

# Sequence analysis of cell-free DNA derived from cultured human bone osteosarcoma (I43B) cells

Tumor Biology

September 2018: 1–18

© The Author(s) 2018

Article reuse guidelines:

[sagepub.com/journals-permissions](http://sagepub.com/journals-permissions)

DOI: 10.1177/1010428318801190

[journals.sagepub.com/home/tub](http://journals.sagepub.com/home/tub)

Abel Jacobus Bronkhorst<sup>1</sup> , Johannes F Wentzel<sup>2</sup>, Vida Ungerer<sup>1</sup>,  
Dimetrie L Peters<sup>3</sup>, Janine Aucamp<sup>4</sup>, Etienne Pierre de Villiers<sup>5</sup> ,  
Stefan Holdenrieder<sup>1</sup> and Piet J Pretorius<sup>3</sup>

## Abstract

The true importance of cell-free DNA in human biology, together with the potential scale of its clinical utility, is tarnished by a lack of understanding of its composition and origin. In investigating the cell-free DNA present in the growth medium of cultured I43B cells, we previously demonstrated that the majority of cell-free DNA is neither a product of apoptosis nor necrosis. In the present study, we investigated the composition and origin of this cell-free DNA population using next-generation sequencing. We found that the cell-free DNA comprises mainly of repetitive DNA, including  $\alpha$ -satellite DNA, mini satellites, and transposons that are currently active or exhibit the capacity to become reactivated. A significant portion of these cell-free DNA fragments originates from specific chromosomes, especially chromosomes 1 and 9. In healthy adult somatic cells, the centromeric and pericentromeric regions of these chromosomes are normally densely methylated. However, in many cancer types, these regions are preferentially hypomethylated. This can lead to double-stranded DNA breaks or it can directly impair the formation of proper kinetochore structures. This type of chromosomal instability is a precursor to the formation of nuclear anomalies, including lagging chromosomes and anaphase bridges. DNA fragments derived from these structures can recruit their own nuclear envelope and form secondary nuclear structures known as micronuclei, which can localize to the nuclear periphery and bud out from the membrane. We postulate that the majority of cell-free DNA present in the growth medium of cultured I43B cells originates from these micronuclei.

## Keywords

Cell-free DNA, biomarker, liquid biopsy, transposons, satellite DNA, anaphase bridges, micronuclei

Date received: 17 April 2018; accepted: 23 August 2018

## Introduction

Molecular analysis of cell-free DNA (cfDNA) marks a new point of departure in the application of genomic methods and ultra-sensitive technologies for the non-invasive diagnosis, prognosis, and monitoring of a wide range of diseases, especially cancer.<sup>1,2</sup> Many studies have demonstrated the translational potential of cfDNA as a powerful and multifaceted biomarker. However, virtually no tests have yet been incorporated into routine clinical testing.<sup>3,4</sup> While the difficulty to do so can, to a certain extent, be attributed to the lack of analytical consensus among research groups,<sup>5–8</sup> a

<sup>1</sup>Institute for Laboratory Medicine, German Heart Centre, Technical University Munich, Munich, Germany

<sup>2</sup>Centre of Excellence for Nutrition (CEN), North-West University, Potchefstroom, South Africa

<sup>3</sup>Human Metabolomics, Biochemistry Division, North-West University, Potchefstroom, South Africa

<sup>4</sup>Centre of Excellence for Pharmaceutical Sciences, North-West University, Potchefstroom, South Africa

<sup>5</sup>Centre for Tropical Medicine, University of Oxford, Oxford, United Kingdom

### Corresponding author:

Abel Jacobus Bronkhorst, Institute for Laboratory Medicine, German Heart Centre, Technical University Munich, Lazarettstraße 36, Munich D-80636, Germany.

Email: [bronkhorst@dhm.mhn.de](mailto:bronkhorst@dhm.mhn.de)



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons

Attribution-NonCommercial 4.0 License (<http://www.creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial

use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

growing number of reports indicate that a dearth of knowledge on the origin and molecular properties of cfDNA is another obstacle that substantially precludes the rapid translation of basic research to clinical practice.<sup>9–15</sup>

Next generation sequencing (NGS) is one of the most powerful methods for elucidating the molecular properties of cfDNA. While sequencing of cfDNA has increased considerably in recent years, research has been mainly clinically motivated, focusing on size evaluation,<sup>16,17</sup> ultra-deep amplicon<sup>18</sup> or exome sequencing,<sup>19</sup> and methylation-specific sequencing.<sup>20–22</sup> In contrast, few researchers perform thorough sequence characterization. One group sequenced the cfDNA obtained from the serum of 51 healthy individuals. Although most sequences indicated an apoptotic origin, an uneven distribution of apoptotic and necrotic DNA across the genome, as well as an overrepresentation of *Alu* and L1 repetitive elements (REs), was observed.<sup>23</sup> This uneven representation is likely due to the complexity of the *in vivo* setting, that is, both the quantitative and qualitative characteristics of cfDNA in the blood of an individual at any instance are modulated by various internal processes and environmental factors.<sup>9,15</sup> Moreover, since most cells release DNA, the aggregate cfDNA profile comprises a muddled blend of mutated and wild-type DNA released by various cells from different tissues and organs by different mechanisms. This makes it very difficult to investigate the biological properties of cfDNA *in vivo*.

Since cell cultures are insulated from most external elements, we argue that many of the difficulties encountered in *in vivo* experiments can be circumvented by *in vitro* models. For example, DNA is released from only a specific cell type in a typical cell culture experiment, rather than the hundreds of different kinds of cells in a whole organism. In a previous study,<sup>24,25</sup> we characterized the DNA present in the culture medium of human bone osteosarcoma (143B) cells. After 4 h of incubation, only a small number of 166 bp DNA fragments are present. However, after 24 h, there is a significant increase in the amount and size of DNA (~2000 bp). Typically, DNA with a size of 166 bp is a product of apoptotic fragmentation, while a size of 2000 bp can be explained by neither apoptosis nor necrosis. These results were confirmed by two flow cytometric assays.

These observations raise the important question: if not apoptosis or necrosis, which mechanisms may give rise to the presence of cfDNA, with an approximate size of 2000 bp, in the growth medium of 143B cells after 24 h of incubation? Therefore, the major aim of the experiments presented in this article was to characterize the nucleotide sequence of the cfDNA that is released by cultured 143B cells after 24 h of incubation. Using this information, we infer a possible origin and function.

## Methods

### Cell culturing and growth medium processing

The human bone cancer (osteosarcoma) cell line 143B was obtained from the American Type Culture Collection (ATCC® CRL-8303™) and was cultured in accordance with ATCC protocols. Cells were grown in Dulbecco's modified Eagle's medium (Hyclone DMEM/high glucose; Thermo Scientific; #SH30243.01) containing 4 mM L-Glutamine, 4500 mg/L glucose, and sodium pyruvate, and fortified with 10% fetal bovine serum (FBS; Biochrom; #S0615) and 1% penicillin/streptomycin (Lonza; #DE17-602E). Cells were incubated in humidified atmosphere containing 5% CO<sub>2</sub> at 37°C. Cells were seeded (at ~50% confluency) and cultured in 75 cm<sup>2</sup> cell culture flasks (Nunc) and grown for 12 h in 12 mL growth medium. After this time, the growth medium was renewed and cells were further grown for 24 h. This was done in duplicate. At the end of incubation, the growth medium was collected in 15 mL nuclease-free tubes (Ambion; #3108090) and then centrifuged at 10,000 × *g* for 10 min and transferred to fresh 15 mL tubes. Samples were then stored at –80°C.

### Extraction and quantification of cfDNA and genomic DNA

The cfDNA was extracted with the NucleoSpin Gel and PCR Clean-Up kit (Macherey-Nagel, Düren, Germany; #1502/001), according to the manufacturer's PCR clean-up instructions, with slight modifications. Briefly, samples were thawed at 37°C in a temperature-controlled water bath. After incubation, the samples were vortexed and centrifuged briefly. For each biological replicate, 12 mL of growth medium was mixed with 24 mL of binding buffer NTI. Samples were then vortexed, the entire volume of growth media added to the spin column in small regiments, and moved through the filter using a vacuum manifold setup and pump. Hereafter, the columns were washed twice, followed by the elution of cfDNA into 50 µL of elution buffer. The duplicate samples were then pooled. Genomic DNA was isolated from 143B cells using the FlexiGene DNA kit (Qiagen), according to the cultured cells protocol. DNA samples were quantified using the Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies) and Qubit dsDNA HS Assay kit, and then stored at 4°C.

### DNA library generation and sequencing

Extracted cfDNA and genomic DNA (~100 ng) were sheared into 100 to 700 bp fragments by sonication with the Bioruptor UCD-200 (Diagenode), by applying three 5-min cycles of 30 s on/30 s off, on the medium setting. After shearing, the fragment ends were end-repaired

and polished with the Ion Plus Fragment Library kit. Ion A and P1 adaptors were then ligated onto the polished fragments, and 330 bp fragments were size-selected using 2% agarose gels on the Egel® system. These libraries were then amplified using the Ion Plus Fragment Library kit and quantified using the Qubit dsDNA HS Assay kit. The libraries were diluted to 100 pM and a sequencing template for each library was prepared on the Ion OneTouch2 system using the Ion PGM Template OT2 200 kit. Each sample's templated ISPs (Ion Sphere® Particles) were manually loaded onto an Ion 318v2 chip and sequenced on the Ion PGM using the Ion PGM Sequencing 200 kit v2. Except where indicated, all kits and reagents were obtained from Thermo Fisher Scientific. Default parameters on the Ion Torrent Suite (v4.4) were used to perform base calling, trimming and filtering, and alignment to hg19. A good quality run was ensured, with higher than 85% ISP loading density and approximately 6 million reads. Approximately 1.365 GB of data (6,184,354 reads) were generated for cfDNA.

### Sequence-analysis pipeline

Binary alignment matrix (BAM) files, containing raw sequencing reads, were used to perform a strict de novo assembly in the CLC genomics workbench (Version 7.0), followed by read-mapping. Selected contigs were converted into a FASTA format, after which the files were exported and converted to text-file format for further analysis. To screen for REs and regions of low complexity, we used RepeatMasker, a program utilizing RepBase (a service of the Institute for Systems Biology). Hereafter, when applicable, three consecutive local alignment analyses were conducted using the BLAST (Basic Local Alignment Search Tool) search engine to compare the generated array of cfDNA sequences with known sequences, including the human genome. The first search was against the ENSEMBL human GRCh38.p5 genome database. For the queries that did not return any hits, a second search against the National Center for Biotechnology Information (NCBI) nucleotide (nr/nt) database was performed using the megablast algorithm. Then again, for the queries that returned no results, a third search was performed using the blastn algorithm. In all cases, default search parameters were used. This procedure was repeated for the hits that covered less than 50% of the query sequence length in the ENSEMBL BLAST search. For each sequence, the highest scoring BLAST hit, its physical location, the overlapping gene (where relevant), and genomic location (e.g. in the centromere, in telomeres, in one gene) were recorded. This information was used to categorize the cfDNA and genomic DNA sequences.

## Results

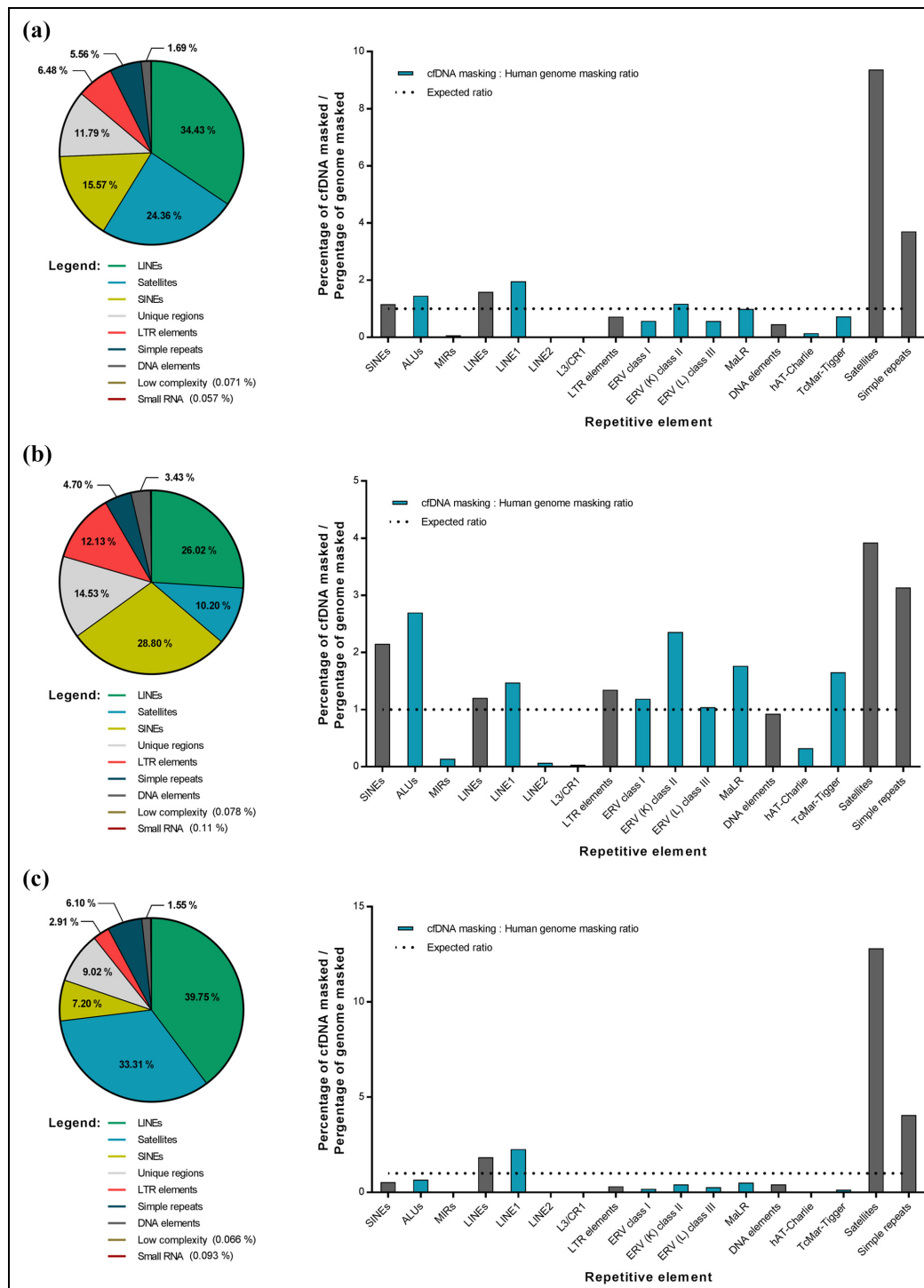
### *The spread of cfDNA sequencing data*

To investigate the composition of the DNA released by 143B cells after 24 h of incubation, the cfDNA present in the culture medium was isolated and sequenced. First, the distribution of cfDNA sequence coverage was evaluated. This showed that the data are skewed significantly to the right, indicating the presence of at least two distinct cfDNA populations, that is, a large number of sequences with a relatively low coverage and a small number of sequences with a very high coverage. The details of these analyses are described in Supplementary file 1, Section 1.

### *Masking and representation of REs*

To screen for REs and regions of low complexity, we used RepeatMasker Open software (4.0). To characterize the REs of the entire cfDNA population and to account for a non-normal distribution (as indicated in Supplementary Fig. 1), three data subsets were investigated, namely, (a) all contigs with coverage greater than 20, (b) contigs with coverage between 20 and 100, and (c) contigs with coverage greater than 100. The details of this method are described in Supplementary file 1, Section 2. In all subsets investigated, only a very small portion of the cfDNA population consists of unique regions, while the amount of REs notably exceeds any value predicted for the human genome<sup>26–28</sup> (Figure 1). Therefore, the DNA released by 143B cells after 24 h of incubation is comprised mainly of REs. Furthermore, when considering the entire cfDNA population (Subset a), long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs), satellites, and simple repeats (mini satellites) make up the majority of the cfDNA population and are overrepresented compared to the human genome (Figure 1(a)). Very interestingly, satellites and mini satellites are significantly overrepresented, while long terminal repeat (LTR) elements and DNA elements are underrepresented. As shown in Subset c (Figure 1(c)), which depicts only the contigs with a coverage greater than 100, satellites, mini satellites, and LINEs are significantly overrepresented in the cfDNA population as a result of a small number of sequences that have a very high coverage. Therefore, when these contigs (which significantly skew the data) are taken out of consideration (Subset b; Figure 1(b)), it becomes clear that, regardless of the overall masking of each repeat class, specific elements in each class are significantly overrepresented, while others are significantly underrepresented or occur at levels comparable to the human genome.

As a control, genomic DNA isolated from 143B cells was sequenced, screened for REs, and compared to 143B cfDNA (Supplementary file 1, Section 2.1). 143B



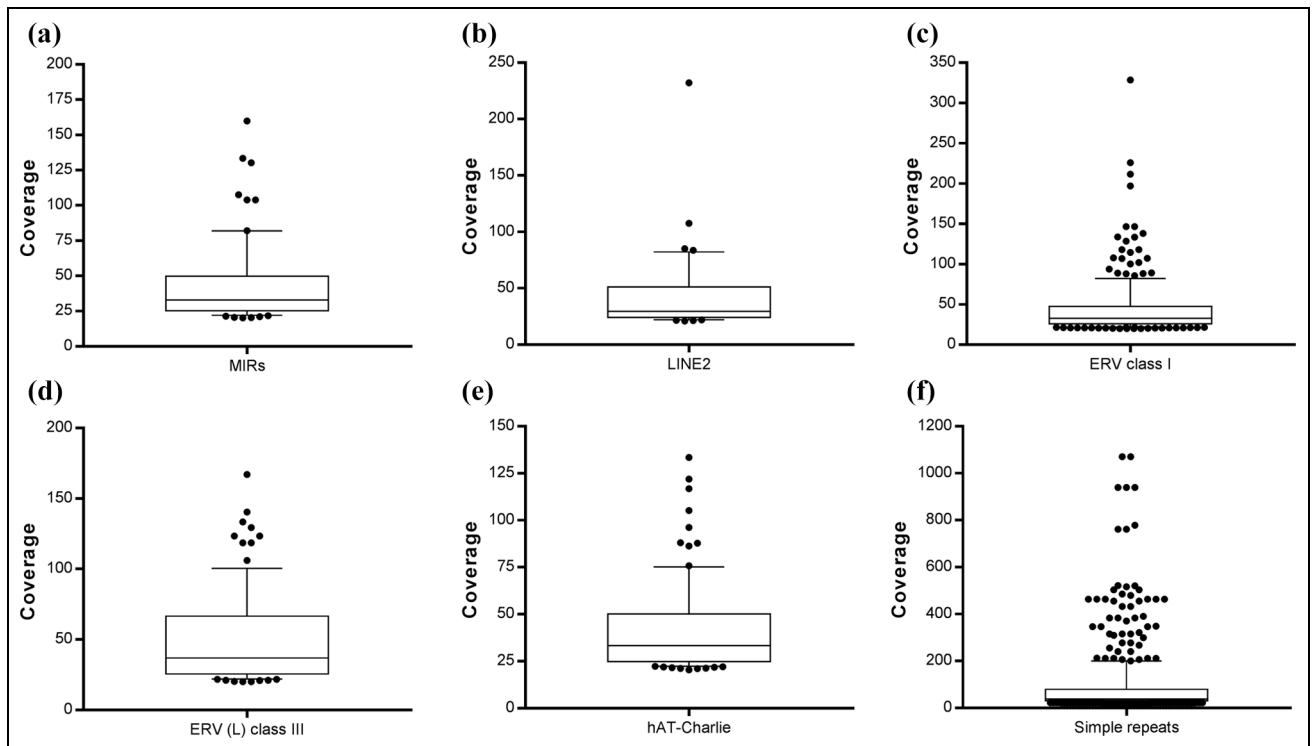
**Figure 1.** Representation of repetitive elements in cell-free DNA. Representation of the repetitive elements in each of the cfDNA subsets that were analyzed (as described in the “Masking and representation of repetitive elements” section), including (a) all contigs with coverage greater than 20, (b) contigs with coverage between 20 and 100, and (c) contigs with coverage greater than 100, respectively. Pie graphs illustrate the representation of the different repetitive element classes in the cfDNA population of each subset. Each repeat class is denoted by a different color, as indicated by the legend. In addition, each chart shows the percentage of the cfDNA population that does not contain repetitive elements (denoted by “unique regions”). Values were obtained by dividing the total number of bases masked by each repetitive element by the number of bases that constitute the entire corresponding subset. Bar graphs illustrate the representation of each repeat class, as well as the repetitive element types in each class, expressed as observed/expected ratios (i.e. the percentage of cfDNA masked by each repetitive element was divided by the percentage of the human genome masked by the corresponding element). Gray bars illustrate the different repeat classes, while blue bars indicate the repetitive element types contained within the preceding class. The dashed line denotes the expected ratio of masking for all repeats. Thus, values that extend above this line indicate an overrepresentation of the repetitive element, while values below this line indicate an underrepresentation of the repetitive element.

genomic DNA shows almost normal levels of REs (~59% of sequences are masked) and normal levels of non-repetitive DNA (gene sequences). In contrast, it is clear that REs are significantly overrepresented in cfDNA (~82% of sequences are masked) while cfDNA contains virtually no unique gene sequences (Supplementary Fig. 2). This suggests that cfDNA does not represent merely fragmented genomic DNA but is actively released from specific parts of the genome. This is further discussed in the “Possible pathways for the extrusion of cfDNA” section.

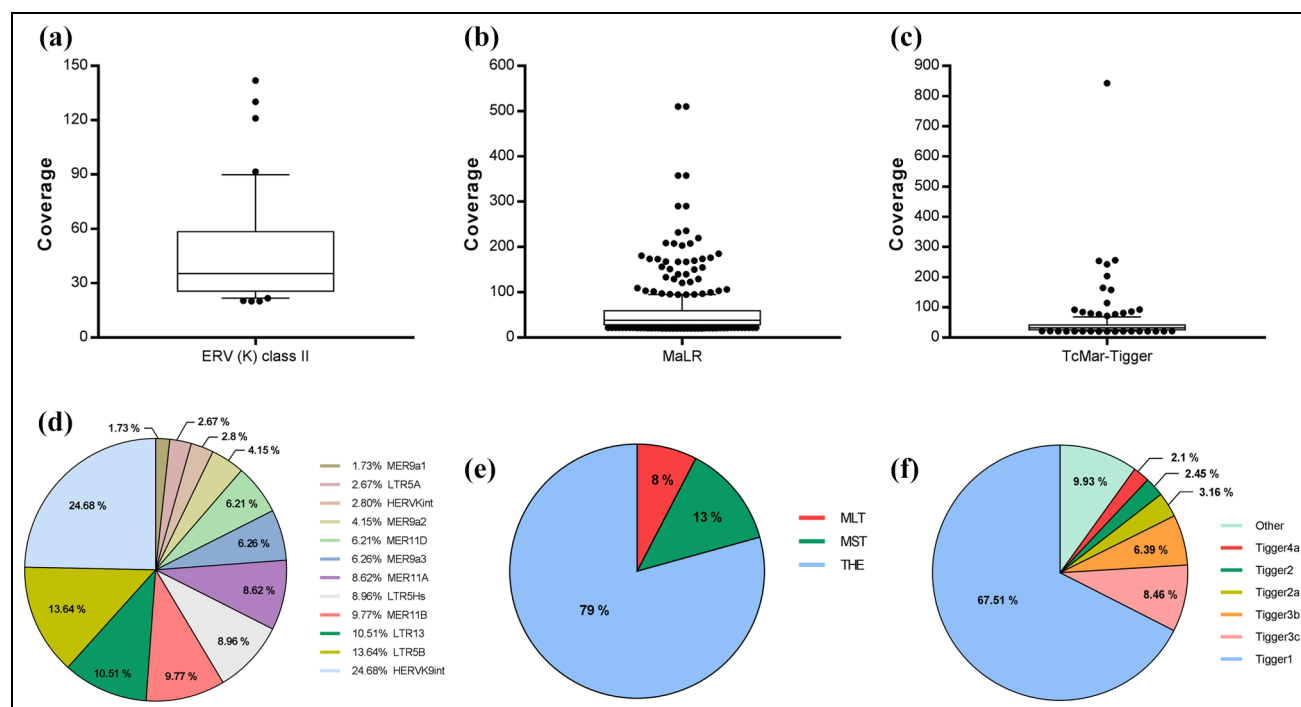
### Evaluation of individual repeat classes and associated subfamilies

Since the different RE families are not equally represented in the human genome, the coverage distribution of the contigs that comprise each RE type/family was evaluated. The purpose of this was to identify sequences that are significantly overrepresented within the context of each group. To do this, the coverage distribution of the contigs that comprise each RE population was displayed as a box plot, where data points above the 90th percentile are considered to be significantly overrepresented. Figure 2 illustrates the RE populations that were not overrepresented, including mammalian-wide

interspersed repeat (MIR; Figure 2(a)), L2 (Figure 2(b)) elements, endogenous retrovirus (ERV) class I (Figure 2(c)), ERV (L) class II (Figure 2(d)), and hAT-Charlie (Figure 2(e)). Although mini satellites are also overrepresented, the sequences of this population are too diverse to be grouped and displayed, and were therefore included in (Figure 2(f)) and not in Figure 3. Figure 3 illustrates the RE populations that were overrepresented, including ERV (K) class II (Figure 3(a)), mammalian non-LTR retrotransposons (MaLR; Figure 3(b)), and TcMar-Tigger (Figure 3(c)). Box plots could not satisfactorily illustrate the spread of the data/identify outliers for *Alu*, L1, and satellites. Alternatively, these data were displayed as scatterplots, wherein 5% of the contigs with the highest coverage (which are considered to be significantly overrepresented) were distinguished from the remaining 95% of the data (Figure 4). Significantly overrepresented elements are summarized in Supplementary file 2. This includes the contig ID, the matching repeat, the length that it masked, the coverage of the sequence with which it aligned, as well as the total bases masked. In addition, the FASTA sequences of each of these overrepresented elements are indexed in Supplementary file 3 and can be located by its contig ID. To evaluate the composition of each RE type, the representation of its corresponding subfamily was



**Figure 2.** Identification of significantly overrepresented repetitive elements—Part I. Coverage distribution of the contigs that constitute each of the different repetitive element populations, including (a) MIRs, (b) LINE2, (c) ERV class I, (d) ERV (L) class III, (e) hAT-Charlie, and (f) simple repeats. Contigs with coverage values above the 90th percentile are significantly overrepresented. Significantly overrepresented elements are summarized in Supplementary file 2.



**Figure 3.** Identification of significantly overrepresented repetitive elements—Part II. Coverage distribution of the contigs that constitute the repetitive element populations that are significantly overrepresented (see Figure 1), including (a) ERV (K) class II, (b) MaLR, and (c) TcMar-Tigger. Contigs with coverage above the 90th percentile are considered significantly overrepresented (significantly overrepresented repetitive elements are summarized in Supplementary file 2). Furthermore, pie charts illustrate the representation of the subfamilies that comprise each of the aforementioned repetitive element populations, including (d) ERV (K) class II, (e) MaLR, and (f) TcMar-Tigger. Although simple repeats are also overrepresented, they were not included in this figure because the sequences of this population are tremendously diverse.

determined. This was done only for the RE types that were shown to be overrepresented in comparison with the human genome (see Figure 1), namely, ERV (K) class II (Figure 3(d)), MaLR (Figure 3(e)), TcMar-Tigger (Figure 3(f)), *Alus* (Figure 4(a)), L1 (Figure 4(b)), and satellites (Figure 4(c)). Furthermore, regarding the three most abundant REs in the cfDNA population, *Alus*, L1, and satellites, the representation of their corresponding subfamilies within the top 5% and bottom 95% of the data was determined. This shows that there are individual subfamilies that contain contigs with a significantly high coverage.

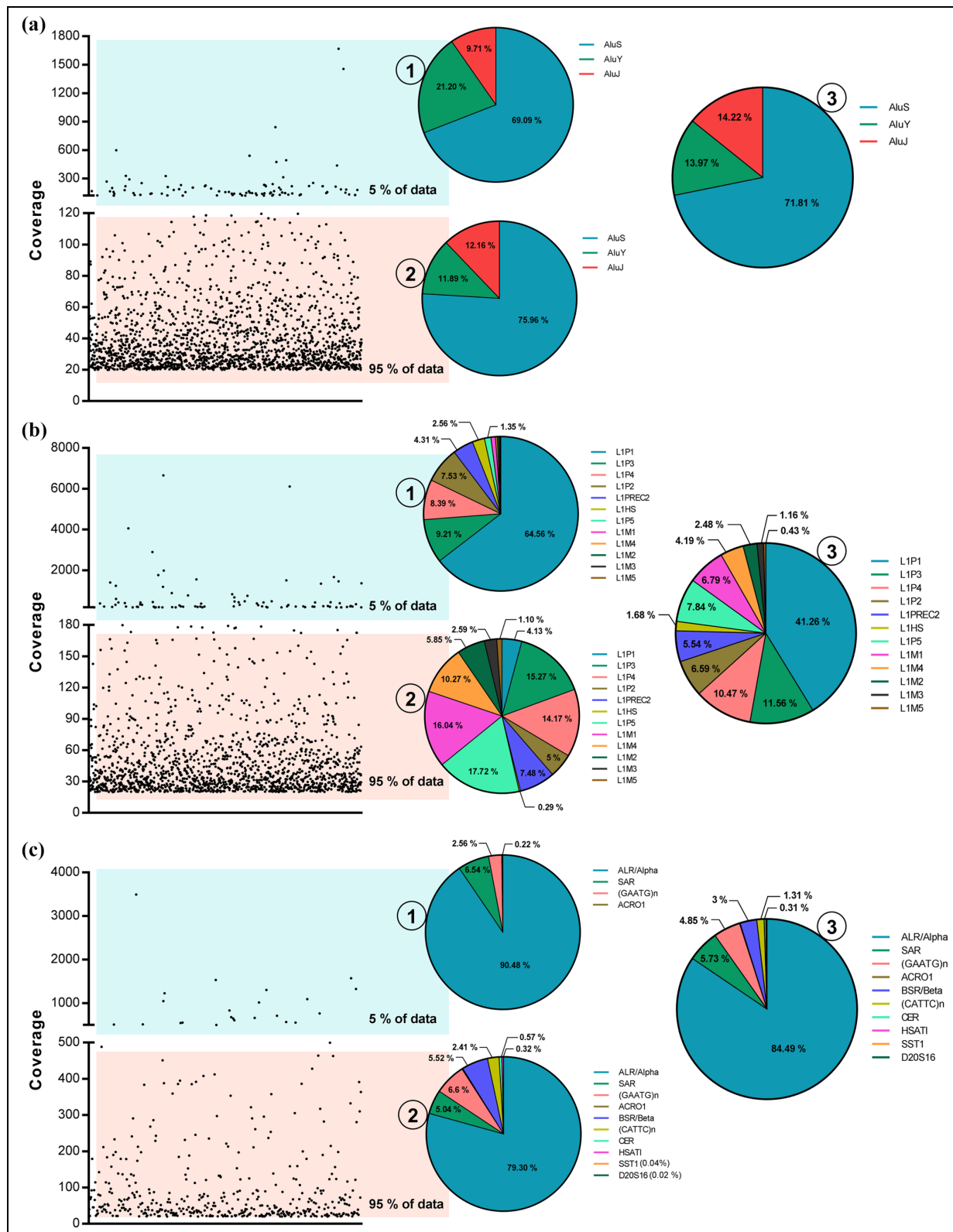
### Chromosomal distribution

To further investigate the origin of the cfDNA sequences, chromosomal distribution was evaluated for each of the different RE populations that were shown to be significantly overrepresented (Figure 5). The number of bases of each RE type on each chromosome of the human genome differs notably. Therefore, to normalize the data, the total number of bases for each RE type on each chromosome (the number of bases multiplied by coverage) was divided by the number of

bases that the corresponding RE type occupies on each chromosome of the human genome. The number of bases that each RE type occupies on each chromosome of the human genome was calculated by subjecting the pre-masked human genome to repeat masking analysis. Furthermore, the chromosomal distribution of 143B cfDNA was compared with the chromosomal distribution of 143B genomic DNA (Figure 5). Interestingly, specific chromosomes were significantly overrepresented, especially chromosomes 1 and 9, while the remaining chromosomes were significantly underrepresented. Moreover, the chromosomal distribution patterns of the cfDNA sequences very closely mirror those of 143B genomic DNA sequences. This observation is further discussed in the “Possible pathways for the extrusion of cfDNA” section.

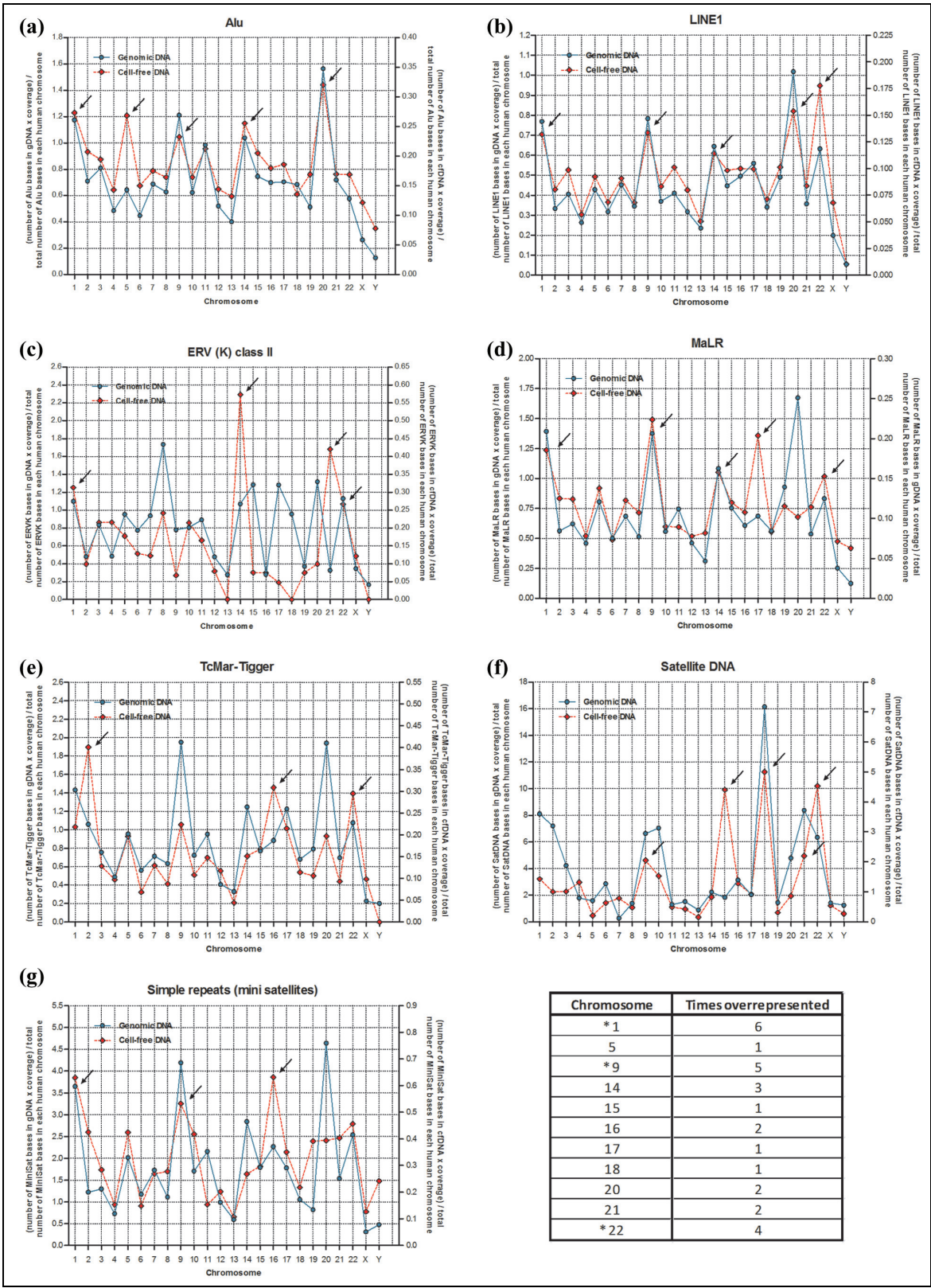
### Additional analyses and supplementary material

Additional analyses that have been performed on the cfDNA sequences are described in detail in supplementary material. This includes (1) local alignment analyses and annotation of unique (unmasked) cfDNA sequences (Supplementary file 1, Section 3 and



**Figure 4.** Identification of significantly overrepresented repetitive elements—Part III. Scatterplots illustrating the coverage distribution of the contigs that constitute each of the repetitive element populations of which significantly overrepresented elements could not be distinguished using box plots, including (a) Alus, (b) LINE1, and (c) satellites. Each dot represents a single contig. The superimposed blue box indicates the contigs that constitute the top 5% of the data when sorted according to increasing coverage, while the pink superimposed box indicates contigs that constitute the bottom 95% of the data. Dots within the blue box, therefore, indicate significantly overrepresented contigs. Significantly overrepresented elements are summarized in Supplementary file 2. Pie charts illustrate the representation of the subfamilies that constitute each of the aforementioned repetitive element populations, including ① the top 5% of the data, ② the bottom 95% of the data, and ③ the entire repeat element population.





**Figure 5.** Chromosomal distribution of cell-free DNA. Chromosomal distribution of significantly overrepresented repetitive element groups (see Figure 1), including (a) Alu, (b) LINE1, (c), ERV (K) class II, (d) MaLR, (e) TcMar-Tigger, (f), satellite DNA, and (g) simple repeats. The chromosomal distribution of these elements in 143B cDNA is plotted on the right y-axis (red) and is compared with 143B genomic DNA, plotted on the left y-axis (blue). Significantly overrepresented chromosomes are indicated by black arrows. Normalized values were obtained by dividing the total number of bases determined for both 143B cDNA and 143B genomic DNA of each repetitive element group on each chromosome by the actual number of bases that each of the different repetitive element groups occupies on each chromosome of the human genome (hg.19).



Supplementary Fig. 3), (2) the assessment of potential sequencing bias and procedural errors (Supplementary file 1, Section 4 and Supplementary Fig. 4), and (3) identification of sequences that did not align with the human genome (Supplementary file 1, Section 5 and Supplementary file 4). All supplementary figures are included in Supplementary file 1, Section 7.

## Discussion

In a previous study, we have shown that the cfDNA present in the growth medium of cultured 143B cells after 24 h of incubation is mainly a product of an active release mechanism and not a consequence of apoptosis or necrosis.<sup>25</sup> The objective of this study was to investigate the origin and purpose of this specific population of cfDNA by means of sequence analysis. After sequencing, cfDNA sequences were re-assembled and then screened for REs, followed by local alignment analyses and annotation. Initial RE screening results showed that most of the cfDNA consists of repetitive DNA (88%; Figure 1), which exceeds any value predicted for the human genome (typical estimates range between 50% and 66%).<sup>26–28</sup> Further analysis of the RE screening data showed that specific RE types are overrepresented (i.e. *Alus*, L1, ERV (K) class II, MaLR, TcMar-Tigger, satellites, and simple repeats (mini satellites)), while others are underrepresented or occur at levels comparable to the human genome (Figure 1). Moreover, it was demonstrated that different subfamilies within each RE type are significantly overrepresented (Figures 2–4). Local alignment analysis of the overrepresented RE types showed that they originate from specific chromosomes, especially 1 and 9 (Figure 5). Interestingly, while 143B genomic DNA show normal levels of REs, some chromosomes, especially 1 and 9, appear to harbor extra copies of satellites, mini satellites, and some other REs, mirroring the cfDNA chromosomal distribution profile. This may be the result of a specific subtype of genomic instability (see “Possible pathways for the extrusion of cfDNA” section). In the following sections, we discuss possible reasons why satellites and each of the different REs is over- or underrepresented, and why they originate mainly from specific chromosomes. We also discuss the potential implications of these findings.

### The representation of class I transposons correlates with transposition activity

**SINEs.** *Alu* elements are notably overrepresented (Figure 1(b)). However, when the different *Alu* subfamilies are compared, it is clear that the cfDNA population contains nearly 10 times more elements originating collectively from the *AluS* and *AluY* subfamilies than

from the *AluJ* family (Figure 4(a)). Interestingly, the former subfamilies harbor more functionally intact elements and are, therefore, more likely to transpose in the human genome, whereas the *AluJ* lineage is the least active and is considered to be functionally extinct.<sup>29</sup> Like *AluJ*, MIR elements are also underrepresented in the cfDNA population (Figure 1(b)) and no longer possess any transposition activity in the human genome.<sup>30</sup> In vitro studies indicate that *Alu* elements are capable of retrotransposing in somewhat differentiated cells,<sup>31,32</sup> which suggests the possibility of activity in somatic tissues.

**LINES.** The underrepresentation of L2 and L3/CR1 also correlates with its inactivity,<sup>33</sup> while the overrepresentation of L1 elements correlates with its active transposition status. Although there are more than 500,000 L1 copies in the human genome, most are considered to be molecular fossils.<sup>34</sup> It has been estimated that an average human genome contains only 80 to 100 retrotransposition-competent copies, which all belong to the human-specific L1 (L1HS) subfamily. L1HS elements are then further stratified into the L1-Ta subfamily (pre-Ta, Ta-0, Ta-1, Ta1-d, Ta1-nd),<sup>35</sup> which is estimated to account for ~31.5% of all L1HS elements,<sup>36</sup> and the older L1P1 subfamily (L1PA2 and L1PA3) that comprises the remainder of the L1HS subfamily.<sup>37</sup> As illustrated in Figure 4(b), the L1P1 subfamily constitutes nearly half of the L1 population. Therefore, it is possible that the representation of the different L1 elements may be related to its transposition activity. Furthermore, there is a relatively small number of these L1P1 elements that have a very high coverage (Figure 4(b)), which may relate to a regulated process, such as gene amplification (see “Possible pathways for the extrusion of cfDNA” section).

**LTRs.** To date, 22 different human endogenous retrovirus (HERV) families have been identified; however, most of these have been rendered inert by the accumulation of mutations and incomplete sequences.<sup>38</sup> Those that do possess some activity are, in the main part, transcriptionally repressed by methylation.<sup>39</sup> However, a number of reports highlight cases in which specific ERVs can become transcriptionally active and even transpose. For example, in some cancers (e.g. bladder, ovarian, and testicular), the degree of HERV demethylation has been found to increase concomitantly with malignancy and is often accompanied by transcript expression.<sup>40,41</sup> Furthermore, the deletion of DNA methyltransferase 3-like (*Dnmt3L*), a gene involved in the de novo methylation of retrotransposons in the germ cells of adult male mice, prevents the methylation of both LTR and non-LTR retrotransposons, which results in the reactivation of transposition activity.<sup>42</sup>

Other researchers have achieved similar effects by the deletion of lymphoid-specific helicase (Lsh or HELLS).<sup>43</sup>

### ***The representation of class II transposons correlates with evolutionary age***

As shown in Figure 3(f), TcMar-Tigger elements are overrepresented in the cfDNA population, while hAT-Charlie elements are underrepresented. Most DNA transposons are thought to be fixed within the human genome due to internal deletions and/or end-truncations,<sup>44</sup> and there is currently no evidence that they are currently active.<sup>45</sup> However, we can suggest one plausible explanation for the differential representation of these elements in the cfDNA. The TcMar-Tigger subfamily is comprised of several more subfamilies (Tigger1, Tigger2, Tigger2a, Tigger3, etc.), where the numerical suffix correlates with the age of the element and refers to the point in time when it diverged structurally from older subfamilies. This divergence then correlates with transposition activity.<sup>46</sup> In other words, the youngest elements are those that have transposed the most recently. Thus, when the different TcMar-Tigger subfamilies were further analyzed, it was found that two-thirds consist of Tigger1 (Figure 3(f)), the youngest member of the TcMar-Tigger family. In addition, the representation of the remaining Tigger elements also correlates with age.<sup>46</sup> Therefore, the representation of DNA elements appears to correlate with its capacity to transpose.

Although there is currently no evidence for the movement of DNA transposons in the human genome, there is a possibility that it has been occurring unnoticed. In Supplementary file 1, Section 6, we summarize some points to illustrate that the movement of DNA transposons in the human genome, within and between cells, is possible.

Taken together, the literature discussed above indicates that the cfDNA present in the growth medium of 143B cells after 24 h of incubation is composed primarily of satellites, mini satellites, and transposons that exhibit the capacity to become activated and transpose (i.e. *AluS*, *AluY*, *L1P1*, *THE*, and Tigger1). This raises the following important questions: why is the cfDNA enriched in satellites, mini satellites, and potentially active transposons? And why do they originate only from specific chromosomes?

### ***The association between hypomethylation of repetitive DNA and cancer***

The cancer genome is often characterized by a decrease in global 5-methylcytosine levels.<sup>47</sup> Since single-copy genes account for less than 5% of human DNA,<sup>26</sup> this

large decrease of methylcytosine content mainly reflects methylation changes of CpG-rich non-coding DNA, in particular peri- and juxtacentromeric sat2 and sat3 (classical satellites), *Alu*, *L1*, and LTR-containing repeats (especially HERVs), but also includes the less CpG-rich centromeric  $\alpha$ -satellite DNA.<sup>48</sup> At the basis of their q-arms, chromosomes 1, 9, and 16 have large satellite-repeat arrays. In healthy adult somatic cells, these chromosomal regions are normally densely methylated;<sup>49</sup> however, they are preferentially hypomethylated in many cancers, such as ovarian epithelial tumors,<sup>50</sup> breast adenocarcinomas,<sup>49</sup> and Wilms tumors.<sup>51</sup> This same phenomenon occurs in Immunodeficiency, Centromeric instability, and Facial anomalies (ICF) syndrome. ICF is a very rare autosomal recessive disorder characterized by variable immunodeficiency, mild facial anomalies, and centromeric decondensation.<sup>52,53</sup> The most consistent molecular observation of this disease is the hypomethylation of the centromeric and pericentromeric regions of chromosomes 1, 9, and 16, while global DNA hypomethylation patterns present within the normal range.<sup>53–55</sup> Typically, hypomethylation of these specific chromosomal loci in lymphocytes from ICF patients are accompanied by significant loss of organization, increased instability, followed by cytogenetic abnormalities, including (1) extension of the juxtacentromeric heterochromatin of metaphase chromosomes 1, 9, and 16; (2) formation of complex multiradiate chromosomes with multiple p- and q-arms derived from chromosomes 1 and 16; (3) breakage and rejoining of Sat2 resulting in the duplication of q-arms on chromosome 1; and (4) extrusion of self-associated Sat2 into secondary nuclear structures called micronuclei or nuclear buds.<sup>56</sup> Similarly, many cancers exhibit a very high frequency of non-random rearrangements in the centromeric and pericentromeric heterochromatin of chromosomes 1, 9, and 16,<sup>57</sup> which constitute chromosomal aberrations that are identical to those observed in ICF syndrome.<sup>56</sup>

Further evidence that the centromeric and pericentromeric regions of chromosomes 1, 9, and 16 are predisposed to hypomethylation and chromosomal instability comes from numerous in vitro studies in which the genotoxic effects of various clastogens, in particular base analogs, have been evaluated. In these studies, administration of various compounds such as 5-Azacytidine, 5-azadeoxycytidine, 2,6-diaminopurine, idoxuridine, and mitomycin C to cells resulted in increased hypomethylation of the centromeric and pericentromeric regions mainly of chromosomes 1, 9, and 16. Moreover, this hypomethylation was usually accompanied by various chromosomal abnormalities, including heterochromatic decondensation, DNA breaks, and inclusion of DNA fragments into micronuclei (as observed in ICF and cancer).<sup>58–62</sup> Perhaps it is also worthy to mention that the same chromosomal

abnormalities arise in cells of which the centromeric and pericentromeric regions of chromosomes 1, 9, and 16 become spontaneously hypomethylated over time.<sup>63</sup>

Therefore, we propose the hypothesis that the cfDNA present in the growth medium of 143B cells after 24 h of incubation is associated with the hypomethylation, and consequently the instability, of the centromeric and pericentromeric regions mainly of chromosomes 1 and 9 (some other chromosomes may also be involved to a lesser extent). This raises the next important question: why and how does the hypomethylation of these regions result in the presence of these specific cfDNA fragments in the extracellular environment? A detailed discussion of the mechanisms that are likely involved in this phenomenon is beyond the scope of this article and is the subject of a prospective review paper. Here, we will briefly outline possible explanations for this phenomenon.

### *Possible pathways for the extrusion of cfDNA*

The structural fidelity of the centromere/kinetochore is crucial for the proper segregation of chromosomes during mitosis.<sup>64</sup> As discussed above, hypomethylation of the centromeric and pericentromeric regions of chromosomes 1 and 9 alters the structure of centromeric DNA, and, therefore, directly affects the formation of kinetochores on centromeres, consequently altering its interaction with microtubules. When these components malfunction, various microtubule attachment errors arise.<sup>65,66</sup> Here, an important example is merotelic attachments, in which one kinetochore binds to microtubules growing from opposite spindle poles. Although most merotelic chromosomes segregate correctly during anaphase (i.e. move toward the poles), a small fraction may remain at the spindle equator, yielding lagging-chromosomes.<sup>65</sup> Depending on where the cleavage furrow ingresses during cytokinesis, the lagging chromosome can be included in either of the daughter cells. Interestingly, when a lagging chromosome is far enough from the remaining chromatin mass, it recruits its own nuclear envelope at the end of mitosis, forming a micronucleus. Consequentially, an interphase daughter cell contains two types of nuclei: the primary nucleus, which encapsulates the main chromatin mass, and an indefinite number of micronuclei that contain the mis-segregated chromosomes. While chromosome lagging and micronucleation can occur at low levels in normal cells, increased levels of micronuclei have been found in various pathologies, especially in cancer cells with chromosomal instability.<sup>67–70</sup>

Micronuclei can also arise during mitosis as a result of mis-repaired double-stranded DNA breaks (DSBs).<sup>71</sup> When a chromosome that has lost its telomere, due to a DSB, is replicated the two ends of the sisters fuse, producing a chromosome with two

centromeres (dicentric chromosome).<sup>72,73</sup> During anaphase, the two centromeres are pulled to opposite poles, generating an anaphase bridge.<sup>74–76</sup> When an anaphase bridge breaks irregularly and in a gene region, which is very often the case, the broken chromosomes lose their telomere sequences at the broken end and usually fuse with their copy after replication.<sup>77–80</sup> When the centromeres of the fused sister chromatids are pulled apart during the next anaphase, the resulting bridge breaks again. Following the next round of replication, the sister chromatids fuse once again, perpetuating the breakage-fusion-bridge (BFB) cycle through several rounds of mitosis, thus amplifying the DNA located near the break- or fusion point.<sup>81,82</sup> This process usually proceeds until the marker chromosome acquires a new telomere. This may explain why certain REs are overrepresented in 143B genomic DNA. Interestingly, these amplified genes are eventually eliminated (looped out) from the aberrant chromosome when homotypic sequences within the amplified genes recombine to form mini-circles of acentric and atelomeric DNA (double minutes (DMs)). These DMs can either be replicated or localize to the nuclear periphery where they are eliminated from the nucleus via budding during S-phase, after which they can be extruded from the cell.<sup>80,83</sup> Interestingly, these DMs most frequently consist of amplified copies of oncogenes, which may explain why specific transposons are significantly overrepresented in the cfDNA population. As this process can reduce the dosage of tumor suppressor genes,<sup>84–86</sup> it may play a key role in the acquired resistance of cancer cells to different therapies and, therefore, constitute a step in tumorigenesis and propagation of cancer. Alternatively, the generation of DMs and its extrusion from cells may serve as a pathway for eliminating amplified oncogenes, resulting in the loss of the malignant phenotype.<sup>71,87</sup> These nuclear buds generally have the same morphology as micronuclei, except that they are physically connected to the nucleus by a stalk of nucleoplasmic material at some stage of the budding process.<sup>77</sup>

A number of studies have reported the occurrence of BFB intermediates in different osteosarcoma cell lines.<sup>88–91</sup> By quantifying anaphase bridge configurations and dicentric chromosomes in four different osteosarcoma cell lines, it was demonstrated that BFB cycle events occur at a high enough frequency to serve as a mechanism for increasing the de novo formation of cytogenetic abnormalities in osteosarcoma. In the same study, an increase in the occurrence of micronuclei, chromatin strings, and nuclear blebs were observed compared to controls, which represents ruptured anaphase bridges.<sup>92,93</sup> Another significant finding of this study is the observation of interstitial alpha-satellites in dicentric chromosomes that are preferentially located at chromosome ends. There is evidence that the alpha-

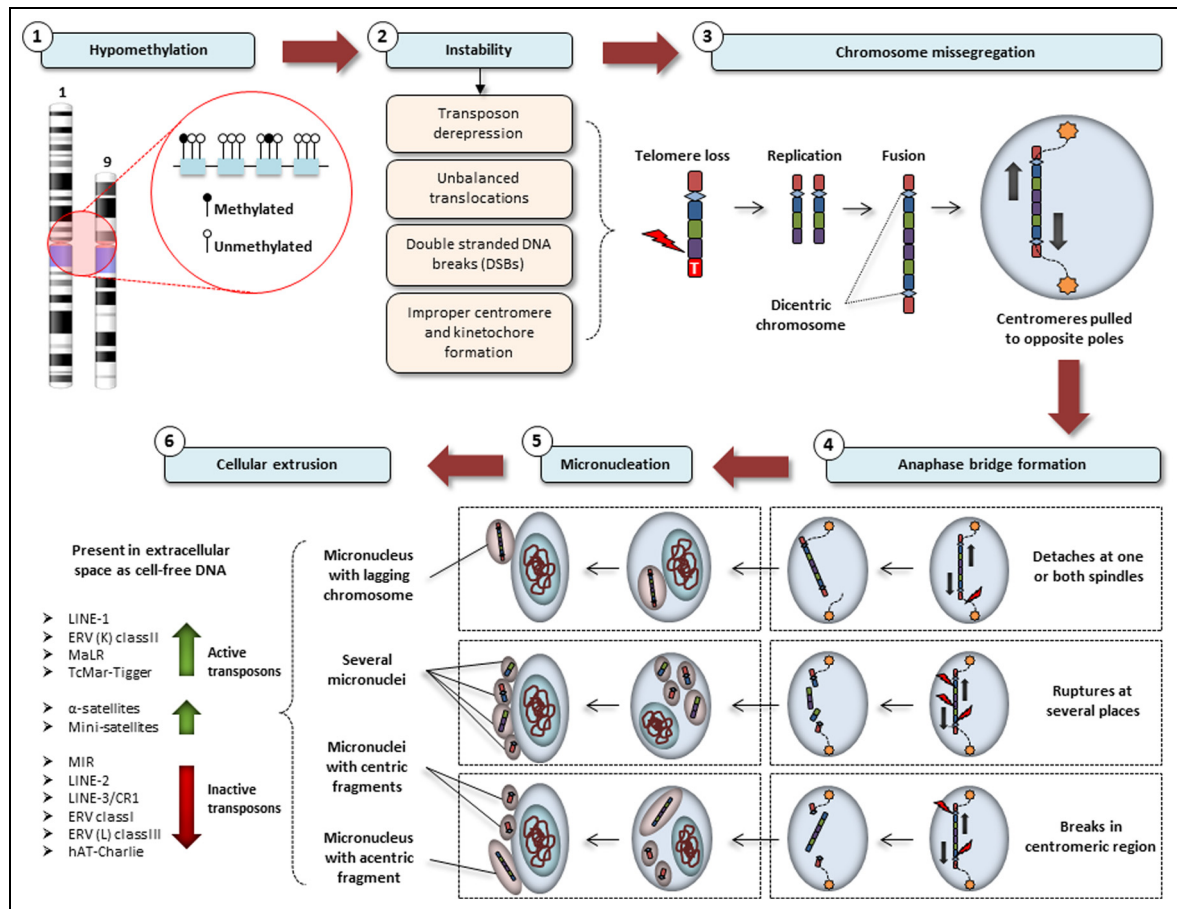
satellites found in the centromeric and pericentromeric regions of human chromosomes represent hotspots for translocations and recombination in some solid tumors.<sup>94,95</sup> Since these sequences have been observed to be specifically amplified in osteosarcoma,<sup>96</sup> this may explain why rearrangements often occur in the pericentromeric regions in osteosarcoma tumors and cell lines.<sup>97</sup> Moreover, in the SAOS-2 osteosarcoma cell line, chromosome 1 exhibits various complex rearrangements, including one rearrangement in which DNA from 1p35-p36 is duplicated and inverted in a der(9)t(1:9).<sup>92</sup> If some micronuclei are derived from ruptured anaphase bridges in which DNA fragmentation was initiated adjacent to the centromeres of these two chromosomes, it may explain why the cfDNA released by 143B cells is overrepresented in DNA originating from the centromeric and pericentromeric regions of chromosomes 1 and 9.

As discussed earlier, transposons are normally repressed by methylation. However, in cancer cells, hypomethylation can result in the reactivation of certain transposons, enabling transcription and transposition. There is evidence that transposons can generate DNA breaks. When replication forks traverse a DNA strand with a single-stranded nick, the fork can collapse and generate a DSB.<sup>98,99</sup> In addition, it has been demonstrated that the nuclease produced by L1 can also generate DSBs.<sup>100</sup> As mentioned in the "LINEs" section, an average human genome contains about 80 to 100 L1 copies that are retrotransposition competent (or contain intact ORFs). By cloning 82 of these L1 copies, it was demonstrated that 40 of them are active in a 143B thymidine kinase negative (TK) TK cultured cell retrotransposition assay.<sup>101</sup> Of these 40 active L1s, 6 hot L1s accounted for 84% of the total activity of all 82 L1s tested. Using sequence divergence and allele frequency of each L1 as surrogate markers for age, it was shown that younger L1s have high activity in cultured cells, while older L1s displayed virtually no activity. As discussed in the "LINEs" section, cfDNA released by 143B cells is overrepresented in young L1 elements. Moreover, only four elements are significantly overrepresented, which may represent hot L1s.

As another example, there is extensive sequence similarity between centromere-protein B (CENP-B), a protein associated with the centromeres of most human chromosomes, and the transposase encoded by the human Tigger DNA transposon. In addition, the terminal inverted repeats (TIRs) of the Tigger2 elements contain a near perfect match to the CENP-B binding site (or CENP-B box),<sup>102</sup> which consist of a 17-bp alpha satDNA motif.<sup>103</sup> These structural similarities suggest two intriguing possibilities. First, CENP-B, akin to transposases, may also possess the ability to cause single-stranded breaks (SSBs).<sup>102</sup> Second, transposases could induce SSBs adjacent to the CENP-B

box. There is strong evidence that both of these phenomena occur in higher organisms.<sup>104–106</sup> It is also possible that larger DNA fragments could be excised by this mechanism.<sup>107–110</sup> Thus, the generation of DSBs by transposons may play a role in the instability of the centromeric and pericentromeric regions of specific chromosomes, especially chromosomes 1 and 9, and result in the formation of dicentric chromosomes, anaphase bridges, and eventually the formation and extrusion of micronuclei that harbor satellites, mini satellites, and transposons that are enriched in the specific region.

An important question is whether the observations made *in vitro* are also reflective of the *in vivo* setting, that is, whether micronuclei are actually released into the bloodstream. As far as we know, there have been no attempts to isolate micronuclei directly from human blood. However, there is some indirect evidence that micronuclei, or at least the contents of micronuclei, are present in the extracellular environment: DNA fragments with a size of ~2000 bp, such as those investigated in this study, and which are also present in the growth medium of several cultured cell lines,<sup>25,111</sup> have been detected in human blood. However, it is readily dismissed as mere cellular DNA contamination (refer to application note<sup>112</sup>). Therefore, *in vitro* studies provide compelling evidence that this may be an erroneous assumption. Interestingly, it has recently been demonstrated that automated cfDNA isolation systems (e.g. KingFisher), which are the most commonly used in clinical assays, are tailored for short fragments and typically fail to isolate DNA fragments in the range of ~2000 bp (refer to aforementioned application note<sup>112</sup>). Since the majority of cfDNA research is clinically motivated (and focused on short cfDNA fragments), it may explain why the larger ~2000 bp cfDNA population is not often encountered. Furthermore, a recent paper reports for the first time the presence of extrachromosomal circular DNA (whose cellular extrusion is facilitated by inclusion into DMs) in human blood.<sup>113</sup> This finding has been corroborated by a different research group, who have demonstrated the presence of a heterogeneous population of extrachromosomal circular DNA, with a size range between 30 and 20,000 bp, in human blood.<sup>114</sup> Since these DNA fragments are larger than apoptotic DNA fragments, it is unlikely that this pool of cfDNA molecules have yet been investigated for clinical purposes. In this regard, the commonly held assumption that apoptosis is the main origin and most relevant fraction of cfDNA in human blood may be incorrect, restrictive, and should be reconsidered. It may be a major breakthrough in the field of clinical diagnostics if the DNA actively released by cultured cells bears any resemblance to the DNA originating from its *in vivo* counterpart. Therefore, this warrants the implementation of comparative studies. Further



**Figure 6.** A provisional hypothesis for the origin of actively released cell-free DNA in cancer. In normal cells, the repetitive elements in the centromeric and pericentromeric regions of chromosomes 1 and 9 are normally densely methylated. ① In cancer cells these regions are predisposed to increased hypomethylation. ② This can result in the derepression of transposons, followed by aberrant translocations and the induction of DSBs. Alternatively, hypomethylation can impair the formation of proper kinetochores and centromeres. ③ When a chromosome that has lost its telomere is replicated the two ends of the sisters fuse, producing a dicentric chromosome. ④ During anaphase, the two centromeres are pulled to opposite poles, generating an anaphase bridge. Anaphase bridges can break in different regions, generating DNA fragments of various sizes and content. ⑤ These DNA fragments can recruit their own nuclear envelope, forming secondary nuclear structures known as micronuclei. ⑥ Micronuclei are then localized to the nuclear periphery, after which they form protrusions and can become eliminated from the cell. Once in the extracellular space, these micronuclei can be referred to as cell-free DNA. Because of a complex mechanism, the generation of double minutes via anaphase breakage-fusion-bridge (BFB) cycles has not been illustrated in this figure. However, the mechanism is shown in Supplementary Fig. 5.

investigation of the molecular characteristics of cfDNA, and inquiry into the mechanisms involved in its release, may provide deeper insight into the correlations observed between the properties of cfDNA and clinicopathological data, and may also expedite the search for different diagnostic, prognostic, and therapeutic cfDNA biomarkers.

## Conclusion

The cfDNA present in the culture medium of 143B cells after 24h of incubation is neither a consequence of apoptosis nor necrosis.<sup>24,25</sup> Herein, we have shown that

this DNA consists mainly of satellites, mini satellites, and transposable elements (TEs) that are either currently active in the human genome or show the capacity to become reactivated. We have also shown that these cfDNA fragments originate mainly from chromosomes 1 and 9. We propose the following hypothesis for the origin of these cfDNA fragments: the centromeric and pericentromeric regions of chromosomes 1 and 9 are predisposed to increased hypomethylation. On one hand, this can lead to the activation of transposons that are capable of inducing DSBs. On the other hand, structural changes within the centromeric and pericentromeric regions can directly impair the formation of



proper kinetochores and subsequently result in chromosome mis-segregation. The type of instability caused by these changes serves as a precursor to the formation of various nuclear anomalies, including lagging whole or fragmented chromosomes, and different types of anaphase bridges. While the exact sequence of events following the formation of these nuclear anomalies remains to be elucidated, numerous studies indicate that DNA fragments derived from these abnormal nuclear structures eventually recruit their own nuclear envelope, forming secondary nuclear structures like micronuclei and DMs (that contain extra-chromosomally amplified DNA). These structures are often localized to the nuclear periphery, after which they form protrusions and can be eliminated from the cell. Therefore, we hypothesize that the cfDNA present in the growth medium of 143B cells originates from these secondary nucleic structures. A summary of this hypothesis is illustrated in Figure 6.

If it turns out that we can in fact measure and distinguish between different kinds of micronuclei that are present in the circulation of a given individual, this will give us a unique window through which we can non-invasively monitor genome stability, as well as peer into the inner workings of cancer (and other genotoxic diseases) in a very non-invasive, more effective, manner. While observations of chromosome 1 and 9 centromeric and pericentromeric instability, anaphase bridging, and micronucleus formation in osteosarcoma tumors and cell lines lend some credibility to our hypothesis, it currently lacks direct experimental evidence. This can be unequivocally tested by correlating the characteristics of the cfDNA population with the biology of cells that are investigated under various conditions, including the (1) application of direct force to kinetochores using micromanipulation, (2) targeted chemical destabilization of kinetochore-microtubule attachments, (3) genetic knock-down of key mitotic machinery proteins and methylation enzymes, (4) chemical- or irradiation-induced DSBs, (5) deprivation of folate or similar metabolic precursors, and (6) prevention of normal methylation using base analogs (e.g. 5-Azacytidene, 2,6-diaminopurine, idoxuridine).

One limitation of this study is the low number of replicates ( $n = 2$ ) that were sequenced and the investigation of only one cell line. It will be of particular interest to compare the cfDNA sequence profiles of different cell lines, especially between healthy and malignant cell lines. One caveat of using cell cultures for elucidating the origin of cfDNA is that slight differences in culture conditions, such as the flask, medium composition, and temperature fluctuations, can result in significant changes in the biology of these cells (even if they are clones). Culturing a large panel of cell lines under identical settings may overcome this problem.<sup>115</sup>

Nevertheless, the information revealed in this study should contribute to our understanding of the origin of cfDNA and the extent of its biological significance. It may also serve as an entry point for bridging in vitro and in vivo studies.

## Acknowledgements

We would like to thank Dr E van Dyk (Thermo Fisher Scientific) for performing the sequencing of cell-free DNA.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: A.J.B. (Grant ID: SFH13092447078) and J.A. (Grant ID: SFH14061869958) were supported by post-graduate scholarships from the National Research Foundation (NRF), South Africa. The financial assistance of the NRF is hereby acknowledged. Opinions expressed and conclusions arrived at are those of the authors and are not to be attributed to the NRF. A.J.B. also thanks the African-German Network of Excellence in Science (AGNES) for granting a Mobility Grant in 2016; the Grant is generously sponsored by the German Federal Ministry of Education and Research and supported by the Alexander von Humboldt Foundation.

## ORCID iDs

Abel Jacobus Bronkhorst  <https://orcid.org/0000-0003-4809-8900>

Etienne Pierre de Villiers  <https://orcid.org/0000-0001-9878-1009>

## References

1. Jung K, Fleischhacker M and Rabien A. Cell-free DNA in the blood as a solid tumor biomarker: a critical appraisal of the literature. *Clin Chim Acta* 2010; 411: 1611–1624.
2. Swarup V and Rajeswari MR. Circulating (cell-free) nucleic acids: a promising, non-invasive tool for early detection of several human diseases. *FEBS Lett* 2007; 581: 795–799.
3. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer* 2017; 17: 223–238.
4. Bettgowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med* 2014; 6: 224ra24.
5. Bronkhorst AJ, Aucamp J and Pretorius PJ. Cell-free DNA: preanalytical variables. *Clin Chim Acta* 2015; 450: 243–253.

6. Bronkhorst AJ, Aucamp J and Pretorius PJ. Methodological variables in the analysis of cell-free DNA. *Adv Exp Med Biol* 2016; 924: 157–163.
7. El Messaoudi S, Rolet F, Mouliere F, et al. Circulating cell free DNA: preanalytical considerations. *Clin Chim Acta* 2013; 424: 222–230.
8. Schmidt B and Fleischhacker M. Is liquid biopsy ready for the litmus test and what has been achieved so far to deal with pre-analytical issues? *Transl Cancer Res* 2017; 7: 130–139.
9. Thierry AR, El Messaoudi S, Gahan PB, et al. Origins, structures, and functions of circulating DNA in oncology. *Cancer Metastasis Rev* 2016; 35: 347–376.
10. Jiang P and Lo YMD. The long and short of circulating cell-free DNA and the Ins and Outs of molecular diagnostics. *Trends Genet* 2016; 32: 360–371.
11. Snyder MW, Kircher M, Hill AJ, et al. Cell-free DNA comprises an in vivo nucleosome footprint that informs its tissues-of-origin. *Cell* 2016; 164: 57–68.
12. Ulz P, Thallinger GG, Auer M, et al. Inferring expressed genes by whole-genome sequencing of plasma DNA. *Nat Genet* 2016; 48: 1273–1278.
13. Murtaza M and Caldas C. Nucleosome mapping in plasma DNA predicts cancer gene expression. *Nat Genet* 2016; 48: 1105–1106.
14. Aucamp J, Bronkhorst AJ and Pretorius PJ. A historical and evolutionary perspective on circulating nucleic acids and extracellular vesicles: circulating nucleic acids as homeostatic genetic entities. *Adv Exp Med Biol* 2016; 924: 91–95.
15. Aucamp J, Bronkhorst AJ, Badenhorst CPS, et al. The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. *Biol Rev Camb Philos Soc* 2018; 93: 1649–1683.
16. Jiang P, Chan CWM, Chan KCA, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A* 2015; 112: E1317–1325.
17. Mouliere F, Robert B, Peyrotte E, et al. High fragmentation characterizes tumour-derived circulating DNA. *PLoS ONE* 2011; 6: e23418.
18. Forshew T, Murtaza M, Parkinson C, et al. Noninvasive identification and monitoring of cancer mutations by targeted deep sequencing of plasma DNA. *Sci Transl Med* 2012; 4: 136ra68.
19. Murtaza M, Dawson SJ, Tsui DWY, et al. Non-invasive analysis of acquired resistance to cancer therapy by sequencing of plasma DNA. *Nature* 2013; 497: 108–112.
20. Taylor KH, Kramer RS, Davis JW, et al. Ultradeep bisulfite sequencing analysis of DNA methylation patterns in multiple gene promoters by 454 sequencing. *Cancer Res* 2007; 67: 8511–8518.
21. Bryzgunova OE, Morozkin ES, Yarmoschuk SV, et al. Methylation-specific sequencing of GSTP1 gene promoter in circulating/extracellular DNA from blood and urine of healthy donors and prostate cancer patients. *Ann NY Acad Sci* 2008; 1137: 222–225.
22. Korshunova Y, Maloney RK, Lakey N, et al. Massively parallel bisulphite pyrosequencing reveals the molecular complexity of breast cancer-associated cytosine-methylation patterns obtained from tissue and serum DNA. *Genome Res* 2008; 18: 19–29.
23. Beck J, Urnovitz HB, Riggert J, et al. Profile of the circulating DNA in apparently healthy individuals. *Clin Chem* 2009; 55: 730–738.
24. Bronkhorst AJ, Wentzel JF, Aucamp J, et al. An enquiry concerning the characteristics of cell-free DNA released by cultured cancer cells. *Adv Exp Med Biol* 2016; 924: 19–24.
25. Bronkhorst AJ, Wentzel JF, Aucamp J, et al. Characterization of the cell-free DNA released by cultured cancer cells. *Biochim Biophys Acta* 2016; 1863: 157–165.
26. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; 409: 860–921.
27. De Koning APJ, Gu W, Castoe TA, et al. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet* 2011; 7: e1002384.
28. Yandell M and Ence D. A beginner's guide to eukaryotic genome annotation. *Nat Rev Genet* 2012; 13: 329–342.
29. Bennett EA, Keller H, Mills RE, et al. Active Alu retrotransposons in the human genome. *Genome Res* 2008; 18: 1875–1883.
30. Krull M, Petrusma M, Makalowski W, et al. Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs). *Genome Res* 2007; 17: 1139–1145.
31. Dewannieux M, Esnault C and Heidmann T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* 2003; 35: 41–48.
32. Comeaux MS, Roy-Engel AM, Hedges DJ, et al. Diverse cis factors controlling Alu retrotransposition: what causes Alu elements to die? *Genome Res* 2009; 19: 545–555.
33. Kapitonov VV, Pavlicek A and Jurka J. Anthology of human repetitive DNA. In: Meyers RA (ed.) *Reviews in cell biology and molecular medicine*. DOI: 10.1002/3527600906.mcb.200300166.
34. Cordaux R and Batzer MA. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* 2009; 10: 691–703.
35. Boissinot S, Chevret P and Furano AV. L1 (LINE-1) retrotransposon evolution and amplification in recent human history. *Mol Biol Evol* 2000; 17: 915–928.
36. Skowronski J, Fanning TG and Singer MF. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol Cell Biol* 1988; 8: 1385–1397.
37. Smit AFA, Tóth G, Riggs AD, et al. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J Mol Biol* 1995; 246: 401–417.
38. Lee J, Cordaux R, Han K, et al. Different evolutionary fates of recently integrated human and chimpanzee LINE-1 retrotransposons. *Gene* 2007; 390: 18–27.
39. Lavie L, Kitova M, Maldener E, et al. CpG methylation directly regulates transcriptional activity of the human endogenous retrovirus family HERV-K(HML-2). *J Virol* 2005; 79: 876–883.
40. Menendez L, Benigno BB and McDonald JF. L1 and HERV-W retrotransposons are hypomethylated in human ovarian carcinomas. *Mol Cancer* 2004; 3: 12.
41. Kleiman A, Senyuta N, Tryakin A, et al. HERV-K(HML-2) GAG/ENV antibodies as indicator for

- therapy effect in patients with germ cell tumors. *Int J Cancer* 2004; 110: 459–461.
42. Bourc'his D and Bestor TH. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* 2004; 431: 96–99.
  43. Huang J, Fan T, Yan Q, et al. Lsh, an epigenetic guardian of repetitive elements. *Nucleic Acids Res* 2004; 32: 5019–5028.
  44. Robertson HM and Martos R. Molecular evolution of the second ancient human mariner transposon, Hsmar2, illustrates patterns of neutral evolution in the human genome lineage. *Gene* 1997; 205: 219–228.
  45. Feschotte C and Pritham EJ. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet* 2007; 41: 331–368.
  46. Pace JK and Feschotte C. The evolutionary history of human DNA transposons: evidence for intense activity in the primate lineage. *Genome Res* 2007; 17: 422–432.
  47. Wilson AS, Power BE and Molloy PL. DNA hypomethylation and human diseases. *Biochim Biophys Acta* 2007; 1775: 138–162.
  48. Hoffmann MJ and Schulz WA. Causes and consequences of DNA hypomethylation in human cancer. *Biochem Cell Biol* 2005; 83: 296–321.
  49. Narayan A, Ji W, Zhang XY, et al. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer* 1998; 77: 833–838.
  50. Qu GZ, Dubeau L, Narayan A, et al. Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. *Mutat Res* 1999; 423: 91–101.
  51. Qu G, Grundy PE, Narayan A, et al. Frequent hypomethylation in Wilms tumors of pericentromeric DNA in chromosomes 1 and 16. *Cancer Genet Cytogenet* 1999; 109: 34–39.
  52. Hansen RS, Wijmenga C, Luo P, et al. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc Natl Acad Sci U S A* 1999; 96: 14412–14417.
  53. Sawyer JR, Swanson CM, Wheeler G, et al. Chromosome instability in ICF syndrome: formation of micronuclei from multibranched chromosomes 1 demonstrated by fluorescence in situ hybridization. *Am J Med Genet* 1995; 56: 203–209.
  54. Ehrlich M, Jackson K and Weemaes C. Immunodeficiency, centromeric region instability, facial anomalies syndrome (ICF). *Orphanet J Rare Dis* 2006; 1: 2.
  55. Stacey M, Bennett MS and Hulten M. Fish analysis on spontaneously arising micronuclei in the ICF syndrome. *J Med Genet* 1995; 32: 502–508.
  56. Xu GL, Bestor TH, Bourc'his D, et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* 1999; 402: 187–191.
  57. Mitelman F. *Catalog of chromosome aberrations in cancer: chromosomes 1-12;[2], chromosomes 13-22, X, Y, references*. Wiley-Liss, 1994.
  58. Jeanpierre M, Turleau C, Aurias A, et al. An embryonic-like methylation pattern of classical satellite DNA is observed in ICF syndrome. *Hum Mol Genet* 1993; 2: 731–735.
  59. Ji W, Hernandez R, Zhang XY, et al. DNA demethylation and pericentromeric rearrangements of chromosome 1. *Mutat Res* 1997; 379: 33–41.
  60. Cimini D, Tanzarella C and Degraffi F. Effects of 5-azacytidine on the centromeric region of human fibroblasts studied by CREST staining and in situ hybridization on cytokinesis-blocked cells. *Cytogenet Cell Genet* 1996; 72: 219–224.
  61. Fauth E and Zankl H. Comparison of spontaneous and idoxuridine-induced micronuclei by chromosome painting. *Mutat Res* 1999; 440: 147–156.
  62. Fauth E, Scherthan H and Zankl H. Chromosome painting reveals specific patterns of chromosome occurrence in mitomycin C- and diethylstilboestrol-induced micronuclei. *Mutagenesis* 2000; 15: 459–467.
  63. Vilain A, Bernardino J, Gerbault-Seureau M, et al. DNA methylation and chromosome instability in lymphoblastoid cell lines. *Cytogenet Genome Res* 2000; 90: 93–101.
  64. Westhorpe FG and Straight AF. Functions of the centromere and kinetochore in chromosome segregation. *Curr Opin Cell Biol* 2013; 25: 334–340.
  65. Cimini D and Degraffi F. Aneuploidy: a matter of bad connections. *Trends Cell Biol* 2005; 15: 442–451.
  66. Godek KM, Kabeche L and Compton DA. Regulation of kinetochore-microtubule attachments through homeostatic control during mitosis. *Nat Rev Mol Cell Biol* 2015; 16: 57–64.
  67. Ford JH and Correll AT. Chromosome errors at mitotic anaphase. *Genome* 1992; 35: 702–705.
  68. Lindholm C, Norppa H, Hayashi M, et al. Induction of micronuclei and anaphase aberrations by cytochalasin B in human lymphocyte cultures. *Mutat Res* 1991; 260: 369–375.
  69. Catalán J, Falck GC and Norppa H. The X chromosome frequently lags behind in female lymphocyte anaphase. *Am J Hum Genet* 2000; 66: 687–691.
  70. Norppa H and Falck GCM. What do human micronuclei contain? *Mutagenesis* 2003; 18: 221–233.
  71. Kisurina-Evgenieva OP, Sutiagina OI and Onishchenko GE. Biogenesis of micronuclei. *Biochemistry* 2016; 81: 453–464.
  72. Stimpson KM, Matheny JE and Sullivan BA. Dicentric chromosomes: unique models to study centromere function and inactivation. *Chromosome Res* 2012; 20: 595–605.
  73. MacKinnon RN and Campbell LJ. The role of dicentric chromosome formation and secondary centromere deletion in the evolution of myeloid malignancy. *Genet Res Int* 2011; 2011: 643628.
  74. Stewenius Y, Gorunova L, Jonson T, et al. Structural and numerical chromosome changes in colon cancer develop through telomere-mediated anaphase bridges, not through mitotic multipolarity. *Proc Natl Acad Sci U S A* 2005; 102: 5541–5546.
  75. Medvedeva NG, Panyutin IV, Panyutin IG, et al. Phosphorylation of histone H2AX in radiation-induced micronuclei. *Radiat Res* 2007; 168: 493–498.
  76. Haaf T, Raderschall E, Reddy G, et al. Sequestration of mammalian Rad51-recombination protein into micronuclei. *J Cell Biol* 1999; 144: 11–20.
  77. Fenech M, Kirsch-Volders M, Natarajan AT, et al. Molecular mechanisms of micronucleus, nucleoplasmic bridge

- and nuclear bud formation in mammalian and human cells. *Mutagenesis* 2011; 26: 125–132.
78. Shimizu N. Molecular mechanisms of the origin of micronuclei from extrachromosomal elements. *Mutagenesis* 2011; 26: 119–123.
  79. Shimizu N, Shingaki K, Kaneko-Sasaguri Y, et al. When, where and how the bridge breaks: anaphase bridge breakage plays a crucial role in gene amplification and HSR generation. *Exp Cell Res* 2005; 302: 233–243.
  80. Shimizu N, Shimura T and Tanaka T. Selective elimination of acentric double minutes from cancer cells through the extrusion of micronuclei. *Mutat Res* 2000; 448: 81–90.
  81. McClintock B. The fusion of broken ends of chromosomes following nuclear fusion. *Proc Natl Acad Sci U S A* 1942; 28: 458–463.
  82. Fenech M. Cytokinesis-block micronucleus assay evolves into a “cytome” assay of chromosomal instability, mitotic dysfunction and cell death. *Mutat Res* 2006; 600: 58–66.
  83. Shimizu N, Itoh N, Utiyama H, et al. Selective entrapment of extrachromosomally amplified DNA by nuclear budding and micronucleation during S phase. *J Cell Biol* 1998; 140: 1307–1320.
  84. Pandis N, Jin Y, Gorunova L, et al. Chromosome analysis of 97 primary breast carcinomas: identification of eight karyotypic subgroups. *Genes Chromosomes Cancer* 1995; 12: 173–185.
  85. Austruy E, Candon S, Henry I, et al. Characterization of regions of chromosomes 12 and 16 involved in nephroblastoma tumorigenesis. *Genes Chromosomes Cancer* 1995; 14: 285–294.
  86. Tymms MJ, Ng AY, Thomas RS, et al. A novel epithelial-expressed ETS gene, ELF3: human and murine cDNA sequences, murine genomic organization, human mapping to 1q32.2 and expression in tissues and cancer. *Oncogene* 1997; 15: 2449–2462.
  87. Utani K, Okamoto A and Shimizu N. Generation of micronuclei during interphase by coupling between cytoplasmic membrane blebbing and nuclear budding. *PLoS ONE* 2011; 6: e27233.
  88. Gisselsson D, Pålsson E, Yu C, et al. Mitotic instability associated with late genomic changes in bone and soft tissue tumours. *Cancer Lett* 2004; 206: 69–76.
  89. Scheel C, Schaefer KL, Jauch A, et al. Alternative lengthening of telomeres is associated with chromosomal instability in osteosarcomas. *Oncogene* 2001; 20: 3835–3844.
  90. Lim G, Karaskova J, Vukovic B, et al. Combined spectral karyotyping, multicolor banding, and microarray comparative genomic hybridization analysis provides a detailed characterization of complex structural chromosomal rearrangements associated with gene amplification in the osteosarcoma cell line MG-63. *Cancer Genet Cytogenet* 2004; 153: 158–164.
  91. Lim G, Karaskova J, Beheshti B, et al. An integrated mBAND and submegabase resolution tiling set (SMRT) CGH array analysis of focal amplification, microdeletions, and ladder structures consistent with breakage-fusion-bridge cycle events in osteosarcoma. *Genes Chromosomes Cancer* 2005; 42: 392–403.
  92. Selvarajah S, Yoshimoto M, Park PC, et al. The breakage-fusion-bridge (BFB) cycle as a mechanism for generating genetic heterogeneity in osteosarcoma. *Chromosoma* 2006; 115: 459–467.
  93. Hoffelder DR, Luo L, Burke NA, et al. Resolution of anaphase bridges in cancer cells. *Chromosoma* 2004; 112: 389–397.
  94. Jin Y, Jin C, Wennerberg J, et al. Cytogenetic and fluorescence in situ hybridization characterization of chromosome 8 rearrangements in head and neck squamous cell carcinomas. *Cancer Genet Cytogenet* 2001; 130: 111–117.
  95. Eichler EE. Masquerading repeats: paralogous pitfalls of the human genome. *Genome Res* 1998; 8: 758–762.
  96. Gisselsson D, Höglund M, Mertens F, et al. Variable stability of chromosomes containing amplified  $\alpha$ -satellite sequences in human mesenchymal tumours. *Chromosoma* 1999; 108: 271–277.
  97. Bayani J, Zielenska M, Pandita A, et al. Spectral karyotyping identifies recurrent complex rearrangements of chromosomes 8, 17, and 20 in osteosarcomas. *Genes Chromosomes Cancer* 2003; 36: 7–16.
  98. Hedges DJ and Deininger PL. Inviting instability: transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* 2007; 616: 46–59.
  99. Michel B, Ehrlich SD and Uzest M. DNA double-strand breaks caused by replication arrest. *EMBO J* 1997; 16: 430–438.
  100. Gasior SL, Wakeman TP, Xu B, et al. The human LINE-1 retrotransposon creates DNA double-strand breaks. *J Mol Biol* 2006; 357: 1383–1393.
  101. Brouha B, Schustak J, Badge RM, et al. Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci U S A* 2003; 100: 5280–5285.
  102. Kipling D and Warburton PE. Centromeres, CENP-B and Tigger too. *Trends Genet* 1997; 13: 141–145.
  103. Muro Y, Masumoto H, Yoda K, et al. Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol* 1992; 116: 585–596.
  104. Casola C, Hucks D and Feschotte C. Convergent domestication of pogo-like transposases into centromere-binding proteins in fission yeast and mammals. *Mol Biol Evol* 2008; 25: 29–41.
  105. Meštrović N, Pavlek M, Car A, et al. Conserved DNA motifs, including the CENP-B box-like, are possible promoters of satellite DNA array rearrangements in nematodes. *PLoS ONE* 2013; 8: e67328.
  106. Meštrović N, Mravinac B, Pavlek M, et al. Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Res* 2015; 23: 583–596.
  107. Sun C, Wyngaard G, Walton DB, et al. Billions of base-pairs of recently expanded, repetitive sequences are eliminated from the somatic genome during copepod development. *BMC Genomics* 2014; 15: 186.
  108. Streit A. Silencing by throwing away: a role for chromatin diminution. *Dev Cell* 2012; 23: 918–919.
  109. Vogt A, Goldman AD, Mochizuki K, et al. Transposon domestication versus mutualism in ciliate genome rearrangements. *PLoS Genet* 2013; 9: e1003659.
  110. Nowacki M, Higgins BP, Maquilan GM, et al. A functional role for transposases in a large eukaryotic genome. *Science* 2009; 324: 935–938.

111. Aucamp J, Bronkhorst AJ, Peters DL, et al. Kinetic analysis, size profiling, and bioenergetic association of DNA released by selected cell lines in vitro. *Cell Mol Life Sci* 2017; 74: 2689–2707.
112. MagMAX Cell-Free DNA Isolation Kit. *A complete next-generation sequencing workflow for circulating cell-free DNA isolation and analysis*, <https://assets.thermofisher.com/TFS-Assets/LSG/Application-Notes/cfDNA-appnote.pdf>
113. Kumar P, Dillon LW, Shibata Y, et al. Normal and cancerous tissues release extrachromosomal circular DNA (eccDNA) into the circulation. *Mol Cancer Res* 2017; 15: 1197–1205.
114. Zhu J, Zhang F, Du M, et al. Molecular characterization of cell-free eccDNAs in human plasma. *Sci Rep* 2017; 7: 1–11.
115. Kuijjer ML, Hogendoorn PCW and Cleton-Jansen AM. Genome-wide analyses on high-grade osteosarcoma: making sense of a genomically most unstable tumor. *Int J Cancer* 2013; 133: 2512–2521.