



# Breaking Down Cell-Free DNA Fragmentation: A Markov Model Approach

Terence H.L. Tsui<sup>1</sup><sup>\*</sup>, Phil F. Xie<sup>2</sup>, Salvador Chulián<sup>3,4,5</sup>, Víctor M. Pérez-García<sup>5</sup>


**1** Department of Statistics, University of Oxford, Oxford OX1 3LB, United Kingdom

**2** Ludwig Institute for Cancer Research, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7DQ, United Kingdom

**3** Department of Mathematics, Universidad de Cádiz, 11510 Puerto Real, Cádiz, Spain

**4** Biomedical Research and Innovation Institute of Cádiz (INiBICA), Hospital Universitario Puerta del Mar, 11009 Cádiz, Spain

**5** Department of Mathematics, Mathematical Oncology Laboratory (MOLAB), Universidad de Castilla-La Mancha, Ciudad Real, Spain

 These authors contributed equally to this work.

\* Corresponding author [terence.tsui@sjc.ox.ac.uk](mailto:terence.tsui@sjc.ox.ac.uk)

## Abstract

Cell-free DNA (cfDNA) is released into the bloodstream from cells in various physiological and pathological conditions. The use of cfDNA as a non-invasive biomarker has attracted attention, but the dynamics of cfDNA fragmentation in the bloodstream are not well understood and have not been previously studied computationally.

To address this issue, we present a Markovian model called FRIME (Fragmentation, Immigration, and Exit) that captures three leading mechanisms governing cfDNA fragmentation in the bloodstream. The FRIME model enables the simulation of cfDNA fragment profiles by sampling from the stationary distribution of FRIME processes. By varying the parameters of our model, we generate fragment profiles similar to those observed in liquid biopsies and provide insight into the underlying biological mechanisms driving the fragmentation dynamics.

To validate our model, we compare simulated FRIME profiles with mitochondrial and genomic cfDNA fragment profiles. Our simulation results are consistent with experimental and clinical observations and highlight potential physicochemical differences between mitochondrial and genomic cfDNA.

The FRIME simulation framework provides an initial step towards an improved computational understanding of DNA fragmentation dynamics in the bloodstream and may aid in the analysis of liquid biopsy data.

## Author summary

Cell-free DNA (cfDNA) released into the bloodstream from cells in different conditions is the basis for liquid biopsies, thus being non-invasive biomarkers of potential interest. However, the dynamics of cfDNA fragmentation in the bloodstream remain poorly understood and are yet to be studied computationally. To address this issue, we developed a Markovian model which captures the three leading mechanisms governing cfDNA fragmentation in the bloodstream: FRagmentation, IMmigration, and Exit (FRIME). The FRIME model enables the simulation of cfDNA fragment profiles by sampling from the stationary distribution of FRIME processes. Simulation results are compared with mitochondrial and genomic cfDNA fragment profiles, which show consistency with experimental and clinical observations. The FRIME simulation framework is a significant step towards understanding DNA fragmentation dynamics in the bloodstream, uncovering potential physicochemical differences between mitochondrial and genomic cfDNA, and aiding the analysis of liquid biopsy data.

## Introduction

Cell-free DNA (cfDNA) is extracellular DNA present in biofluids, particularly blood, that originates mainly from dying cells as part of normal cellular turnover or pathological processes, including cancer [1]. Assays, such as PCR and sequencing, can detect cfDNA in the bloodstream, providing clinicians with molecular information on the cell-of-origin. This created a new field called “liquid biopsy”, which is more accessible and less risky than traditional biopsies, which require invasive procedures that may cause complications such as infection, bleeding, or pain. The decreasing cost of sequencing technologies has made cfDNA a promising material for prenatal genetic diagnosis, early cancer detection, and cancer treatment decision-making [2]. Currently, over 1000 clinical trials are exploring the potential of cfDNA, and there are a few FDA-approved tests [3].

Apart from being low-risk and easier to obtain, cfDNA can provide real-time monitoring of the disease status, allowing for prompt therapy adjustments. However, several challenges must be addressed before cfDNA can become a standard clinical tool. These include the need for further validation of the methods used for cfDNA detection, identification of the most appropriate clinical applications, optimization of the cost-effectiveness of cfDNA analysis and an improved understanding of the processes behind cfDNA generation and dynamics [4].

There are many molecular features in cfDNA data, such as DNA mutation, copy number variation [5], DNA methylation [6], and histone modifications [7]. When focusing on the oncological applications of cfDNA, identifying pathological signals from physiological activities or technical background in cfDNA is non-trivial. This is because tumor-derived cfDNA usually makes up less than 10% of plasma cfDNA even at late tumor stages [8], resulting in low signal-to-noise ratio. Sometimes, tumor cfDNA cannot be detected despite using highly sensitive methods such as digital PCR [9]. Strategies to overcome this problem include enriching for tumor-derived DNA [10], as well as combining multi-modal information to increase the number of tumor-specific features that can be captured or analyzed [11, 12].

CfDNA fragment length profile is one of the keys in both strategies. It has been demonstrated in mice xenograft models and cancer patients that tumor-derived cfDNA are in general shorter and more fragmented [10, 13], therefore selecting for short DNA fragments between 90-150 base pairs (bp) can enrich tumor signals. The short-to-long fragment ratio has also been proposed to be a diagnostic marker [14]. Yet, studies so far have been focusing on fragment sizes under approximately 200bp, and fragment length

profiles beyond this have not been thoroughly described. Moreover, how to best integrate the fragmentation profile with other multi-modal information remains an open question.

Currently, cfDNA data is commonly perceived by biologists as a fixed snapshot, with testing for specific markers of interest, such as mutation frequency. Nevertheless, cfDNA is part of a dynamic equilibrium similar to other biological processes, constantly being generated, broken down, and eliminated from the body. Although the kinetics of cfDNA can potentially be analyzed during tumor resection [15] or caesarean section [16], these approaches are not practical for diagnostic or molecular characterization purposes.

Fragmentation patterns have been proposed to have a high potential of containing information of clinical relevance in what it is currently being referred to as cfDNA fragmentomics. Further insights into the fragmentomics of plasma cfDNA may shed light on the origin and fragmentation mechanisms of cfDNA during pathological processes in diseases and enhance our ability to take the advantage of plasma cfDNA as a molecular diagnostic tool. This includes extracting information from fragment length, end motifs, jagged ends, preferred end coordinates, as well as nucleosome footprints, open chromatin region, and gene expression inferred by the cfDNA fragmentation pattern across the genome [20–22].

However, to unveil the cfDNA properties, including that of fragmentation patterns, as a potential biomarker for early diagnosis, diagnosis and prognosis, it is crucial to characterize the intrinsic kinetics of cfDNA within the organism. This is so especially when using cfDNA analysis as longitudinal diagnostic tool for monitoring the course of treatment [17]. Consequently, computational models that capture the dynamics of cfDNA may serve as a valuable tool to extract biological insights from the static cfDNA data.

Mathematicians and physical chemists have used fragmentation equations to model the degradation of polymers [18,19]. Fragmentation equations are differential equations that describe how the concentration of fragments at different length scales changes over time. However, these equations did not account for the biological noise that occurs naturally or the advent of DNA sequencing technology. While fragmentation with immigration processes have been studied [23,24], no mathematical or computational model has been built including an exit mechanism to study the effect of cfDNA clearance mechanisms on the fragment profile. Simple approaches have estimated basic averaged kinetic parameters for different processes [17] but without incorporating the complexities of fragmentation processes. Recent studies based on non-dynamical approaches have correlated tumor detection size with circulating tumor DNA shedding [27] and developed a mathematical equation relating the distribution profile of a stochastically fragmented DNA sample to the probability that a DNA fragment within that sample can be amplified by any PCR assay of arbitrary length [26]. Finally, data analytic mathematical and machine learning method have been used extensively to analyze cfDNA fragmentation datasets [25,28,29]. Nevertheless, those approaches have not incorporated mechanistically the dynamical interplay of the different processes playing a role in the fragmentation dynamics.

CfDNA samples collected before and after treatment can provide an expanded view of the genetic response of a patient's tumor, including the dynamic changes in the mutational landscape as well as the heterogeneity that develops due to the selective pressure of therapy. In addition, it is clear that the potential of cfDNA analyses is enhanced when combined with results obtained from tissue biopsies and integrated with mathematical modelling of tumor evolution. [30].

Therefore, simulation-based computational frameworks are required to gain a deeper understanding of the biological processes that occur during DNA fragmentation and to provide a more robust clinical analysis, and thus develop biomarkers, of sequenced cfDNA

fragment profiles. 88

In this study, we develop a mathematical/computational stochastic Markovian model and propose that the dynamic behaviour of cfDNA can be retrieved from the fragment length distribution of cfDNA, obtained from sequencing. When a long DNA molecule is cleaved by DNase, the resulting shorter molecules remain in the system, which would be further split into even smaller molecules. When the system reaches dynamic equilibrium, the abundance of molecules of each size would be dependent on the fragmentation behaviour. 89  
90  
91  
92  
93  
94  
95

The structure of this paper is as follows: In the Methods section, we describe the biological factors that determine the dynamics of cfDNA and propose the FRAGMENTATION with IMMIGRATION and EXIT (FRIME) framework, along with an algorithm to implement it. In the Results section, we use the FRIME framework to simulate the distribution of cfDNA fragment lengths and test various biological assumptions by adjusting the parameters of the model. We then compare our simulated results with real-world cfDNA data obtained using state-of-the-art sequencing techniques. In the Discussion section, we explore the biological implications of our results. Our simulations elucidate the shape of the entire cfDNA fragmentation profile, including the fragment counts of high molecular weight fragments. Moreover, our study identifies a difference in fragmentation kinetics between genomic and mitochondrial cfDNA, revealing new insights into how cfDNA is metabolized. We also provide a novel framework to study the dynamic parameters of cfDNA from existing data, which may improve analytic strategies. By shedding light on the underlying biological mechanisms of cfDNA fragmentation, our study contributes to the development of cfDNA as a noninvasive biomarker in clinical settings. 96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111



## Materials and methods

### Biological determinants of cfDNA dynamics

Various pathways of cell death, degradation, and regulated extrusion, partial or complete genomes of various origins (e.g., host cells, fetal cells, and infiltrating viruses and microbes) contribute to the shedding into human body fluids, most notably blood, of segmented cell-free DNA (cfDNA) molecules [32]. Active and passive release and elimination mechanisms contribute to the composition of cfDNA. In this manuscript we assumed these macromolecules to be continuously produced, metabolized, and removed from blood through the interplay of the following processes:

**Production.** CfDNA originates mainly from dying cells, which can be part of normal cellular turnover or pathological processes such as cancer [1]. When cells die, mitochondrial and genomic DNAs are released into the bloodstream. There are variations in blood cfDNA levels among patients with different tumor types and stages. In general patients with benign lesions or with early-stage cancer have lower amounts of cfDNA compared to patients with advanced or metastatic tumors of comparable size, what could be a result of the increased cell turnover and metabolic properties of progressing cancer [35]. It has been found that cfDNA levels were correlated with metabolic disease volume, estimated with  $^{18}\text{F}$ -labelled fluorodeoxyglucose positron emission tomography, in some cancers. In our modelling framework, cfDNA input mechanisms, whatever their origin, were incorporated as an input term feeding large fragments sizes into the system.

**Fragmentation rate.** CfDNA can be digested both intracellularly during cell death and extracellularly through different processes. The fragmentation rate of cfDNA depends on the DNase activity, which can differ across different body fluids [31], health conditions [33], and individuals. In the context of plasma cfDNA, it is digested extracellularly by circulating DNases that break partially digested cfDNA into even smaller fragments.

**Fragmentation pattern.** Healthy and diseased cells exhibit a structured genomic DNA arrangement, with nucleosomes tightly binding most of the DNA except for short linker stretches [34]. DNases selectively digest the linker DNA, creating a periodic pattern of fragment sizes, where each peak corresponds to DNA associated with nucleosomes. This pattern is referred to as a nucleosomal pattern. In contrast, mitochondrial DNA is not bound by nucleosomes and lacks a nucleosomal fragmentation pattern.

**Elimination.** The elimination mechanisms of circulating cfDNA are not fully understood. Nucleosome-bound cfDNA is thought to mainly be cleared by the liver, whereas naked single-stranded DNA may be cleared through the kidney in a size-dependent manner [35]. Since the concentration of cfDNA is more or less in a dynamic equilibrium, it can also be inferred that elimination of cfDNA is concentration-dependent. The reported half-life of cfDNA is between 0.25 to 2 hours [15, 16].

## Technical determinants of cfDNA fragmentation data

153

There are multiple factors that may cause a systematic bias in cfDNA fragment lengths readings, including the choice of cfDNA extraction kit [36], single-stranded versus double-stranded library preparation [37], and most notably, short-read versus long-read sequencing platform. Short-read sequencing platforms are designed for short DNA molecules, and are not ideal for our study for a few reasons. Firstly, the sequencing length is not guaranteed to cover the entire DNA molecule, and fragment length is inferred from the reference genome after sequence alignment. Thus, fragment length inference is affected by alignment errors as well as host-specific genetic variations. Secondly, there is a fragment length bias in short-read sequencing, where large DNA fragments are captured very inefficiently. This may be an effect of PCR during library preparation or the sequencer itself [38]. On the other hand, long-read sequencing is capable of directly measuring the fragment length by reading through the whole DNA molecule. It can also capture large DNA fragments efficiently with a modest size bias [39].

154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167

## Computational model

To model the dynamics of cell-free DNA fragments in the bloodstream, we proposed the FRIME framework which integrated the mechanisms mentioned above into a continuous time Markov process.

See Fig 1 for a schematic diagram of the process.

**FRAGMENTATION MECHANISMS.** CfDNA fragmentation is known to occur through DNase digestion, however it remains unclear how exactly these biomolecules are broken down and how does the reaction rate of cfDNA digestion depend on their size. Reaction rates of other polymers are known to be proportional to polymer sizes raised to some power [40]. Similar results can also be deduced mathematically under the Zimm chain model which considers random movement of polymers [41]. In this paper we considered different types of fragmentation rates and tested their outcomes against human data to see which one provided a faithful description of the observed phenomenology. Such comparison with clinical data can help us find possible biological mechanisms for cfDNA digestion.

We also assumed that a fragment of length  $x$  reacts with an enzyme at a rate  $C_F x^{\alpha_F}$ . Upon reaction, the fragment breaks into two pieces of length  $xr, x(1-r)$ , where  $r \sim \text{Beta}(a, 1)$ . In this paper, the parameters  $C_F, \alpha_F, r$  will denote the *fragmentation speed*, *fragmentation power* and *fragmentation ratio*, respectively.

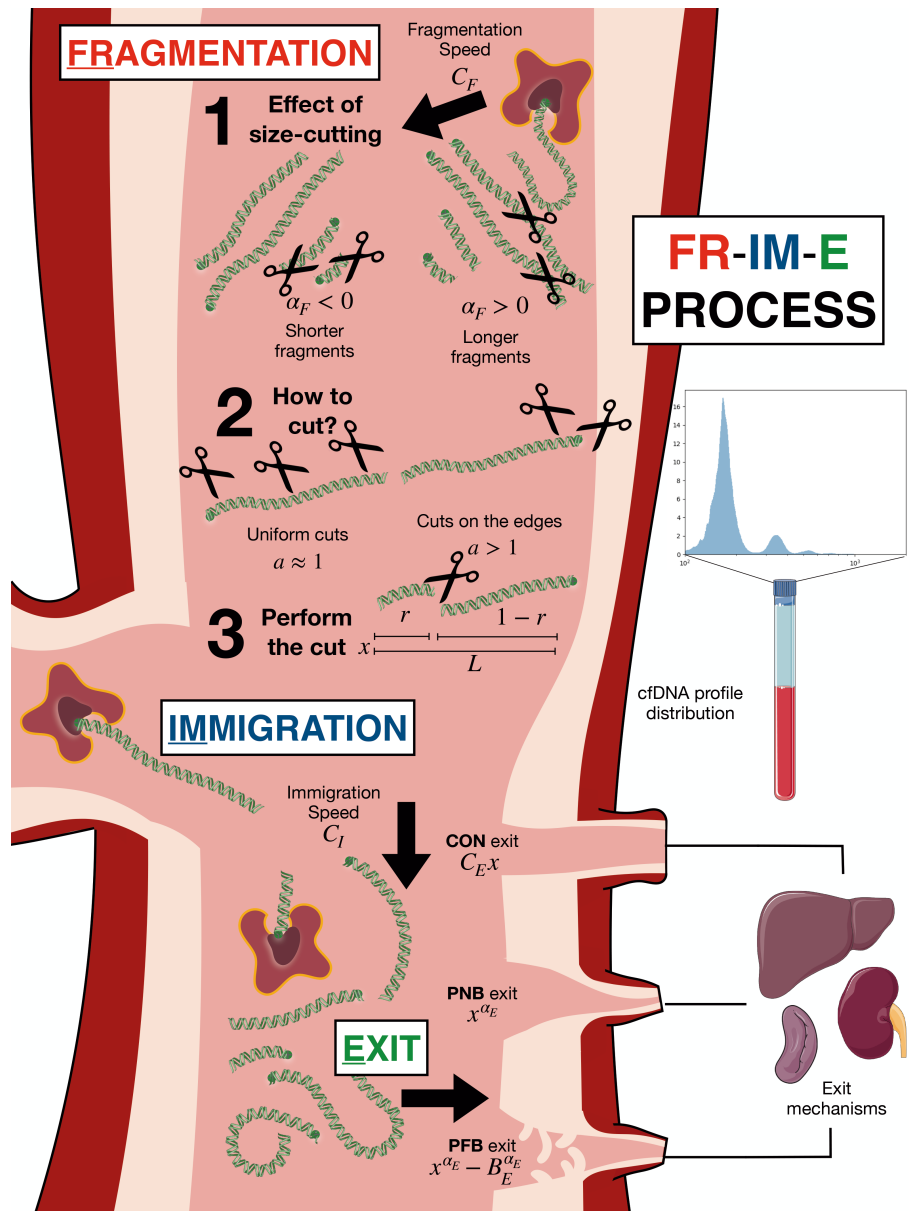
The case of  $\alpha_F < 0$  corresponds to an increase of fragmentation rate with shorter fragment lengths. This corresponds to the hypothesis that the reaction rate of cfDNA polymers are limited by diffusion, and longer cfDNA polymers diffuse slower than shorter polymers due to its size and possible secondary structures.

The case  $\alpha_F > 0$  conversely assumes that fragmentation rate increases with longer fragment length. This scenario corresponds to cfDNA digestion by surface exposure, where the rate would be proportional to the number of available reaction sites.

The parameter  $a$  determines the distribution of the fragmentation ratio  $r$ . When  $a = 1$ , the fragmentation ratio  $r$  is given by  $\text{Beta}(1,1)$ , and is thus uniformly distributed in  $[0, 1]$ . Biologically, this would correspond to the situation where cfDNA is digested by endonucleases, i.e. fragments are cut in the middle. When  $a \gg 1$ , the fragmentation ratio  $r$  is much biased to values close to 1. This last situation would be where the dominant fragmentation mechanism is digestion by exonucleases, i.e. fragments are cut near the edges. It is known that the major intracellular and extracellular DNases responsible for cutting cfDNA are not exonucleases [1, 42], thus in we chose  $a = 1$ . As described in the paragraph *Fragmentation pattern*, genomic DNA is nucleosome-protected, and fragmentation is not uniform. However, we still assumed uniform cutting on a nucleosomal scale.

**IMMIGRATION MECHANISMS.** We assumed that the maximum possible length of an immigrated fragment is  $L$ , that would be equal to or smaller than the size of a full DNA molecule. Fragments of all sizes between  $[0, L]$  immigrate into the system with rate  $C_I$ . As there was no prior knowledge on the length distribution of freshly released cell-free DNA, the length of new fragments were assumed to be uniformly distributed in the interval  $[0, L]$ . The parameter  $C_I$  was chosen to denote the *Immigration speed*.

An alternative is to assume uniform cutting by intracellular DNases of DNA before entering the bloodstream. This scenario would result in more short fragments than long fragments, and could be modelled using a  $\text{Beta}(1, n)$  distribution for the immigrating fragment size distribution. The stationary fragment distribution profiles under such an assumption were also studied in this manuscript and described in detail in Section 1 of the supplementary material S1 Appendix.



**Fig 1. Schematic diagram of the FRIME process of the distribution of cfDNA fragments with length  $x$ .** The FRAGMENTATION mechanisms included the effect of: 1) fragmentation speed  $C_F$  and fragmentation power  $\alpha_F$  (negativity and positivity related to shorter and longer fragment lengths, respectively); 2) the cut location, with  $a = 1$  as uniform cuts, and  $a \gg 1$  as cuts near the edges; 3) the fragmentation ratio  $r$ , which followed a distribution  $Beta(a, 1)$  for any possible fragment of length  $L$ . The IMMIGRATION processes were modelled by an immigration speed  $C_I$ . For the EXIT mechanisms we considered three different possibilities. First a constant exit rate (CON, with  $C_E$  as the constant), next a Power decay with No Boundary (PNB, with  $\alpha_E$  as the exit rate), and finally a Power decay exit with Fixed Boundary (PFB, with  $\alpha_E$  as the exit rate and  $B_E$  as the boundary). Icons used are originally from <https://bioicons.com>, with license <https://creativecommons.org/licenses/by/3.0/>.

**Exit mechanisms.** We considered three different classes of exit mechanisms: a constant exit rate (CON), Power decay with No Boundary (PNB), and a Power decay exit model with a Fixed Boundary (PFB). 217  
218  
219

For the CON model, we assumed that fragments of all sizes exit at the same rate  $C_E$ . This implies that cfDNA clearance mechanism is agnostic to cfDNA size. The PNB model corresponds to the assumption that fragments of length  $x$  exit the system at a rate proportional to  $x^{\alpha_E}$  with  $\alpha_E < 0$ . That functional dependence describes situations where all fragments leave the system but smaller fragments are easier to release from circulation. Finally, the PFB model is based on the assumption that only fragments smaller than  $B_E$  exit the system and do so at a rate proportional to  $x^{\alpha_E} - B_E^{\alpha_E}$ . This modelling scenario is in-line with the the renal clearance hypothesis, where only short cell-free DNA fragments can pass through the glomerular capsules and be cleared. 220  
221  
222  
223  
224  
225  
226  
227  
228

In the following sections of this paper, the parameter  $\alpha_E$  is referred as the *exit power* and the parameter  $B_E$  as the *exit boundary*. A summary of all three exit mechanisms to be studied in this paper can be found in Table 1. One of our goals in this paper was to study which mechanism is more likely to be the one responsible for cfDNA clearance. 229  
230  
231

Exit Mechanism	$\mathcal{E}(x)$
CON	$C_E$
PNB	$L^{-\alpha_E} x^{\alpha_E}, \alpha_E < 0$
PFB	$L^{-\alpha_E} \max\{x^{\alpha_E} - B_E^{\alpha_E}, 0\}$

**Table 1. Functional dependencies of the exit rates  $\mathcal{E}(x)$  on the fragment sizes  $x$  for the different exit mechanisms considered in this paper.** 232

**FRIME Process and equilibrium measure.** To model cfDNA fragmentation profile, a Markov model with the above fragmentation, immigration, and exit mechanisms was studied. Fragment profiles were sampled from the equilibrium measure of the model. More specifically, Monte Carlo simulations were run for a continuous-time Markov Chain (FRIME process) satisfying the following properties: 233  
234  
235  
236  
237

- The process starts off with a fragment with length  $x_i(0) \sim U(0, L)$ . 238
- Fragments of length  $x$  at rate  $C_F x^{\alpha_F}$  fragment into two pieces of length  $xr$  and  $x(1-r)$ , where  $r \sim \text{Beta}(a, b)$ . 239  
240
- Fragments of length  $x < B_E$  leave the system at rate  $\mathcal{E}(x)$ . 241
- New fragments immigrate into the system as a Poisson process with intensity  $C_I$ . The length  $z$  of a new fragment is given by the uniform distribution between  $[0, L]$ . 242  
243

## Algorithm

The following algorithm was used to generate one simulation of a stationary profile of the FRIME process.

1. Generate  $x_0 \sim U(0, L)$ . Assign  $\mathcal{L} = [x_0]$ . Assign  $\mathcal{L}^{hist} = \mathcal{L}$ . Assign  $d = 1, T = 0$ .
2. While  $d < \kappa$  :
  - (a) Generate  $t_{imm} \sim Exp(C_I L)$ ,  $t_{frag} \sim Exp(\sum_{x \in \mathcal{L}} C_F x^{\alpha_F})$  and  $t_E \sim Exp(\sum_{x \in \mathcal{L}} \mathcal{E}(x))$
  - (b) Assign  $t_{min} = \min(t_{imm}, t_{frag}, t_{exit})$ . Assign  $T = T + t_{min}$ .
    - If  $t_{min} = t_{frag}$ , sample  $x_{frag} \in \mathcal{L}$ , such that the probability of sampling  $x_i$  is proportional to  $x_i^{\alpha_F}$ .  
Generate  $r \sim Beta(a, 1)$ .  
Assign  $\mathcal{L} = \mathcal{L} - [x_{frag}] + [rx_{frag}, (1-r)x_{frag}]$
    - If  $t_{min} = t_{imm}$ , generate  $x_{new} \sim U(0, L)$ . Assign  $\mathcal{L} = \mathcal{L} + [x_{new}]$ .
    - If  $t_{min} = t_E$ , sample  $x_E \in \mathcal{L}$ , such that the probability of sampling  $x_i$  is proportional to  $\mathcal{E}(x_i)$ .  
Assign  $\mathcal{L} = \mathcal{L} - [x_E]$ .
  - (c) Assign  $\mathcal{L}^{hist} = \mathcal{L}^{hist} + \mathcal{L}$ . Assign  $d = KS_{Dist}(\mathcal{L}, \mathcal{L}^{hist}/T)$ .

In step 1, the initial conditions of the FRIME process are set up. The list  $\mathcal{L}$  is the current fragment length distribution and the list  $\mathcal{L}^{hist}$  is the list of fragment lengths throughout the history of the simulation.

In the while loop,  $\kappa$  is the threshold of tolerance for the metric  $d$ , which is the Kolmogorov-Smirnov distance between the current fragment list  $\mathcal{L}$  and the historical average  $\mathcal{L}^{hist}/T$ , where  $T$  corresponds to the current time step.

In step 2(a), the next immigration, fragmentation and exit times  $t_{imm}, t_{frag}, t_E$  are generated and in step 2(b) the present time  $T$  is updated to  $T + \min\{t_{imm}, t_{frag}, t_E\}$  by the theory of competing exponential. The fragment list  $\mathcal{L}$  is also updated by different mechanisms according to which simulated event comes first.

In step 2(c), the present fragment profile is loaded into the historical fragment list, and the metric  $d$  is calculated as the Kolmogorov-Smirnov distance between the current fragment list  $\mathcal{L}$  and the historical average  $\mathcal{L}^{hist}/T$ . The process terminates when  $d$  is smaller than some preset tolerance threshold  $\kappa > 0$ .

Note that  $T$  is an auxiliary variable and does not reflect biological time, and is independent of the stationary profile of a FRIME process.

## Computing machines and software

All simulations and clinical data analysis were run on Python 3.9.12 on a MacBook Air M2 2022 with 8GB Memory and the Apple M2 Chip with operating system macOS Ventura 13.3.1. Python packages required for the simulations are Numpy version 1.23.4, Scipy version 1.7.3, Matplotlib version 3.5.1, Sklearn version 1.1.3, Seaborn version 0.11.2 and Pandas version 1.4.2. Simulations of FRIME processes were computationally light; all simulations were run within 3 hours. Clinical data analysis took fewer than 2 hours. All codes for this manuscript are accessible on <https://github.com/thltsui/cfDNA-FRIME>.

## Data availability and preprocessing

Published cfNano data from Katsman et. al. [44] were used in this study, comprising of cfDNA Nanopore long-read sequencing data from 7 healthy control subjects and 6 lung adenocarcinoma patients. Shallow sequencing was done at about  $0.2\times$  genome coverage and varies across samples. Anonymised sequence alignment files in BAM format were downloaded from [43].

Alignment and data pre-processing have been previously described in the original publication [44]. Fragment lengths in base pairs were extracted from properly mapped reads using samtools version 1.8 by excluding unmapped reads, non-primary alignment reads, and supplementary alignment reads. Fragment length information was then aggregated to obtain the number of fragments for each unique length in genomic DNA (gDNA) and mitochondrial DNA (mtDNA). We considered gDNA and mtDNA separately because they exhibit very different fragmentation profiles, as the former is nucleosome-bound and has a multi-peak nucleosomal pattern, while the latter is not nucleosome-bound and only has one peak.

## Model fitting and comparison

**Numerical experiment.** To study the effect of each parameter of the FRIME model on its stationary profile, numerical experiments were run where one parameter of interest was assigned multiple values with all other parameters fixed.

For each parameter regime, a stationary profile was generated by running 50 simulations with different seeds. For all experiments, we set  $L = 1, \kappa = 0.05$ . The sample average, lower quartile and upper quartile of fragment count across samples at each length scale were plotted. The simulated data were also compared to several model-specific best fitting curves for which analytic expressions were available from the theoretical analysis (Table 2). In Section 4 of S1 Appendix, we showed that the evolution of the fragment size distribution of a FRIME process could be described as a solution to a class of partial differential equation. Furthermore, the mean of the stationary distribution of FRIME processes could also be approximated by the stationary solution to the class of partial differential equations. As a result, we could justify that FRIME processes converge to a stationary fragment profile with mean fragment size given by the expression set out in Table 2. For a detailed construction of these stationary best fit curves, we refer to Section 4.2 of the supplementary material.

**FRIME simulation fitting.** To test how well FRIME model fits clinical data, stationary profiles from simulated FRIME processes were compared against clinical mitochondrial DNA (mtDNA) fragment profiles from [43].

As fragments longer than  $10^4$ bp are scarce, and sequencing accuracy for very long fragments is low, all fragments longer than  $10^4$  in clinical data were discarded. FRIME



Model	Extra Assumptions	Stationary solutions ('Best Fit' Curves)
CON	$r \sim U(0, 1)$	$C_1(C_F x^{\alpha_F} + C_E)^{-\frac{\alpha_F+2}{\alpha_F}}$
PNB	$\alpha_F = 1$	$C_3 z^{-\alpha_E} (C_F z^{1-\alpha_E} + 1)^{\frac{3-\alpha_E}{\alpha_E-1}}$
PFB	$r \sim U(0, 1), x > B_E$	$C_2 x^{-(\alpha_F+2)}$

**Table 2. Stationary distributions of fragment profiles under different exit mechanisms.** For each model, the functions presented describe the expected fragment count per unit length for the stationary distribution of the FRIME process given the extra assumptions on the basic FRIME model. For the CON model, we assumed that the fragmentation ratio was uniformly distributed. For the PFB model, the best fit curve only applies on the region  $x > B_E$  and only if the fragmentation ratio is uniform. For the PNB model, the stationary distribution curve is applicable if the fragmentation power is  $\alpha_F = 1$ . The details of the derivations can be found at Supplementary Section S4.2.

processes were simulated with maximum length  $L = 10^4$ . Furthermore, since cfDNAs are digested through endonuclease, we took  $r \sim \text{Beta}(1, 1)$  for our simulations.

Four clinical mitochondrial cell-free DNA fragment profiles sequenced by nanopore technology with more than 500 fragment counts were analyzed in detail. These fragment profiles were compared against simulated fragment profiles from FRIME model under the Kolmogorov-Smirnov test. The simulated fragment profiles were sampled after running a FRIME process under specific parameter regimes after  $10^6$  fragmentation / exit / immigration events.

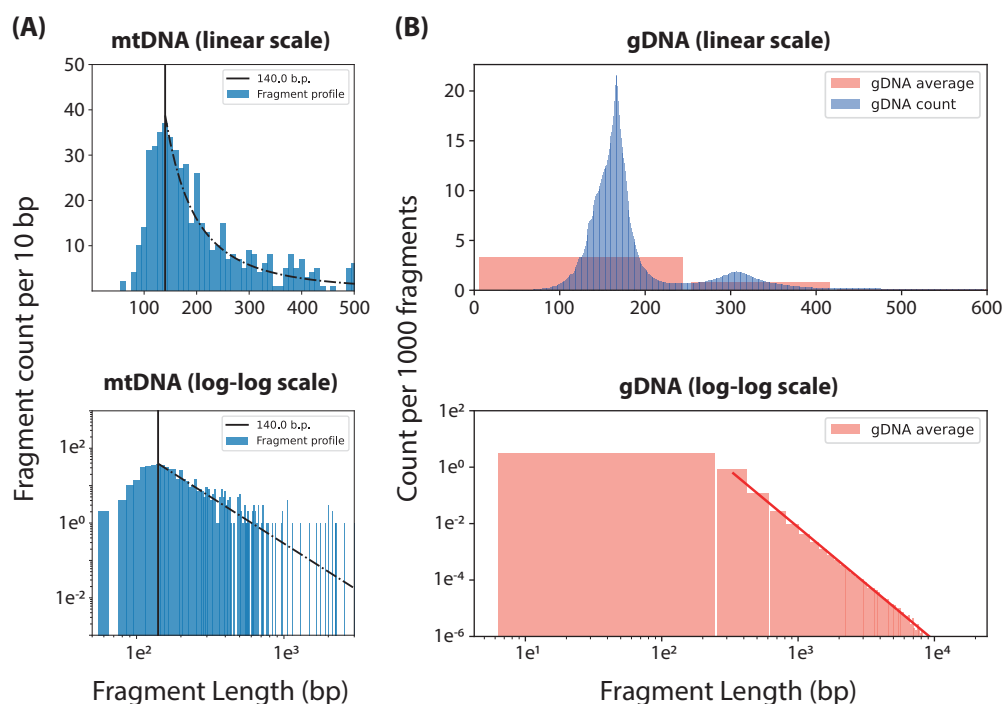
The genomic cell-free DNA fragment profile for sample ISPRO.bc05 was analyzed, which has the highest sequencing depth among healthy control samples. To adjust for the cyclical fragmentation pattern of genomic DNA (gDNA) data, fragment counts were aggregated according to nucleosomal bins (see S3 Fig for details). The similarity between simulated FRIME profiles and the genomic cfDNA fragment profile was evaluated through a  $\chi^2$  goodness of fit test.

**PFB model fitting.** Note that in Table 2, the best-fit curve for PFB model is the simplest among the three options. Furthermore, the PFB model is useful for a broad range of  $\alpha_F$  values, while the PNB model is only applicable for  $\alpha_F = 1$ . Therefore, a total number of 13 mitochondrial and genomic cell-free DNA fragment profiles in [43] were fitted with the PFB model. Specifically, the tail distributions of these fragment profiles were fitted to a curve of the form  $Cx^\beta$  (Table 2).

## Results

### CfDNA fragment profile data exhibits a peak and a linear decay trend under log-log scale

During initial data exploration, we observed that cell-free mtDNA fragment profile generated from Nanopore long-read sequencing exhibits a linear decay trend under log-log scale. While gDNA fragment profile has a cyclical, nucleosomal pattern, the linear trend remains true for gDNA if fragment counts are averaged over nucleosomal bins. The findings were illustrated using a selected sample in Fig 2. Furthermore, a clear peak in fragment count was observed for all mitochondrial and genomic cfDNA fragment profiles around 100-200bp. Plots for the full set of data can be found in supplementary material S5 Fig S6 Fig S7 Fig.



**Fig 2. Fragment size profiles of a selected sample ISPRO.bc01 in linear and log-log scale. (A)** Count per 10bp plotted against mtDNA fragment size shown in linear (top) and log-log scale (bottom). A linear line is fitted against interpolated counts in log-log scale, and is then plotted as a black dashed line in both linear and log-log scale. **(B)** Count per 1000 fragment plotted against gDNA fragment size, shown in linear (top) and log-log scale (bottom). The actual count is shown in blue, whereas the average count for each nucleosomal bin is shown in pink. A best-fit linear line for the average count is plotted in red for log-log scale.

## Exit mechanism determines existence and position of fragmentation profile peak

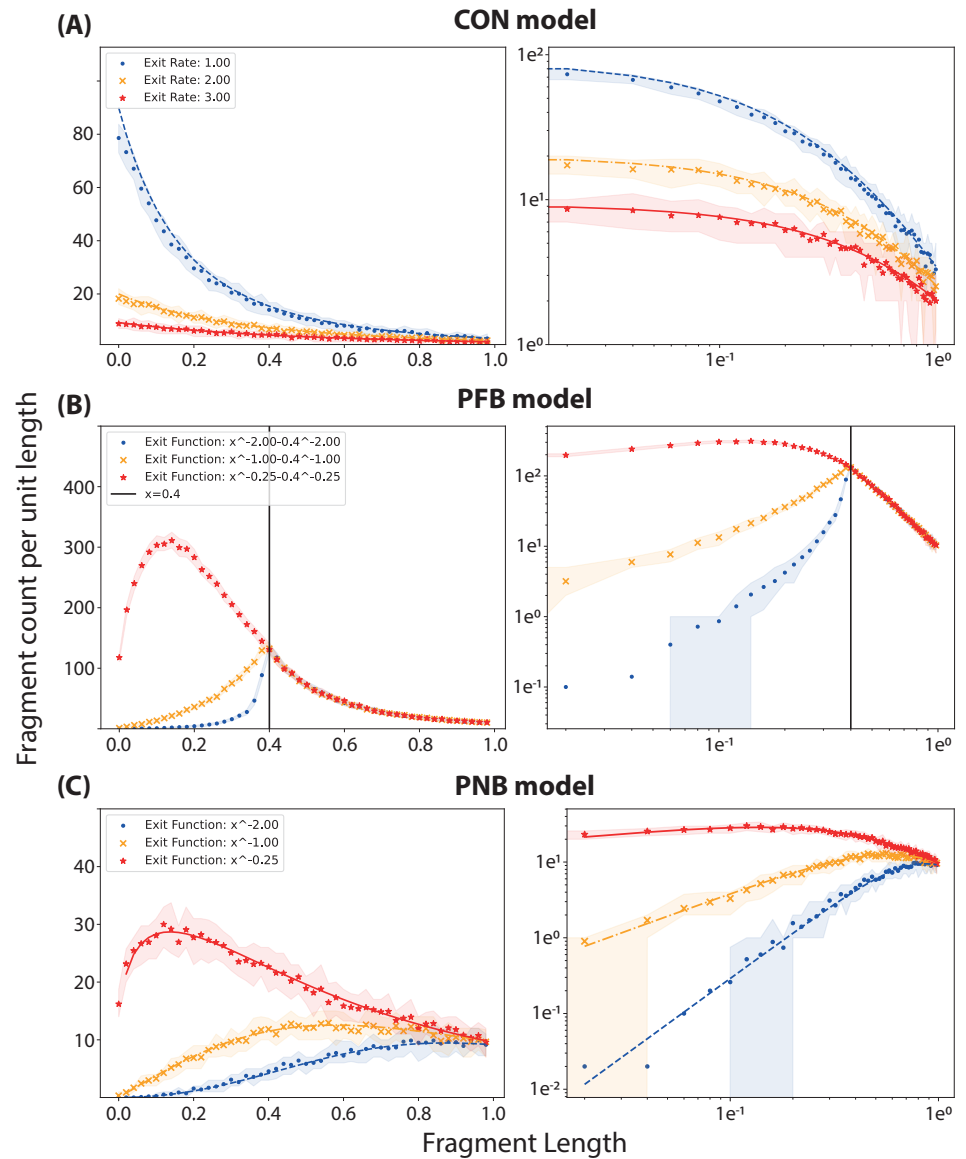
To study the stationary fragment distribution as a function of fragment size, numerical experiments were run for different parameter values for each of the three classes of exit mechanisms described in Computational model. Some results for specific parameter values are shown in Fig 3.

The results in Fig 3 showed relevant trends and features of the different models/exit mechanisms considered. Firstly, the CON model led to a monotonic decrease of the fragment count with the fragment size and did not exhibit a peak in fragment count ubiquitous in clinical cfDNA profiles.

The PFB model showed a peak in the dependence of the fragment count with size. When the boundary played a substantial role, in this case  $\alpha_E \leq -1$ , the peak was obtained on the boundary  $B_E$ , and within the interval  $[0, B_E]$  for larger  $\alpha_E$  values. For a theoretical justification of the result, refer to Section 4 of S1 Appendix.

Finally, the PNB model displayed a peak number of counts within the interval. The peak moved to smaller fragment counts and was more pronounced as the parameter  $\alpha_E$  grew. We note that the log-log plots showed that the tail-distributions for PFB profiles were linearly decreasing in a broad range of small to medium fragment sizes. However, profiles obtained using the CON and PNB models did not exhibit linearity in the log-log plots.

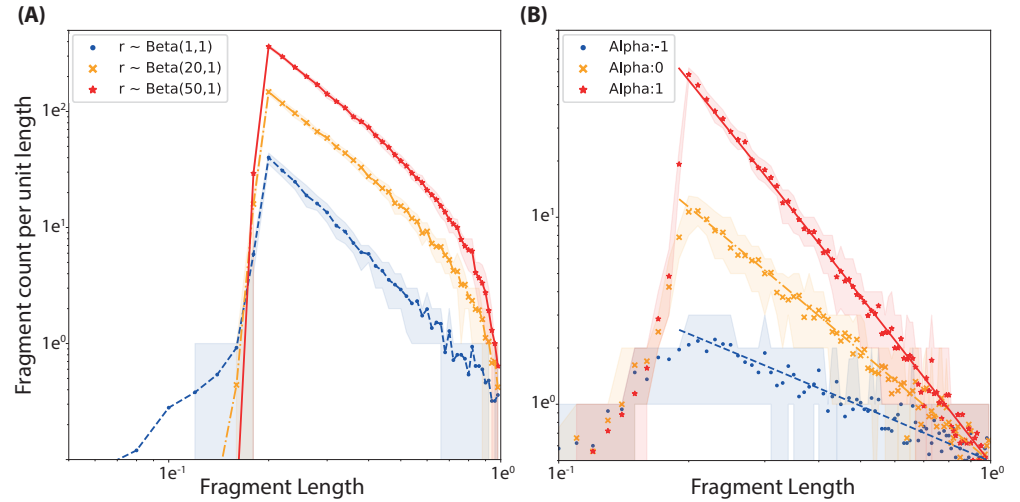
The presence of peaks and the shape of the tail distribution in the log-log plots for the PFB model simulations are consistent with the previous data exploration. Thus, in the following, we perform the subsequent analyses focusing only on the PFB model.



**Fig 3. Simulated distribution of fragment sizes for different exit mechanisms.** Stationary distributions of fragment sizes obtained from simulations of the FRIME process for parameter values  $C_I = 500$ ,  $C_F = 2$ ,  $\alpha_F = 1$ ,  $a = 1$ ,  $\kappa = 0.05$ . 50 points were plotted for each configuration. Each point corresponds to the fragment count per unit length over an interval of length 0.02. Different choices of the exit function were taken and 50 simulations run for each choice of exit function. The plots on the left column are in normal scale, while the plots on the right column are in log-log scale. **(A)** CON model with exit functions  $\mathcal{E}(x) = 1$  (blue circles), 2 (orange crosses), 3 (red stars). **(B)** PFB model,  $\mathcal{E}(x) = x^{\alpha_E} - B_E^{\alpha_E}$ , with  $\alpha_E = -2$  (blue circles),  $-1$  (orange crosses),  $-0.25$  (red stars). Parameter  $B_E$  was taken to be equal to 0.4. The vertical line was plotted at the exit boundary  $B_E = 0.4$ . **(C)** PNB model,  $\mathcal{E}(x) = x^{-2}$  (blue circles),  $x^{-1}$  (orange crosses),  $x^{-0.25}$  (red stars). Shaded regions were plotted using upper and lower quartile fragment count across all simulations. Best fit lines were also plotted for comparison for the CON and PNB model.

## Fragmentation mechanism determines tail distribution of fragment profile for PFB model

**Fragmentation ratio and fragmentation power.** To study the effect of fragmentation mechanism on stationary fragment profile, numerical experiments were simulated for three different fragmentation ratios and three different fragmentation power rates, as seen in Fig 4.



**Fig 4. Simulated fragment profiles with different fragmentation ratios and fragmentation power rates.** 50 simulations were run for each fragmentation mechanism. 50 points were plotted for each configuration corresponding to average fragment density at intervals sampled in range  $[0, 1]$  of length 0.02. Shaded regions were plotted using upper and lower quartile fragment count across all simulations. PFB best fit lines were also plotted for comparison in panel (B) with blue-dashed, orange point-dashed and red-solid lines respectively for  $\alpha_F \in \{-1, 0, 1\}$ . All plots were in log-log scale. (A) The fragmentation ratios were  $r \sim Beta(1, 1)$ ,  $Beta(20, 1)$ ,  $Beta(50, 1)$  respectively for blue-circled, orange-crossed and red-starred dots. Fixed parameters were ( $C_I = 20$ ,  $C_F = 1$ ,  $\alpha_F = 1$ ,  $\mathcal{E}(x) = x^{-2} - 0.2^{-2}$ ,  $\kappa = 0.05$ ). (B) Fragmentation power were  $\alpha_F \in \{-1, 0, 1\}$  respectively for blue-circled, orange-crossed and red-starred dots. Fixed parameters were ( $C_I = 20$ ,  $C_F = 1$ ,  $r \sim Beta(1, 1)$ ,  $\mathcal{E}(x) = x^{-2} - 0.2^{-2}$ ,  $\kappa = 0.05$ ).

In Fig 4(A), several fragmentation profiles with different fragmentation ratio were explored. A larger value in the first coefficient  $a$  of the  $Beta(a, b)$  distribution (corresponding to exonuclease hypothesis, see Fragmentation mechanisms in Computational model) leads to a steeper decay in the fragment profile tail distribution. In particular when  $a = 1$ , the fragment profile under the log-log scale is linear.

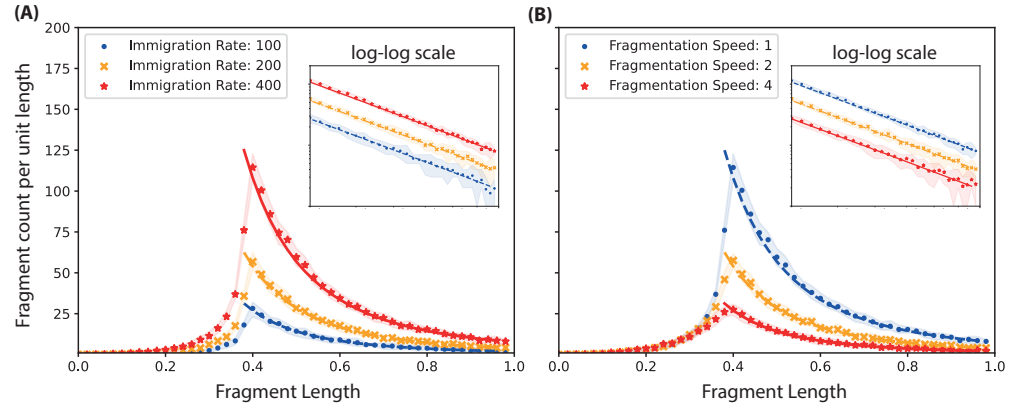
For fragmentation profile when  $r \sim Beta(1, 1)$ , the PFB best fit expression of form  $Cx^{-\alpha_F-2}$  lies within the shaded region, (see Fig 4(B)). This corresponds to a linear decrease under the log-log scale. Mathematically, if we denote  $\beta_{PFB}$  to be the slope of the best-fit line in log-log scale, then

$$\beta_{PFB} = -\alpha_F - 2. \quad (1)$$

For fragmentation profiles where  $r \sim Beta(n, n)$  for large  $n$ , refer to S2 Fig.

**Fragmentation speed and immigration speed.** To study the effect of fragmentation speed on stationary fragment profile, numerical experiments were simulated for three different fragmentation and three different immigration speeds, namely  $C_F$  and  $C_I$ . The fragmentation speeds were chosen as  $C_F \in \{1, 2, 4\}$  with fixed  $C_I = 400$ . Similarly, the immigration rates were chosen as  $C_I = 100, 200, 400$  with fixed  $C_F = 1$ .

393  
394  
395  
396  
397  
398



**Fig 5. Simulated fragment profiles with different fragmentation and immigration speeds.** 50 simulations were run for each fragmentation mechanism. 50 points were plotted for each configuration corresponding to average fragment density at  $[0,0.02], \dots, [0.98,1]$ . Shaded regions were plotted using upper and lower quartile fragment count across all simulations. The small plots were the tail distribution but in log-log scale. **(A)** The immigration rates were  $C_I = 100, 200, 400$  for blue circles, orange crosses, and red stars respectively. Fixed parameters were  $C_F = 1, \alpha_F = 1, r \sim \text{Beta}(1, 1), \mathcal{E}(x) = x^{-2} - 0.4^{-2}, \kappa = 0.05$ . **(B)** The fragmentation speed were  $C_f = 1, 2, 4$  for blue circles, orange crosses, and red stars respectively. Fixed parameters were  $C_I = 400, \alpha_F = 1, r \sim \text{Beta}(1, 1), \mathcal{E}(x) = x^{-2} - 0.4^{-2}, \kappa = 0.05$ .

The fragment profiles under the log-log scale are parallel to each other. The fragment profiles are scaled at ratio proportional to the immigration rate and inversely proportional to the fragmentation speed.

399  
400  
401  
402  
403

For fragment profiles with non-uniform immigration distribution (e.g. Exponential, Normal distributions), refer to S1 Fig.

## Simulated FRIME profiles fit clinical data

Next we wanted to study if the FRIME model was able to describe correctly the different features observed in clinical data. To do so, FRIME processes with different parameters were simulated and compared against the mtDNA fragment profile of the 4 samples ISPRO.S1, ISPRO.bc01, ISPRO.bc03 and ISPRO.bc05 from the nanopore study found in [43]. The 4 samples were chosen according to the criteria described in Model fitting and comparison.

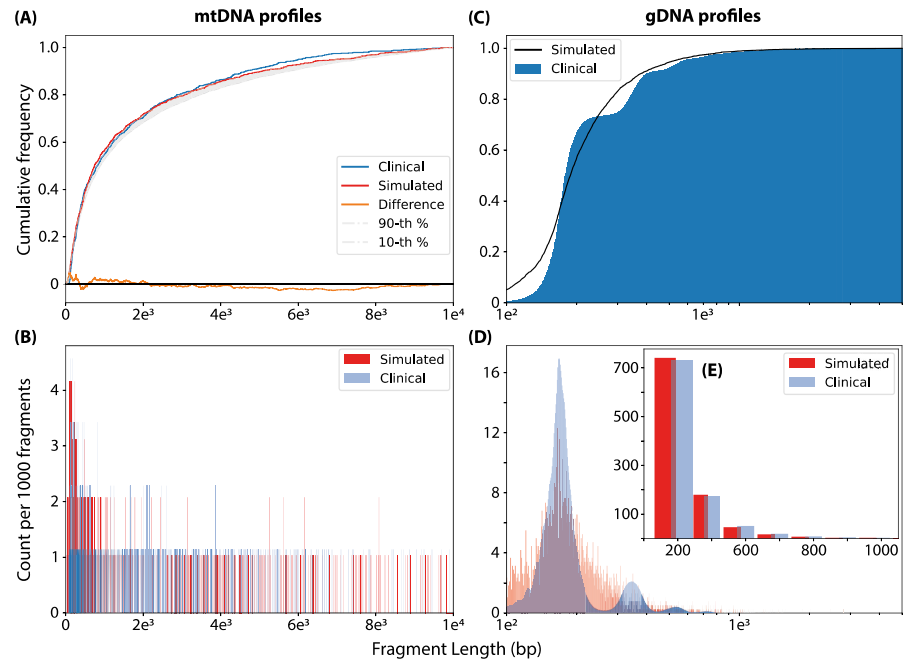
For each of the 4 mtDNA fragment profiles, there are parameter regimes which produce simulated fragment profiles with  $p$ -values larger than 0.1 under the Kolmogorov-Smirnov test, see Table 3. As the  $p$ -values are larger than 0.1, the simulated fragment profiles cannot be considered statistically different from clinical data. See Fig 6(A-B) for an example of a simulated FRIME process that fits well with clinical data.

id	$C_F$	$\alpha_F$	$C_I$	$\alpha_E$	Model	$B_E$	$d_{KS}$	$p$
ISPRO.S1	$10^6$	-1.0	11.0	2.05	PNB	N/A	0.022	0.287
ISPRO.bc01	$10^3$	0.36	10.0	2.5	PFB	115	0.049	0.345
ISPRO.bc03	$10^5$	-0.5	10.0	2.10	PNB	N/A	0.046	0.261
ISPRO.bc05	$7.5 \times 10^5$	-0.78	8.0	2.10	PNB	N/A	0.048	0.216

**Table 3. Parameters of mtDNA best fit of the clinical datasets with FRIME processes.** Simulated FRIME processes were run for  $10^6$  number of events with a fixed initial seed for generating FRIME events. The fragmentation ratio was fixed to be uniform, i.e.  $r \sim Beta(1, 1)$ , and new fragment lengths were distributed as  $U(0, L)$ , with  $L = 10^4$ . The column id corresponds to the id of the patient whose mtDNA fragment profile was studied. The constant  $C_F$  is the fragmentation speed of the FRIME process and  $\alpha_F$  is the fragmentation power. The constant  $C_I$  measures the immigration speed. The constant  $B_E$  is the exit boundary and  $\alpha_E$  the exit power. For PNB model, the exit function is of the form  $L^{-\alpha_E} x^{\alpha_E}$ ; for the PFB model, the exit function is of the form  $L^{-\alpha_E} (x^{\alpha_E} - B_E^{\alpha_E})$ . The value  $d_{KS}$  is the Kolmogorov-Smirnov distance between the simulated fragment distribution selected (best fit) and the clinical mtDNA data. The value  $p$  refers to the  $p$ -value in the Kolmogorov-Smirnov test.

Similarly, simulated FRIME profiles were found to be statistically close to genomic cfDNA fragment profiles when adjusted for nucleosomal cycles. Under the binning outlined in S3 Fig, simulated FRIME profiles and gDNA profiles were statistically similar for the datasets available. An example is shown in Fig 6(C-E).





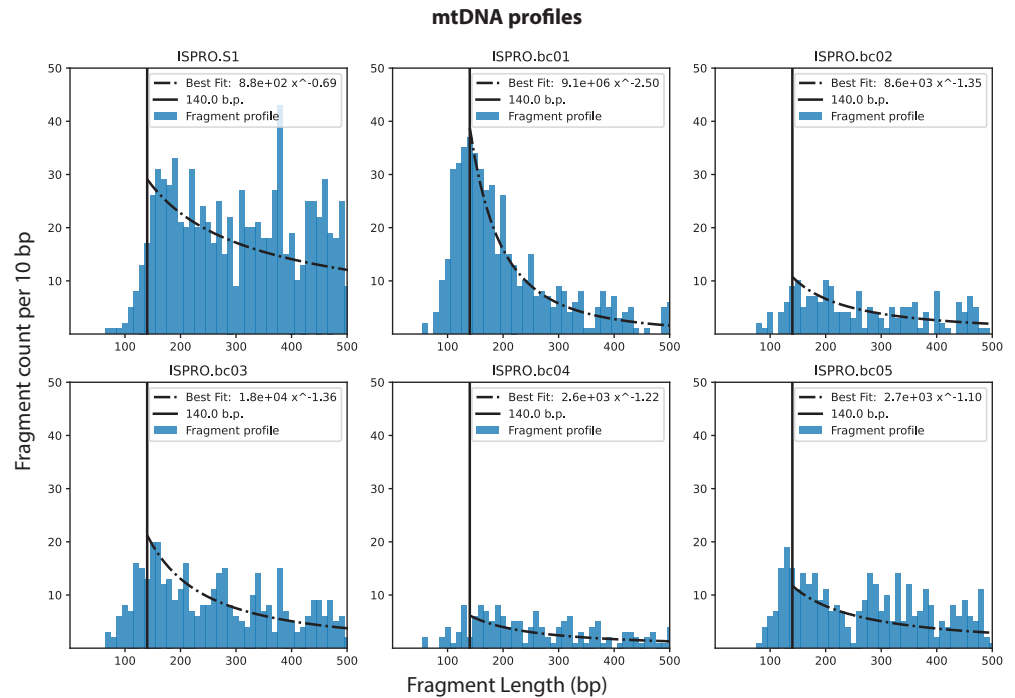
**Fig 6. Fitting clinical data for sample ISPRO.bc05 with the FRIME model.**

**mtDNA** - The mtDNA fragment profile for sample ISPRO.bc05 in nanopore study was compared against a simulated FRIME profile. The simulated profile was generated by running a FRIME model with parameters  $C_F = 10^6$ ,  $C_I = 11$ ,  $\mathcal{E}(x) = L^{2.05}x^{-2.05}$ ,  $\alpha_F = -1$ ,  $L = 10^4$ ,  $r \sim \text{Beta}(1, 1)$ , with seed = 42 until  $10^6$  events occurred. Fragment lengths were rounded to the nearest integer and plotted against clinical data. **(A)** The cumulative fragment count for the mtDNA profile of sample ISPRO.bc05 was plotted as the blue line. The mean cumulative fragment count for the simulated FRIME profile was plotted as the red line, and the shaded region was plotted using 10th and 90th percentile fragment count across all simulations. The difference between the two cumulative count was plotted as the orange line. **(B)** The fragment count (per 1000 fragments) for the mtDNA profile of sample ISPRO.bc05 was plotted as a histogram in blue colour. The fragment count (per 1000 fragments) for the simulated FRIME profile was plotted as a histogram in red colour. Fragment length was measured in base pairs. **gDNA** - The gDNA fragment profile for sample ISPRO.bc05 was compared against a simulated FRIME profile. The simulated profile was generated by running a FRIME model with parameters  $C_F = 2.5$ ,  $C_I = 10$ ,  $\mathcal{E}(x) = L^2(x^{-2} - 167^{-2})$ ,  $\alpha_F = 1.2$ ,  $L = 10^4$ ,  $r \sim \text{Beta}(1, 1)$  until  $10^6$  events occurred. Fragment lengths were rounded to the nearest integer and plotted against clinical data. **(C)** The cumulative fragment count for the gDNA profile of sample ISPRO.bc05 was plotted with the blue area. The cumulative fragment count for the simulated FRIME profile was plotted as a black line. The  $x$ -axis is under the log scale. **(D)** The fragment count for the gDNA profile of sample ISPRO.bc05 was plotted as a histogram in blue colour. The fragment count for the simulated FRIME profile was plotted as a histogram in red colour. Fragment length is measured in base pairs and the  $x$ -axis is under the log scale. **(E)** The number of fragments in the region  $[0, 250]$ ,  $[250, 420]$ ,  $[420, 620]$ ,  $[620, 820]$ , ...,  $[1220, 1420]$  for both profiles were plotted as histograms. The simulated profile was plotted in red and the fragment profile for the gDNA profile of sample ISPRO.bc05 was plotted in blue.

## Differences in PFB best-fit lines for gDNA and mtDNA profiles 420

The tail distribution of gDNA fragmentation profiles for all 13 samples in [43] were 421  
 fitted with the PFB best-fitting expression of the form  $Cx^\beta$ . Similarly, the mtDNA 422  
 fragment profiles for 9 samples in [43] were fitted. The PFB best fit lines align with the 423  
 overall pattern for all of the 9 mtDNA and 13 gDNA fragment profiles as discussed in 424  
 detail below. 425

**Best-fit statistics for mtDNA data.** For mtDNA data, the parameter  $\beta_{mtDNA}$  426  
 (slope of the best-fitting line under log-log scale) has an average value of  $-1.56$  with 427  
 range between  $[-2.5, -0.69]$ . According to Equation (1), this corresponds to a negative 428  
 fragmentation power  $\alpha_F$  of around  $-0.5$ . See Fig 7 for the best fit expressions of 6 429  
 mtDNA profiles. See S5 Fig, S6 Fig for the best-fit curves for all profiles.

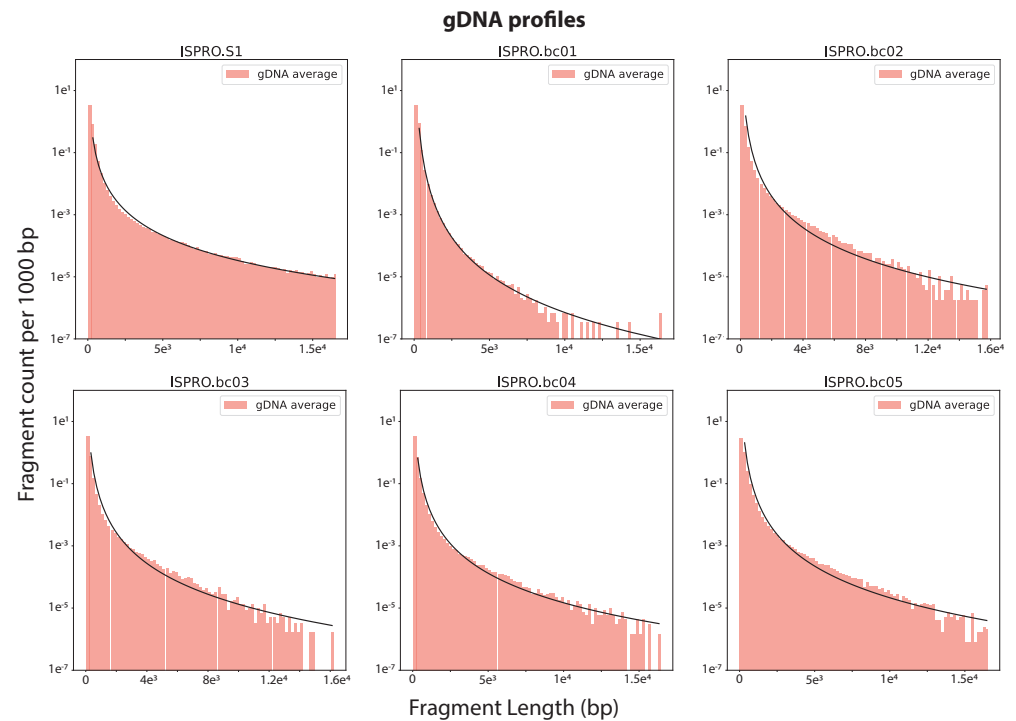


**Fig 7. Cell-free mtDNA fragments profiles and PFB best fit lines.**

Mitochondrial DNA fragment profiles for samples ISPRO.S1, ISPRO.bc01, ISPRO.bc02, ISPRO.bc03, ISPRO.bc04, ISPRO.bc05 from nanopore studies are plotted out as bar charts in blue on the top panels after linear interpolation at intervals of 10bp. Best fit curves of the expression  $Cx^\beta$  are plotted as black dot-dashed lines. The coefficients of the best fit curve are given by a linear fit under the log-log scale on the counts of fragments longer than 140bp. A black vertical line at 140bp indicating the fixed boundary is also plotted on each profile. Both  $x$  and  $y$  axes are in linear scale.

430

**Best-fit statistics for gDNA data.** For gDNA data, the parameter  $\beta_{gDNA}$  has an average value of  $-3.32$  with range between  $[-4, -2.5]$ . According to Equation (1), this corresponds to a positive fragmentation power  $\alpha_F$  of around 1. See Fig 8 for the best fit expressions of 6 gDNA profiles. See S7 Fig for the best-fit curves for all profiles. 431  
432  
433  
434



**Fig 8. Cell-free gDNA fragment profile with PFB best fit line.** Genomic DNA fragment profiles for samples ISPRO.S1, ISPRO.bc01, ISPRO.bc02, ISPRO.bc03, ISPRO.bc04, ISPRO.bc05 from nanopore studies are plotted out as bar charts in pink. Best fit curves of the expression  $Cx^\beta$  are also plotted out. The coefficients of the best fit curve are given by a linear fit under the log-log scale after binning the data. The  $y$ -axis is in log-scale.

## Discussion

In this paper, we proposed a Markov process (FRIME process) that can model cfDNA production, digestion and clearance in the bloodstream. Our FRIME model is robust and allows us to simulate cfDNA fragment profiles as the equilibrium measure of a FRIME process. Using the FRIME model, we can examine how different assumptions on fragmentation, immigration and exit mechanism produce stationary fragment profiles and compare them against real clinical data. As a result, we can deduce possible cfDNA digestion and clearance mechanisms that are impossible to study under an experimental setting. We can draw the following biological insights from our results.

**Removal mechanism of mitochondrial cell-free DNA is size-dependent and explains the existence of fragment peak.** When determining the removal mechanism of cell-free DNA, we considered the cases of size-independent exit (CON), size-dependent exit with no boundary (PNB), and with fixed boundary (PFB) and showed that size-dependent exit mechanism alone is sufficient to result in a peak. On the other hand, size-independent exit results in no peak in the absence of other factors such as protein protection.

Circulating mtDNA, unlike gDNA, is not nucleosome-bound, but demonstrate a single peak at around 100-150bp in the fragment profile, which is variable across individuals. We therefore propose that the existence of a modal fragment size in mitochondrial cfDNA can be explained by its removal mechanism, where shorter fragments are removed faster than longer fragments. Alternatively, the peak can be due to failure to capture ultra-short fragments during sequencing, or protection by protein. However, both cases should result in a uniform peak across individuals, which is not the case.

**Tail distribution of cfDNA fragment profile provides rich information when examined on a log-log scale.** In the initial data exploration, we observed that cfDNA fragment profiles exhibit a linear decay trend under log-log scale. In the subsequent modeling of exit mechanism, we showed that only the PFB model has a log-log linear tail distribution, i.e. to the right of the exit boundary. Furthermore, we examined the effect of endonuclease and exonuclease activity on the fragment profile. Log-log linearity is preserved only when all positions have the same chance of being cut (i.e. endonuclease model). In addition, we showed that the slope of the tail distribution is determined by the fragmentation power, which reflects the effect of DNA length on reaction kinetics. Therefore, log-log linearity is a very unique property that can only be met by a few mechanisms, allowing us infer the underlying biological processes.

**Fully nucleosome-bound cell-free DNA is rarely removed from circulation.** While there are periodic peaks in genomic cfDNA data, we observed that the fragment distribution binned in nucleosomal units is linear on a log-log scale, until the first nucleosome. From our simulations, this suggests that genomic cfDNA has an exit boundary at around the size of 1 or few nucleosome(s), above which the clearance rate is 0 (corresponding to the PFB model), or very small compared to the fragmentation rate (corresponding to the PNB model with small  $\alpha_E$ ). We hypothesize that short gDNA, especially those shorter than 167bp, may no longer maintain a stable complex with histone proteins, which leads to a change in physical properties and clearance rate. Our result seems to contradict with previous mice experiments, which have shown that the liver is capable of clearing, or at least trapping, circulating mononucleosomes and chromatin with radiolabelled histones [45, 46]. However, we argue that previous studies

only radiolabelled the histone component, and it is unknown whether the nucleosomes are cleared from the circulation as an intact unit with the DNA. 482  
483

**mtDNA and gDNA undergo different size-biased fragmentation mechanisms.** 484  
Parameter inference of the tail distribution from clinical data using PFB model showed 485  
that mtDNA and gDNA have systematically different slopes, as shown in S8 Fig The 486  
fragmentation power  $\alpha_F$  can be derived from the slope, and is negative for mtDNA and 487  
positive for gDNA. Hence for each molecule, smaller mtDNA fragments react faster 488  
than longer mtDNA fragments. Meanwhile, gDNA exhibits the opposite reaction 489  
kinetics, and longer fragments react faster than smaller fragments. This suggests that 490  
the degradation rate of mtDNA may be diffusion-limited. Smaller fragments react faster 491  
because they have a higher diffusion rate than longer ones, which increases the chance 492  
of colliding with a DNase. On the other hand, although mtDNA and gDNA are 493  
chemically identical, gDNA forms stable complex with histone proteins. This provides 494  
an explanation for the positive  $\alpha_F$ , which suggests that gDNA fragmentation is limited 495  
by surface area instead of diffusion. The longer the DNA molecule, the higher 496  
availability of fragmentation sites, hence higher chance of reacting with a DNase. Taken 497  
together with the discussion on the clearance mechanism of nucleosome-bound cfDNA, 498  
we propose a unifying model where nucleosome-bound cfDNA are not free-floating in 499  
circulation, but may have low solubility and form aggregates or phase-separated 500  
droplets. This mesoscale organization may sequester the nucleosomes, making them 501  
difficult for the body to remove, and cfDNA fragmentation only happens at the contact 502  
surface with plasma. This is supported by the existing literature on cell-surface bound 503  
DNA, where high molecular weight DNA is found on surfaces of blood cells [50], 504  
although the authors did not find low molecular weight DNA by gel electrophoresis, 505  
which contradicted with our hypothesis. 506

**FRIME model allows the interpretation of cfDNA concentration as biomarker of tumor.** 507  
Plasma cfDNA concentration alone is able to predict patient 508  
prognosis in multiple cancer types [47–49] and correlates well with overall tumor 509  
burden [47]. Therefore, one hypothesis is that the increased plasma cfDNA comes from 510  
increased production. However, literature also claimed that the DNase activity in cancer 511  
patients may be reduced, contributing to higher cfDNA concentration [33]. The 512  
contributing factors of cfDNA concentration has not been quantitatively studied 513  
holistically. From Fig 5, we showed that when the cfDNA fragmentation kinetics remain 514  
unchanged, the rate of production of cfDNA is directly proportional to the total amount 515  
of fragments in a patient’s bloodstream. This explains the linear correlation between 516  
cfDNA concentration and tumor burden. On the other hand, when cfDNA production 517  
rate is unchanged, the fragmentation speed is inversely proportional to cfDNA fragment 518  
counts. The FRIME model, therefore, provides a numerical framework for 519  
understanding the concentration of plasma cfDNA. 520

**Best fit coefficients in log-log scale can serve as potential biomarkers.** We 521  
showed in S5 Fig, S6 Fig that all fragment profiles in our study have an excellent fit 522  
under the PFB model visually. However, there are variations among individuals on the 523  
slope of linearity for such trends. We believe it is possible for this to be a potential 524  
biomarker for diseases. Although we do not have enough data to fully examine this 525  
hypothesis, preliminary analysis suggests that best-fit gDNA coefficients of patients 526  
with lung adenocarcinoma have higher variance than healthy patients. For further 527  
detail, refer to S8 Fig. 528

## Limitation and Future Work

We generated several testable hypotheses from our FRIME model, including the exit mechanisms, physical and chemical kinetic properties for nucleosome-bound and free-floating cfDNA. We hope these ideas can be tested with *in vitro* and *in vivo* experiments in future.

Apart from drawing biological insights, we were also able to infer different model parameters for individual patients, which are related to their cfDNA fragmentation and production rates, which may be more meaningful parameters than cfDNA concentration alone. However, one of the biggest limitations here is the lack of high depth long read cfDNA sequencing data. Most of the published cfDNA sequencing data were done using short-read sequencing platforms, which are very inefficient in capturing DNA molecules beyond 1 kb, thus not suitable for comparing with our model. At the time of our study, only one long read cfDNA sequencing dataset was available without restrictions [43]. Unfortunately, shallow sequencing was done for this dataset. We set an arbitrary cutoff of at least 5000 mtDNA molecules being sequenced in order to statistically compare with our simulation, and only 4 samples passed this cutoff. With more clinical long read sequencing data we will be able to evaluate whether the fitted model parameters from the fragmentation profile are good biomarkers.

Another limitation is that we were not able to reach a definite conclusion about the mode of cfDNA exit mechanism. Although we obtained a good fit using PNB model statistically under the FRIME framework, it cannot be guaranteed that the solution is the best fit.

Finally, our model inferred parameters based on an auxiliary time variable  $t$ . For example, we may infer  $x$  fragmentation events per unit time, but further work needs to be done to translate this into the commonly used units for DNase activity. Common used assays of DNase activity involve comparing samples to a reference with known activity [31, 51], which is often specified by the manufacturers as the amount of enzyme required to completely degrade 1  $\mu\text{g}$  of plasmid DNA in the 10 minutes at optimal working conditions. We propose that a simulation can be performed to convert our auxiliary time unit to the degradation assay units. As such, in future studies where the plasma DNase activity is measured alongside with cfDNA sequencing, reasonable fragmentation parameters can be specified prior to fitting our FRIME model, thus allowing a more accurate estimation of the cfDNA production rate.

## Conclusion

We have shown in our clinical data analysis that FRIME model highly recapitulates real life data with a few simple hypotheses, and can be used to draw inferences from mtDNA fragment profile. Although gDNA fragment profile has an inherent cyclical nature, once we adjust our fragment profile by binning data, i.e. by effectively measuring fragment lengths in units of nucleosomes, the gDNA fragment profile also fits into the FRIME model and the log-log linearity model.

We believe that our model reveals insights about cfDNA kinetics that would otherwise be unethical to measure in patients. We look forward to the general application of the FRIME model in other means of liquid biopsy as well, such as circulating RNA and proteins.

## Supporting information

**S1 Appendix Supplementary text for the manuscript.** This supplementary test includes further FRIME simulations, a detailed explanation of the statistical test used for clinical data analysis, as well as a mathematical derivation of the best fit curve for CON, PFB, PNB model.

**S1 Fig. FRIME profile with exponential and normal distributed immigration mechanism.** Stationary distributions of fragment sizes obtained from simulations of the FRIME process with immigration and fragmentation parameter values  $C_I = 500, C_F = 1, \alpha_F = 1, a = 1$ , PFB exit mechanism with parameters  $\alpha_E = -2, B_E = 0.4$ , and simulation threshold  $\kappa = 0.05$ . 50 points were plotted for each configuration. Each point corresponds to the fragment count per unit length over an interval of length 0.02. Different choices of the immigration function were taken and 50 simulations run for each choice of immigration function. Shaded regions were plotted using upper and lower quartile fragment count across all simulations. **(A)** EXP model with immigrating length  $z \sim Exp(\lambda)$  where  $\lambda = 0.1$  (dots), 0.5 ( $\times$ ), 0.7 (stars). **(B)** The simulated fragment profiles look sufficiently linear under the log-log scale. **(C)** NORM model with immigrating length  $z \sim N(\mu, 0.1\mu)$  with  $\mu = 0.5$  (dots), 0.7 ( $\times$ ), 0.9 (stars). The simulated fragment profiles clearly exhibit a peak at the mean fragment length size.

**S2 Fig. FRIME profile with strong fragmentation around midpoint.** Stationary distributions of fragment sizes obtained from simulations of the FRIME process with uniform immigration and fragmentation parameter values  $C_I = 100, C_F = 1, \alpha_F = 1$ , PFB exit mechanism with parameters  $\alpha_E = -2, B_E = 0.2$ , and simulation threshold  $\kappa = 0.05$ . 50 points were plotted for each configuration. Each point corresponds to the fragment count per unit length over an interval of length 0.02. Three different choices of the fragmentation ratios were taken and 50 simulations run for each choice of fragmentation ratio. Shaded regions were plotted using upper and lower quartile fragment count across all simulations. Right panel is normal scale, left panel is log-log scale. **(A)** Fragmentation ratio  $r \sim Beta(a, a)$  with  $a = 1$  (blue circles), 100 (orange crosses), 1000 (red stars). For large  $a$ , fragments tend to be split into two roughly equal proportions. The simulated fragment profiles exhibit big drops and flat decays. **(B)** In log-log scale, the fragment profiles for large values of  $a$  look like steps.

**S3 Fig. Genomic DNA data on log-log scale, showing the first 9 partitions.**

**S4 Fig. Evolution of  $p$ -value under KS test on FRIME simulations.** We ran two FRIME simulations and compared them against the mtDNA fragment profile and gDNA fragment profile for sample ISPRO.bc05 under the two-sample Kolmogorov-Smirnov test. **(A)** A FRIME process with parameters  $C_F = 7.5 * 10^5, \alpha_F = -0.78, L = 10^4, C_I = 8, \mathcal{E}(x) = L^{2.10}x^{-2.10}, r \sim Beta(1, 1)$  is run for  $10^5$  events. After every  $10^4$  events, the fragment profile of the process is compared against the mtDNA fragment profile under the Kolmogorov-Smirnov test. The  $p$ -value of the result is plotted in the graph. **(B)** A FRIME process with parameters  $C_F = 2.5, \alpha_F = 1.2, L = 10^4, C_I = 10, \mathcal{E}(x) = L^2x^{-2}, r \sim Beta(1, 1)$  is run for  $10^5$  events. After every  $10^4$  events, the fragment profile of the process is compared against the gDNA fragment profile under the  $\chi^2$  two-sample test. The  $p$ -value of the result is plotted in the graph.

**S5 Fig. Mitochondrial cfDNA fragment profiles with best fit curve.** Mitochondrial DNA fragment profiles for all samples from nanopore studies are plotted



out as bar charts in blue on the top panels. Best fit curves of the expression  $Cx^\beta$  are plotted as black dot-dashed lines (-). The coefficients of the best fit curve are given by a linear fit under the log-log scale on the counts of fragments longer than 140bp. A black vertical line at 140bp is also plotted on each profile. The cumulative fragment count for all fragment profiles were plotted out as bar charts in blue. Black lines plotted on top of the bars were linear interpolation of the cumulative fragment count at intervals of 10bp.

**S6 Fig. Mitochondrial cfDNA fragment profiles with best fit curve, log-log scale.** Mitochondrial DNA fragment profiles for all samples from nanopore studies are plotted out. Best fit curve of the expression  $Cx^\beta$  is also plotted out. The coefficients of the best fit curve are given by a linear fit under the log-log scale on the counts of fragments longer than 140bp. A black vertical line at 140bp is also plotted on each profile. Both axes are in log scale.

**S7 Fig. Genomic cfDNA fragment profiles with best fit curve.** Genomic DNA fragment profiles for all samples from nanopore studies are plotted out as bar charts in pink. Best fit curves of the expression  $Cx^\beta$  are also plotted out. The coefficients of the best fit curve are given by a linear fit under the log-log scale after binning the data. The  $x$ -axis is in log-scale.

**S8 Fig. gDNA best fit coefficients distribution plot.** The best-fit coefficients for linearity in log-log scale were plotted as a scatter plot. The  $x$ -axis is the gDNA slope, and the  $y$ -axis is the gDNA intercept. Healthy patients are in blue and patients with lung adenocarcinoma is in pink. The scatter plot suggests that the data for patients with lung adenocarcinoma has higher variability.

**S9 Fig. FRIME profile with different linear exit mechanism.** Stationary distributions of fragment sizes obtained from simulations of the FRIME process with uniform immigration and fragmentation parameter values  $C_I = 500, C_F = 1, \alpha_F = 1, a = 1, b = 1$  and simulation threshold  $\kappa = 0.05$ . 50 points were plotted for each configuration. Each point corresponds to the fragment count per unit length over an interval of length 0.02. Different choices of the immigration function were taken and 50 simulations were run for each choice of immigration function. Shaded regions were plotted using upper and lower quartile fragment count across all simulations. The simulations were run with linear exit mechanism  $\mathcal{E}(x) = B_E m_E - m_E x$ , where  $B_E = 0.4$  and  $m_E = 2$  (blue), 3 (orange), 4 (green). A vertical line at  $B_E = 0.4$  was plotted in black. A best-fit line of  $100x^3$  was plotted in red in the region  $x > 0.4$ . Note that when  $m_E = 4 > 3C_F$ , the peak coincides with 0.4 and when  $m_E = 2 < 3C_F$ , the peak is to the left of 0.4.

## Acknowledgments

655

T. H.L. Tsui would like to thank Jere Koskela and Jaromir Sant for their feedback and suggestions. P. F. Xie would like to thank Chi Wu for the interpretation of polymer degradation kinetics and Xin Lu for the encouraging research environment and support. T. H.L. Tsui was partially supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1), the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2047/1 – 390685813, St. John's College Oxford and the Rhodes Trust. P. F. Xie was partially supported by the Croucher Scholarship for Doctoral Study, Ludwig Institute for Cancer Research Oxford branch, Mabel Churn Scholarship, and NDM Prize Studentship. S. Chulián and V. M. Pérez-García are supported by the Spanish Ministerio de Ciencia e Innovación, MCIN/AEI/10.13039/501100011033 (grant number PID2019-110895RB-I00). V. M. Pérez-García is also supported by Junta de Comunidades de Castilla-La Mancha (SBPLY/21/180501/000145).

656

657

658

659

660

661

662

663

664

665

666

667

668

669

## Author contributions

670

<b>Role</b>	<b>Contributing Author</b>
Conceptualization	Tsui, Xie
Data Curation	Xie
Formal Analysis	Tsui
Funding Acquisition	N/A
Investigation	Tsui, Xie
Methodology	Tsui
Project Administration	Tsui
Resources	N/A
Software	Tsui
Supervision	Pérez-García
Validation	Chulián
Visualization	Chulián, Tsui, Xie
Writing – Original Draft Preparation	Tsui, Xie
Writing – Review & Editing	Chulián, Pérez-García

## Competing interests

671

All authors have no competing interests.

## References

1. Heitzer E, Auinger L, Speicher MR. Cell-Free DNA and Apoptosis: How Dead Cells Inform About the Living. *Trends in Molecular Medicine*. 2020;26(5):519–528. doi:10.1016/j.molmed.2020.01.012.
2. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science*. 2021;372(6538):eaaw3616. doi:10.1126/science.aaw3616.
3. Cisneros-Villanueva M, Hidalgo-Pérez L, Rios-Romero M, Cedro-Tanda A, Ruiz-Villavicencio CA, Page K, et al. Cell-free DNA analysis in current cancer clinical trials: a review. *British Journal of Cancer*. 2022;126(3):391–400. doi:10.1038/s41416-021-01696-0.
4. Rolfo C, Cardona AF, Cristofanilli M, Paz-Ares L, Diaz Mochon JJ, Duran I, et al. Challenges and opportunities of cfDNA analysis implementation in clinical practice: Perspective of the International Society of Liquid Biopsy (ISLB). *Crit Rev Oncol Hematol*. 2020;151:102978. doi:10.1016/j.critrevonc.2020.102978.
5. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications*. 2017;8(1):1324. doi:10.1038/s41467-017-00965-y.
6. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Liu MC, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Annals of Oncology*. 2020;31(6):745–759. doi:10.1016/j.annonc.2020.02.011.
7. Sadeh R, Sharkia I, Fialkoff G, Rahat A, Gutin J, Chappleboim A, et al. ChIP-seq of plasma cell-free nucleosomes identifies gene expression programs of the cells of origin. *Nature Biotechnology*. 2021;39(5):586–598. doi:10.1038/s41587-020-00775-6.
8. Molparia B, Nichani E, Torkamani A. Assessment of circulating copy number variant detection for cancer screening. *PLOS ONE*. 2017;12(7):e0180647. doi:10.1371/journal.pone.0180647.
9. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Science Translational Medicine*. 2014;6(224):224ra24–224ra24. doi:10.1126/scitranslmed.3007094.
10. Mouliere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Science Translational Medicine*. 2018;10(466):eaat4921. doi:10.1126/scitranslmed.aat4921.
11. Siejka-Zielińska P, Cheng J, Jackson F, Liu Y, Soonawalla Z, Reddy S, et al. Cell-free DNA TAPS provides multimodal information for early cancer detection. *Science Advances*. 2021;7(36):eabh0534. doi:10.1126/sciadv.abh0534.
12. Peneder P, Stütz AM, Surdez D, Krumbholz M, Semper S, Chicard M, et al. Multimodal analysis of cell-free DNA whole-genome sequencing for pediatric cancers with low mutational burden. *Nature Communications*. 2021;12(1):3230. doi:10.1038/s41467-021-23445-w.

13. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, et al. Fragment Length of Circulating Tumor DNA. *PLOS Genetics*. 2016;12(7):e1006162. doi:10.1371/journal.pgen.1006162.
14. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*. 2019;570(7761):385–389. doi:10.1038/s41586-019-1272-6.
15. Diehl F, Schmidt K, Choti MA, Romans K, Goodman S, Li M, et al. Circulating mutant DNA to assess tumor dynamics. *Nature Medicine*. 2008;14(9):985–990. doi:10.1038/nm.1789.
16. Yu SC, Lee SW, Jiang P, Leung TY, Chan KA, Chiu RW, et al. High-Resolution Profiling of Fetal DNA Clearance from Maternal Plasma by Massively Parallel Sequencing. *Clinical Chemistry*. 2013;59(8):1228–1237. doi:10.1373/clinchem.2013.203679.
17. Khier S, Lohan L Kinetics of circulating cell-free DNA for biomedical applications: critical appraisal of the literature *Future Sci OA*. 2018;4(4):FSO295 doi:10.4155/fsoa-2017-0140
18. Ziff RM, McGrady ED. The kinetics of cluster fragmentation and depolymerisation. *Journal of Physics A: Mathematical and General*. 1985;18(15):3027. doi:10.1088/0305-4470/18/15/026.
19. Ederer HJ, Basedow AM, Ebert KH. Modelling of Polymer Degradation Reactions. In: Ebert KH, Deuflhard P, Jäger W, editors. *Modelling of Chemical Reaction Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1981. p. 189–215.
20. Qi T, Pan M, Shi H, Wang L, Bai Y, Ge Q. Cell-Free DNA Fragmentomics: The Novel Promising Biomarker. *Int J Mol Sci*. 2023; 24(2):1503. doi: 10.3390/ijms24021503
21. Liu, Y At the dawn: cell-free DNA fragmentomics and gene regulation. *Br J Cancer* 2022; 126, 379–390. doi:10.1038/s41416-021-01635-z
22. Mouliere F, Smith CG, Heider K, Su J, van der Pol Y, Thompson M, et al Fragmentation patterns and personalized sequencing of cell-free DNA in urine and plasma of glioma patients *EMBO Mol Med*. 2021;13(8):e12881 doi: 10.15252/emmm.202012881
23. Ben-Naim E, Krapivsky PL. Fragmentation with a steady source. *Physics Letters A*. 2000;275(1):48–53. doi:10.1016/S0375-9601(00)00570-3.
24. Haas B. Equilibrium for Fragmentation with Immigration. *The Annals of Applied Probability*. 2005;15(3):1958–1996.
25. Zhou Z, Ma ML, Chan RWY, Lam WKJ, Peng W, Gai W, et al Fragmentation landscape of cell-free DNA revealed by deconvolutional analysis of end motifs. *Proc Natl Acad Sci U S A*. 2023;120(17):e2220982120. doi: 10.1073/pnas.2220982120.
26. Johnston AD, Lu J, Korbie D, Trau M. Modelling clinical DNA fragmentation in the development of universal PCR-based assays for bisulfite-converted, formalin-fixed and cell-free DNA sample analysis. *Sci Rep*. 2022 Sep 26;12(1):16051. doi:10.1038/s41598-022-18196-7.

27. Avanzini S, Kurtz DM, Chabon JJ, Moding EJ, Hori SS, Gambhir SS, et al. A mathematical model of ctDNA shedding predicts tumor detection size. *Sci Adv*. 2020;6(50):eabc4308. doi:10.1126/sciadv.abc4308
28. Renaud G, Nørgaard M, Lindberg J, Grönberg H, De Laere B, Bjerggaard J, et al. Unsupervised detection of fragment length signatures of circulating tumor DNA using non-negative matrix factorization *eLife* 2020; 11:e71569. doi:10.7554/eLife.71569
29. Moser T, Kühberger S, Lazzeri I, Vlachos G, Heitzer E. Bridging biological cfDNA features and machine learning approaches. *Trends Genet*. 2023; 39(4):285-307. doi: 10.1016/j.tig.2023.01.004
30. Khan KH, Cunningham D, Werner B, Vlachogiannis G, Spiteri I, Heide T, et al. Longitudinal Liquid Biopsy and Mathematical Modeling of Clonal Evolution Forecast Time to Treatment Failure in the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer Discov*. 2018; 8(10):1270-1285. doi: 10.1158/2159-8290.CD-17-0891
31. Nadano D, Yasuda T, Kishi K. Measurement of deoxyribonuclease I activity in human tissues and body fluids by a single radial enzyme-diffusion method. *Clinical Chemistry*. 1993;39(3):448-452. doi:10.1093/clinchem/39.3.448.
32. Grabuschnig S, Bronkhorst AJ, Holdenrieder S, Rosales Rodriguez I, Schliep KP, Schwendenwein D, et al. Putative Origins of Cell-Free DNA in Humans: A Review of Active and Passive Nucleic Acid Release Mechanisms. *Int J Mol Sci*. 2020 Oct 29;21(21):8062. doi:10.3390/ijms21218062
33. Tamkovich SN, Cherepanova AV, Kolesnikova EV, Rykova EY, Pyshnyi DV, Vlassov VV, et al. Circulating DNA and DNase Activity in Human Blood. *Annals of the New York Academy of Sciences*. 2006;1075(1):191-196. doi:10.1196/annals.1368.026.
34. Ivanov, M., Baranova, A., Butler, T. et al. Non-random fragmentation patterns in circulating cell-free DNA reflect epigenetic regulation *BMC Genomics* 16 (Suppl 13), S1 doi:10.1186/1471-2164-16-S13-S1
35. Kustanovich A, Schwartz R, Peretz T, Grinshpun A. Life and death of circulating cell-free DNA. *Cancer Biology & Therapy*. 2019;20(8):1057-1067. doi:10.1080/15384047.2019.1598759.
36. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, et al. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. *Analytical and Bioanalytical Chemistry*. 2014;406(26):6499-6512. doi:10.1007/s00216-014-7835-3.
37. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. *Cell*. 2016;164(1):57-68. doi:10.1016/j.cell.2015.11.050.
38. Gohl DM, Magli A, Garbe J, Becker A, Johnson DM, Anderson S, et al. Measuring sequencer size bias using REcount: a novel method for highly accurate Illumina sequencing-based quantification. *Genome Biology*. 2019;20(1):85. doi:10.1186/s13059-019-1691-6.

39. Yu SCY, Deng J, Qiao R, Cheng SH, Peng W, Lau SL, et al. Comparison of Single Molecule, Real-Time Sequencing and Nanopore Sequencing for Analysis of the Size, End-Motif, and Tissue-of-Origin of Long Cell-Free DNA in Plasma. *Clinical Chemistry*. 2023;69(2):168–179. doi:10.1093/clinchem/hvac180.
40. Basedow AM, Ebert KH, Ederer HJ. Kinetic Studies on the Acid Hydrolysis of Dextran. *Macromolecules*. 1978;11(4):774–781. doi:10.1021/ma60064a031.
41. Doi M, Edwards SF, Edwards SF. The theory of polymer dynamics. vol. 73. oxford university press; 1988.
42. Han DSC, Ni M, Chan RWY, Chan VWH, Lui KO, Chiu RWK, et al. The Biology of Cell-free DNA Fragmentation and the Roles of DNASE1, DNASE1L3, and DFFB. *The American Journal of Human Genetics*. 2020;106(2):202–214. doi:10.1016/j.ajhg.2020.01.008.
43. Katsman E, Orlanski S, Martignano F, Conticello SG, Berman BP. Data for for Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing; 2022.
44. Katsman E, Orlanski S, Martignano F, Fox-Fisher I, Shemer R, Dor Y, et al. Detecting cell-of-origin and cancer-specific methylation features of cell-free DNA from Nanopore sequencing. *Genome Biology*. 2022;23(1):158. doi:10.1186/s13059-022-02710-1.
45. Gauthier VJ, Tyler LN, Mannik M. Blood clearance kinetics and liver uptake of mononucleosomes in mice. *The Journal of Immunology*. 1996;156(3):1151–1156. doi:10.4049/jimmunol.156.3.1151.
46. Du Clos TW, Volzer MA, Hahn FF, Xiao R, Mold C, Searles RP. Chromatin clearance in C57Bl/10 mice: interaction with heparan sulphate proteoglycans and receptors on Kupffer cells. *Clinical and Experimental Immunology*. 1999;117(2):403–411. doi:10.1046/j.1365-2249.1999.00976.x.
47. Dziadziuszko R, Peters S, Mok T, Camidge DR, Gadgeel SM, Ou SHI, et al. Circulating Cell-free DNA as a Prognostic Biomarker in Patients with Advanced ALK+ Non-small Cell Lung Cancer in the Global Phase III ALEX Trial. *Clinical Cancer Research*. 2022;28(9):1800–1808. doi:10.1158/1078-0432.CCR-21-2840.
48. Fernandez-Garcia D, Hills A, Page K, Hastings RK, Toghil B, Goddard KS, et al. Plasma cell-free DNA (cfDNA) as a predictive and prognostic marker in patients with metastatic breast cancer. *Breast Cancer Research*. 2019;21(1):149. doi:10.1186/s13058-019-1235-8.
49. Chen E, Cario CL, Leong L, Lopez K, Márquez CP, Chu C, et al. Cell-free DNA concentration and fragment size as a biomarker for prostate cancer. *Scientific Reports*. 2021;11(1):5040. doi:10.1038/s41598-021-84507-z.
50. Laktionov PP, Tamkovich SN, Rykova EY, Bryzgunova OE, Starikov AV, Kuznetsova NP, et al. Cell-Surface-Bound Nucleic Acids: Free and Cell-Surface-Bound Nucleic Acids in Blood of Healthy Donors and Breast Cancer Patients. *Annals of the New York Academy of Sciences*. 2004;1022(1):221–227. doi:10.1196/annals.1318.034.
51. Sinicropi D, Baker DL, Prince WS, Shiffer K, Shak S. Colorimetric Determination of DNase I Activity with a DNA-Methyl Green Substrate. *Analytical Biochemistry*. 1994;222(2):351–358. doi:10.1006/abio.1994.1502.