sample FREE-C at different bin sizes.

c. Pearson correlation at pixel level between Hi-C (WBC) and single sample FREE-C at different bin sizes.

d. Spearman correlation at pixel level between Hi-C (WBC) and single sample FREE-C at different bin sizes.

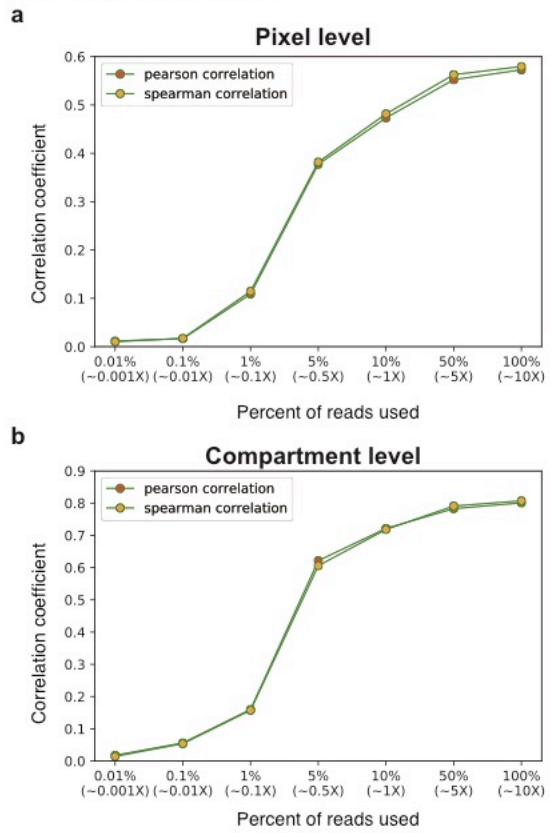
e. Pearson correlation at compartment level between Hi-C (WBC) and multi-sample FREE-C at different bin sizes.

f. Spearman correlation at compartment level between Hi-C (WBC) and multi-sample FREE-C at different bin sizes.

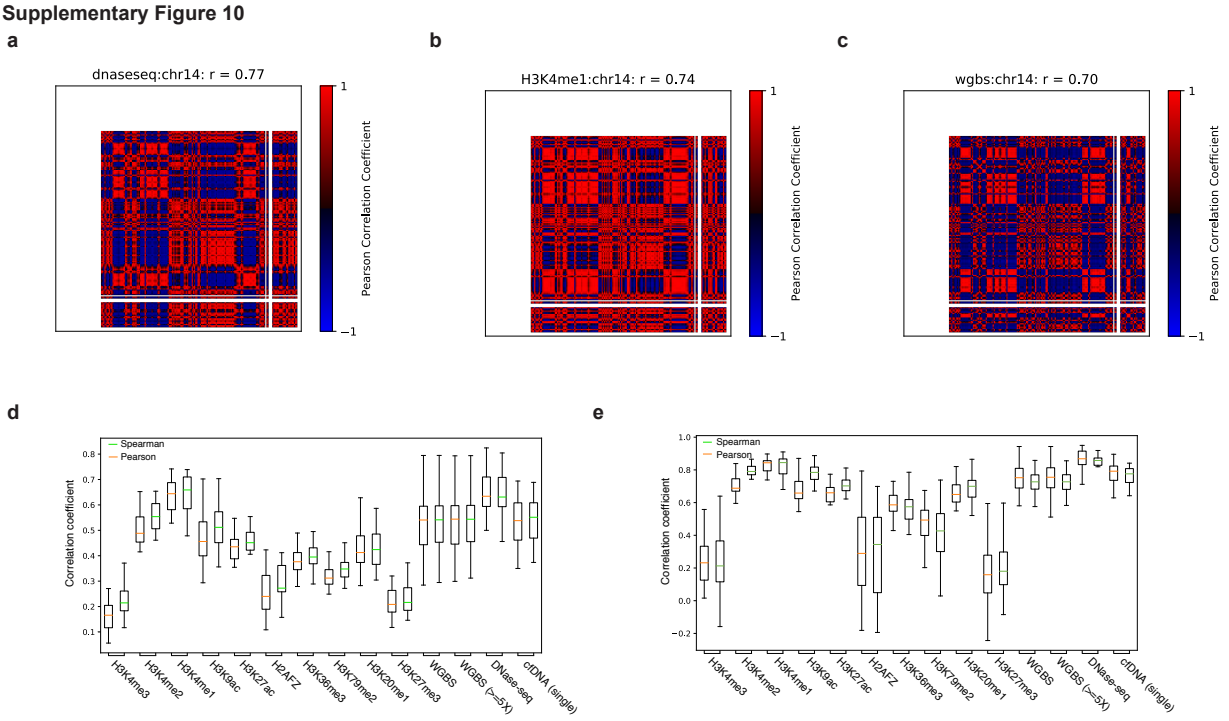
g. Pearson correlation at compartment level between Hi-C (WBC) and single sample FREE-C at different bin sizes.

h. Spearman correlation at compartment level between Hi-C (WBC) and single sample FREE-C at different bin sizes.

Supplementary Figure 9



Supplementary Figure 9: Correlation between Hi-C and single-sample FREE-C at different coverages. Pearson and Spearman correlation at pixel level (a) and compartment level (b) between Hi-C (WBC) and single-sample FREE-C (FH1AB8) at different read number after downsampling.



Supplementary Figure 10. Correlation between two-dimensional epigenetic signals similarity and Hi-C.

a. Heatmap of DNase accessibility correlation in two dimensions (chr14).

b. Heatmap of H3K4me1 correlation in two dimensions (chr14).

c. Heatmap of DNA methylation from WGBS, two dimensional correlation (chr14).

d. Boxplot of pixel level correlation between different epigenetic signals (two dimensions) and Hi-C from the same cell type (GM12878) across all autosomes.

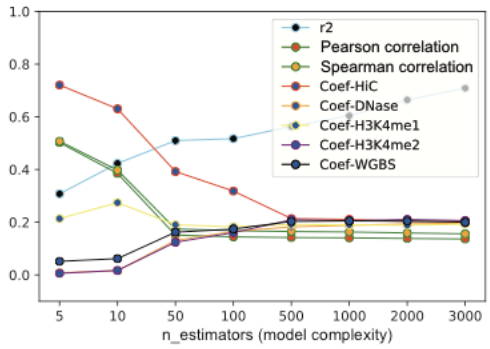
e. Boxplot of compartment level correlation between different epigenetic signals (two dimensions) and Hi-C from the same cell type (GM12878) across all autosomes.

Supplementary Figure 11

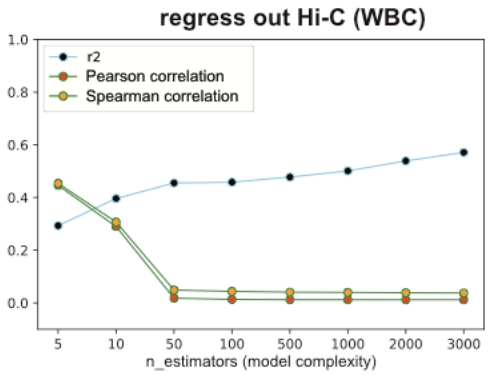
a

	coef	std err	t	P>|t|	[0.025	0.975]
const	-0.2058	0.002	-109.422	0.000	-0.210	-0.202
Hi-C	0.3848	0.002	218.020	0.000	0.381	0.388
DNase-seq	0.0386	0.002	20.025	0.000	0.035	0.042
H3K4me1	0.2039	0.003	72.825	0.000	0.198	0.209
H3K4me2	0.0419	0.003	12.238	0.000	0.035	0.049
WGBS	0.0953	0.002	62.049	0.000	0.092	0.098

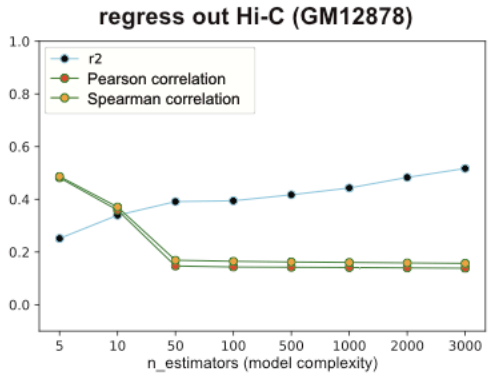
b



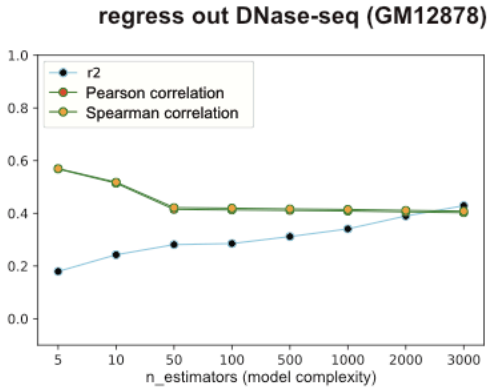
c



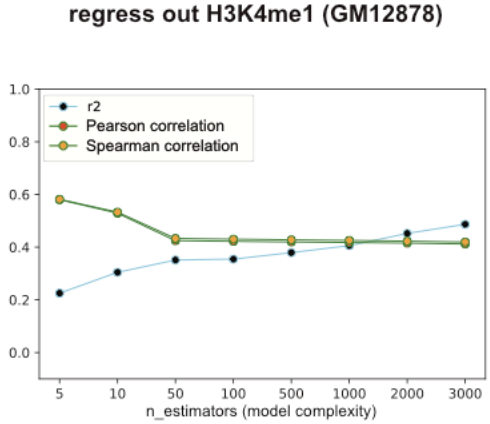
d



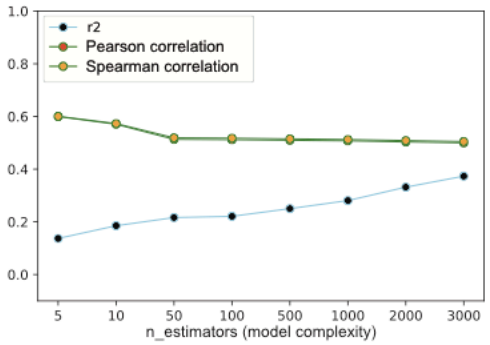
e



f



g



Supplementary Figure 11: The contributions of different epigenetic signals to the signal of FREE-C.

a. Multivariate regression analysis to explain the signal of multi-sample FREE-C.

b. Regression tree is performed to regress out Hi-C (GM12878), DNase-seq (GM12878), H3K4me1 (GM12878), H3K4me2 (GM12878) and WGBS (GM12878) signals together from multi-sample FREE-C. The importance of each feature is evaluated at different model complexities.

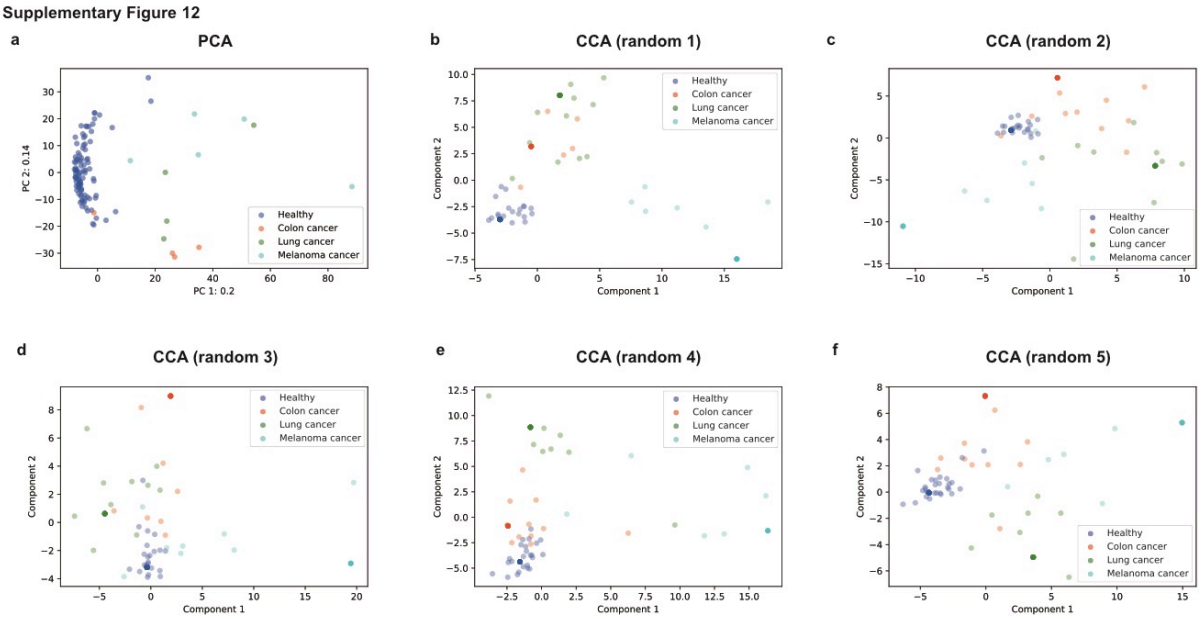
c. Regression tree is performed to regress out Hi-C (WBC) from multi-sample FREE-C. Residual correlation with WBC Hi-C (green line) and r^2 of the model goodness-of-fit (blue line) is plotted for each case at different model complexities.

d. Regression tree is performed to regress out Hi-C (GM12878) from multi-sample FREE-C.

e. Regression tree is performed to regress out DNase-seq (GM12878) from multi-sample FREE-C.

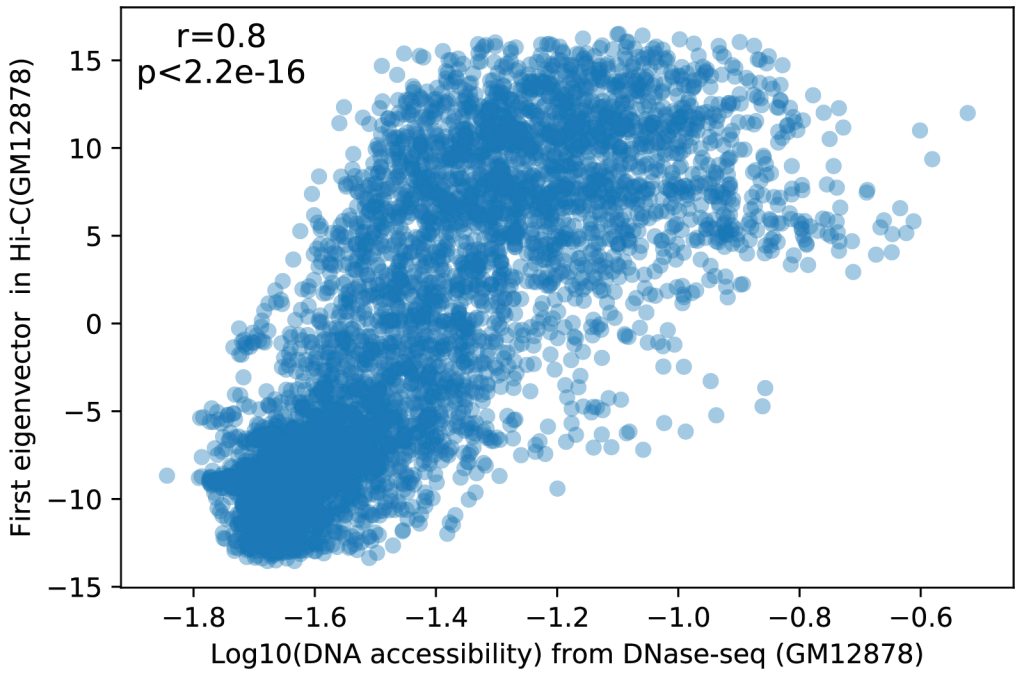
f. Regression tree is performed to regress out H3K4me1 (GM12878) from multi-sample FREE-C.

g. Regression tree is performed to regress out WGBS (GM12878) from multi-sample FREE-C.

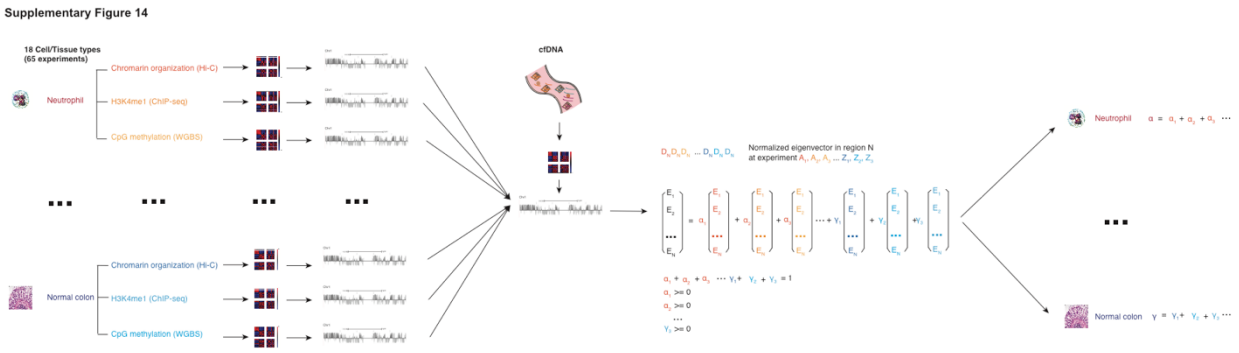


Supplementary Figure 12: Reduced dimensional representation of single-sample FREE-C at compartment level in different physiological and pathological conditions. (a) PCA on healthy and high tumor fraction (TF>0.3) colorectal cancer, lung cancer, and melanoma samples. Contribution of each PC is also plotted. (b-f) CCA on test samples (one-third of all samples) in healthy, colorectal cancer, lung cancer, and melanoma patients. The test was randomly performed five times (labeled as “random 1” through “random 5”).

Supplementary Figure 13

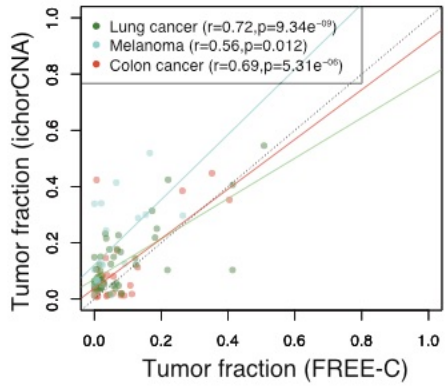


Supplementary Figure 13: Correlation between DNA accessibility from DNase-seq and compartment-level eigenvector from Hi-C in the same cell line (GM12878).



Supplementary Figure 14: Workflow scheme for single-sample FREE-C tissues-of-origin analyses. "Virtual compartment" is calculated from "virtual chromosome organization" in each assay from each cell type as the reference panel. The first eigenvector from FREE-C is deconvoluted by the reference panel as described in Online Methods. The coefficients from each of cell type are added together to represent the fraction of cell type that contributed to the cfDNA.

Supplementary Figure 15



Supplementary Figure 15: Correlation between tumor fraction estimated from ichorCNA and from single-sample FREE-C, using only genomic bins with no overlap of CNVs detected by ichorCNA. Data are plotted as in Figure 4d.

Supplementary Table 1. Summary of internal cfDNA WGS data.

Supplementary Table 2. Summary of internal generated Hi-C data.

Supplementary Table 3. Summary of all publically available data used in this study.

Supplementary Table 4. Summary of the tissues-of-origin results in each sample.

ACKNOWLEDGMENTS

The authors would like to thank the following individuals for thoughtful discussions and suggestions: tissues-of-origin analysis from Drs. Anshul Kundaje and Bing Ren; sequence composition bias analysis from Artur Jaroszewicz; the effect of epigenetic factors analysis from Drs. Olivier Elemento and Loren Hansen; ichorCNA analysis from Nathan Wan and Dr. Gavin Ha; manuscript preparation from Signe Fransen, Dr. Greg Hogan and Riley Ennis; and the design of the schematics from Reed Perkins, Drs. Ellen Heitzer, Peter Ulz and Michael Speicher. The authors gratefully acknowledge National Health Services Research Scotland, Tayside Biorepository, Geneticist Inc., iSpecimen Inc., and Indivumed for support of this research by providing de-identified plasma samples.

AUTHOR CONTRIBUTIONS

YL conceived the idea and wrote the manuscript with input from TYL, DEW, CLT, ADS, SS, LCL, LC, FCB, GP and ISH. YL and ISH designed the experiments with input from DEW. YL and TYL performed the computational analyses with review and infrastructure support from BWW. CJD, CLT, ADS, SS, VT, LCL designed and performed the Hi-C experiment on neutrophils and WBCs. LC and FCB provided the detailed clinical annotations of cancer samples. GP and ISH supervised the project.

DISCLOSURE DECLARATION

A patent application has been filed for aspects of the methods disclosed here (YL, ISH, and TYL: “Method to infer chromosome conformation changes by the abundance of cell-free DNA”). YL, TYL, DEW and BWW are employees of Freenome Inc. GP is an employee of Freenome and an independent contractor/consultant for the MolDX program at Palmetto GBA. ISH is a former employee and owns stock from Freenome Inc. CJD, CLT, ADS, SS are employees of Arima Genomics.

Human research protections statement

WBC samples for Hi-C were obtained under an IRB protocol approved by the UCSD Human Research Protections Program. All plasma samples analyzed in the cfDNA study were de-identified prior to receipt, and no key is available to Freenome to re-identify.

REFERENCES

Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, Gydush G, Reed SC, Rotem D, Rhoades J, et al. 2017. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun* **8**: 1324.

Bickmore WA. 2013. The spatial organization of the human genome. *Annu Rev Genomics Hum Genet* **14**: 67–84.

Borghans JAM, de Boer RJ. 2007. Quantification of T-cell dynamics: from telomeres to DNA labeling. *Immunol Rev* **216**: 35–47.

Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, Chang HY, Greenleaf WJ. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**: 486–490.

Capone M, Giannarelli D, Mallardo D, Madonna G, Festino L, Grimaldi AM, Vanella V, Simeone E, Paone M, Palmieri G, et al. 2018. Baseline neutrophil-to-lymphocyte ratio (NLR) and derived NLR could predict overall survival in patients with advanced melanoma treated with nivolumab. *J Immunother Cancer* **6**: 74.

Chan KCA, Zhang J, Hui ABY, Wong N, Lau TK, Leung TN, Lo K-W, Huang DWS, Lo YMD. 2004. Size distributions of maternal and fetal DNA in maternal plasma. *Clin Chem* **50**: 88–92.

Christina Fan H, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. 2008. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci U S A* **105**: 16266–16271.

Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* **2**: 292–301.

Derrien T, Estellé J, Sola SM, Knowles DG, Raineri E, Guigó R, Ribeca P. 2012. Fast Computation and Applications of Genome Mappability. *PLoS One* **7**: e30377.

Dileep V, Gilbert DM. 2018. Single-cell replication profiling to measure stochastic variation in mammalian replication timing. *Nat Commun* **9**: 427.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376–380.

Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, Aiden EL. 2016. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**: 95–98.

Eagen KP, Aiden EL, Kornberg RD. 2017. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proc Natl Acad Sci U S A* **114**: 8764–8769.

ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Fortin J-P, Hansen KD. 2015. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol* **16**: 180.

Guo S, Diep D, Plongthongkum N, Fung H-L, Zhang K, Zhang K. 2017. Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat Genet* **49**: 635–642.

Heitzer E, Haque IS, Roberts CES, Speicher MR. 2018. Current and future perspectives of liquid biopsies in genomics-driven oncology. *Nat Rev Genet*. http://dx.doi.org/10.1038/s41576-018-0071-5.

Ivanov M, Baranova A, Butler T, Spellman P, Mileyko V. 2015. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation. *BMC Genomics* **16 Suppl 13**: S1.

Kalluri R, LeBleu VS. 2016. Discovery of Double-Stranded Genomic DNA in Circulating Exosomes. *Cold Spring Harb Symp Quant Biol* **81**: 275–280.

Lalani A-KA, Xie W, Martini DJ, Steinharter JA, Norton CK, Krajewski KM, Duquette A, Bossé D, Bellmunt J, Van Allen EM, et al. 2018. Change in Neutrophil-to-lymphocyte ratio (NLR) in response to immune checkpoint blockade for metastatic renal cell carcinoma. *J Immunother Cancer* **6**: 5.

Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheim J, Vaknin-Dembinsky A, Rubertsson S, Nellgård B, Blennow K, Zetterberg H, et al. 2016. Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc Natl Acad Sci U S A* **113**: E1826–34.

Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO, et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**: 289–293.

Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**: 589–595.

Lui YYN, Chik K-W, Chiu RWK, Ho C-Y, Lam CWK, Lo YMD. 2002. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem* **48**: 421–427.

Ma H, Tu L-C, Naseri A, Chung Y-C, Grunwald D, Zhang S, Pederson T. 2018. CRISPR-Sirius: RNA scaffolds for signal amplification in genome imaging. *Nat Methods* **15**: 928–931.

Misteli T. 2007. Beyond the sequence: cellular organization of genome function. *Cell* **128**: 787–800.

Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P, et al. 2018. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun* **9**: 5068.

Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665–1680.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.

Sexton T, Cavalli G. 2015. The role of chromosome domains in shaping the functional genome. *Cell* **160**: 1049–1059.

Sexton T, Schober H, Fraser P, Gasser SM. 2007. Gene regulation through nuclear organization. *Nat Struct Mol Biol* **14**: 1049–1055.

Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. 2016. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell* **164**: 57–68.

Stunnenberg HG, International Human Epigenome Consortium, Hirst M. 2016. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**: 1145–1149.

Summers C, Rankin SM, Condliffe AM, Singh N, Peters AM, Chilvers ER. 2010. Neutrophil kinetics in health and disease. *Trends Immunol* **31**: 318–324.

Sun K, Jiang P, Allen Chan KC, Wong J, Cheng YKY, Liang RHS, Chan W-K, Ma ESK, Chan SL, Cheng SH, et al. 2015. Plasma DNA tissue mapping by genome-wide methylation sequencing for noninvasive prenatal, cancer, and transplantation assessments. *Proc Natl Acad Sci U S A* **112**: E5503–E5512.

Sun K, Jiang P, Wong AIC, Cheng YKY, Cheng SH, Zhang H, Chan KCA, Leung TY, Chiu RWK, Lo YMD. 2018. Size-tagged preferred ends in maternal plasma DNA shed light on the production mechanism and show utility in noninvasive prenatal testing. *Proc Natl Acad Sci U S A* **115**: E5106–E5114.

Ulz P, Thallinger GG, Auer M, Graf R, Kashofer K, Jahn SW, Abete L, Pristauz G, Petru E, Geigl JB, et al. 2016. Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat Genet* **48**: 1273–1278.

von Vietinghoff S, Ley K. 2008. Homeostatic regulation of blood neutrophil counts. *J Immunol* **181**: 5183–5188.

Wan JCM, Massie C, Garcia-Corbacho J, Mouliere F, Brenton JD, Caldas C, Pacey S, Baird R, Rosenfeld N. 2017. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* **17**: 223–238.

Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP. 2018. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**: 591–602.