Characterization of circulating free DNA in healthy and diseased individuals

Maniesh van der Vaart M.Sc

Thesis submitted for the degree Doctor of Philosophy at the Potchefstroom campus of the North-West University

Promoter: Prof. P.J. Pretorius

May 2009
You may never know what results come from your action.

But if you do nothing, there will be no result.

Mahatma Gandhi
Abstract

Circulating DNA is small fragments of DNA present in the blood of vertebrates. Even though the origin, function and significance of these DNA fragments are not elucidated yet, it is widely investigated as a source of biomarkers for cancer as well as other ailments. This is a result of the detection of a higher concentration of circulating DNA as well as a number of tumor related changes in the blood of cancer patients compared to healthy individuals.

The main theories involving the origin of circulating DNA include apoptosis, necrosis and active release of DNA by living cells. A number of arguments against apoptosis and necrosis and in favor of an active release mechanism as the main source of circulating DNA were presented. It was concluded that apoptosis is not the main source of circulating DNA and that an active release mechanism may be a viable alternative. Furthermore, the reason for the increased circulating DNA concentration observed in cancer patients might be the disturbance of the equilibrium between release and clearance of circulating DNA.

The utilization and role of circulating DNA as a biomarker for cancer was investigated, and the major methods being utilized to investigate circulating DNA as a biomarker include quantification, detection of genetic alterations and epigenetic analysis of circulating DNA. Comparison of the results obtained by a number of studies lead to the conclusion that circulating DNA is currently overrated as a biomarker for cancer, and before it can be used as an informative biomarker the significance and cause of fluctuation first need to be determined.

A method to clone and sequence circulating DNA from cancer patients as well as healthy individuals was utilized in two pilot studies to sequence circulating DNA on a small scale. Analysis of the obtained sequences exemplified the need for large scale sequencing. As a result a method for large scale sequencing of circulating DNA via parallel tagged sequencing on a GS FLX sequencer (454 Life sciences, Roche) were standardized and used to generate sequences obtained from healthy as well as diseased individuals. These sequences were compared between the two groups and to the human genome in an attempt to characterize circulating DNA.

The primary aim of this study was the large scale sequencing of circulating DNA, and secondarily, to determine if a significant difference exists between circulating DNA from cancer patients and healthy individuals. A total number of ~5500 usable sequences were obtained and analyses of the sequences revealed that, except for a higher frequency of mutations that could be detected in the circulating DNA of cancer patients compared to those from healthy individuals, a significant difference between these two groups were not evident. Thus it can be concluded that this study was successful, all aims were reached, but further sequence analysis will be pursued.

Key words: Circulating DNA, plasma, sequencing, 454 sequencing™
Uittreksel

Sirkulerende DNA is kort fragmente DNA wat teenwoordig is in die bloed van vertebrate. Selfs al is die oorsprong, funksie en belang of betekenis van hierdie molekules nog nie bekend nie, word dit alom ondersoek as ‘n bron van biomarkers vir kanker sowel as ander siektetoestande as gevolg van die feit dat hoër konsentrasies sirkulerende DNA, sowel as kanker verwante veranderinge, in die bloed van kanker pasiente waargeneem kan word indien dit met gesonde persone vergelyk word.

Die hoof teorie in verband met die oorsprong van sirkulerende DNA sluit apoptose, nekrose en aktiewe vrystelling deur lewende selle in. ‘n Aantal argumente teen apoptose en nekrose en in die guns van ‘n aktiewe vrystellings mekanisme as die hoof bron van sirkulerende DNA is aangebied. Daar is tot die gevolgtrekking gekom dat apoptose nie die hoof bron van sirkulerende DNA is nie en dat aktiewe vrystelling van DNA ‘n werklike alternatief kan wees. Die rede vir die verhoogde konsentrasie sirkulerende DNA wat by kanker pasiente waargeneem kan word is moontlik ‘n verstering in die ekwilibriuim tussen die vrystelling en verwyding van sirkulerende DNA in en uit die bloed.

Die gebruikte en rol van sirkulerende DNA as ‘n biomarker vir kanker is ondersoek en die metodes wat gebruik word om biomarkers op te spoor is hoofsaaklik kwantifisering, opsporing van genetiese veranderinge en epigenetiese analyse van sirkulerende DNA. Deur die resultate van ‘n aantal studies te vergelyk kon die afleiding gemaak word dat sirkulerende DNA tans oorskat word as ‘n biomarker vir kanker. Voordat sirkulerende DNA as ‘n informatiewe bron van biomarkers gebruik kan word is dit nodig om die belang of betekenis, sowel as die oorsaak van fluktuasie in die konsentrasie van sirkulerende DNA, te bepaal.

‘n Metode om sirkulierende DNA te kloon en volgorde bepaling van DNA afkomstig van kanker pasiente sowel as gesonde persone te doen, is gebruik om twee proef studies te doen waarin volgorde bepaling van sirkulerende DNA op klein skaal gedoen is. Die volgorde wat verkry is is geanalyser en daar is bevind dat groot skaalse volgorde bepaling benodig word. Dus is ‘n metode vir die groot skaalse volgorde bepaling deur paralelle merking van volgorde op die GS FLX volgorde bepaler (454 Life sciences, Roche) gestandardiseer en gevolglik gebruik om volgorde van sirkulerende DNA vanaf kanker pasiente en gesonde persone te bepaal. Die twee groep se volgorde wat bekom is, is met mekaar sowel as met die mens genoom vergelyk in ‘n poging om sirkulierende DNA te karakteriseer.

Die pimère doel van hierdie studie was om groot skaalse volgorde bepaling van sirkulerende DNA uit te voer, en sekondêr, om te bepaal of ‘n betekenisvolle verskil tussen sirkulerende DNA vanaf kanker pasiente en gesonde persone bestaan. ‘n Totaal van ~5500 bruikbare volgorde is bekom en volgorde analisie het gewys dat, behalwe vir ‘n vertoging in die hoeveelheid mutasies wat in die volgorde van kanker pasiente voorkom wanneer dit met die van gesonde persone vergelyk word, is daar nie ‘n betekenisvolle verskil tussen hierdie twee groep nie. Dus kan die afleiding gemaak word dat hierdie studie suksesvol afgehandel is, nieteenstaande sal verdere volgorde analisie wel gedoen word.

Sleutelwoorde: Sirkulerende DNA, plasma, volgorde bepaling, 454 sequencing™
Acknowledgements

I thank our heavenly father for giving me the knowledge, perseverance and faith to be able to complete this study. Without Him this would not have been possible, I also thank him for sending the following individuals over my path:

- Prof. Piet Pretorius who are a great inspiration, role model and mentor to me not only in science but also in life. I could not have asked for a better promoter! You gave me great opportunities and advice, you were always there for me with a word of encouragement or wisdom, you are invaluable and I will never forget your contribution in my life.
- Dmitry Semerov at the Institute of Chemical Biology and Fundamental Medicine in Novosibirsk, Russia for much needed help and assistance with bioinformatics and also valuable advice
- Mathias Meyer for making the sequences of the barcoding adapters and the protocol for tagging available to us even before it was published
- Leonard Santana for writing the algorithm used to draw Figure 1 in chapter 3
- Philip Law and Prof. Fourie Joubert at the Bioinformatics and Computational Biology unit (FABI square) of the University of Pretoria for help with bioinformatics
- Dr. Vonkeman for supplying names of willing patients to participate in the study
- Patients and controls that were willing to give blood and personal information for this study
- AmPath, Lancet and PathCare for blood collection from patients and controls
- Inqaba Biotec for use of their fluoro-spectrometer; for sequencing and advice
- Ruan Steyn, the love of my life, for support and always being there for me, but most of all for unconditional love
- My parents, Henk and Martie van der Vaart, for always believing in me, supporting me and for giving me the opportunity to further my studies and enrich my life
- My sister, Nadia and all my friends, especially Amaria, for encouragement, help and support

I thank you from the bottom of my heart, you inspired me, motivated me and helped me to achieve this great accolade.
Table of contents

ABSTRACT

UITREKSEL

ACKNOWLEDGEMENTS

CHAPTER 1

Introduction
  1.1 Problem statement and substantiation
  1.2 Research aims and objectives
  1.3 Structure of thesis

CHAPTER 2

What is free circulating DNA?
  2.1 Paper published in Annals of the New York Academy of Sciences
  2.2 Letter to the editor published in Clinical Chemistry

CHAPTER 3

Circulating DNA as a biomarker, what is currently being done?
  3.1 Paper accepted for publication in Clinical Biochemistry

CHAPTER 4

Sequencing of Free Circulating DNA
  4.1 Letter to the editor published in Clinica Chimica Acta
  4.2 Paper published in Annals of the New York Academy of Sciences

CHAPTER 5

What is the significance of free circulating DNA? More results from a bigger pilot study
  5.1 Prepared manuscript

CHAPTER 6

454 sequencing™ GS FLX – massive sequencing
  6.1 Paper published in Clinica Chimica Acta
  6.2 Some technical aspects
  6.3 Suggested improvements to the sequencing method

CHAPTER 7

Summary and conclusions

REFERENCES
Table of contents (continued)

SUPPLEMENTARY MATERIAL  

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>List of figures, tables and abbreviations</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>List of figures</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>List of tables</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Abbreviations</td>
<td>57</td>
</tr>
<tr>
<td>2</td>
<td>Materials, methods and supporting results</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>i. Blood collection and processing</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>ii. Isolation of circulating DNA</td>
<td>59</td>
</tr>
<tr>
<td></td>
<td>iii. Cloning and sequencing in the pilot study</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>iv. Preparation of circulating DNA for 454 sequencing™ (GS FLX)</td>
<td>62</td>
</tr>
<tr>
<td>3</td>
<td>Informed consent form</td>
<td>70</td>
</tr>
<tr>
<td>4</td>
<td>Conference proceedings</td>
<td>75</td>
</tr>
<tr>
<td>5</td>
<td>Permission from editors to use published papers</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>Supplementary material for published papers</td>
<td>82</td>
</tr>
<tr>
<td>7</td>
<td>Permission from co-authors</td>
<td>89</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Mandel and Métais isolated free circulating DNA and RNA by perchloric acid precipitation in 1948, that was even before Watson and Crick elucidated the structure of DNA (Anker et al., 1999)*; but no attention was given to this fact. In 1965 Bendich et al. (Guadajara et al., 2008)* suggested that circulating DNA might be a factor involved in oncogenesis and in 1966 Tan et al. (Ziegler et al., 2002; Goebel et al., 2005)* proposed that circulating DNA might be from endogenous tissue breakdown as it could be detected in diseases associated with tissue destruction like systemic lupus erythematosus, they also identified DNA antibodies which was used in 1977 by Leon et al. to do sensitive radioimmunoassay. Stroun and Anker reported in 1972 that nucleic acids could be spontaneously released by living frog auricles which was the first in vitro evidence of living cells releasing DNA. Leon et al. (1977) detected higher amounts of circulating DNA in the blood of half the cancer patients compared to the healthy controls they examined, they also associated higher amounts of circulating DNA with cancer metastasis and reported some correlation with response to therapy. In 1983 significant differences in the amounts of circulating DNA in the blood of patients with benign and malignant disease was reported by Shapiro et al. (1983) who identified the possible diagnostic and prognostic value of circulating DNA. Stroun et al. (1987) showed that circulating DNA is double stranded fragments of human origin and some studies could not detect circulating DNA in normal individuals because of poor sensitivity of the methods used (Steinman, 1975; Stroun et al., 1987; Stroun et al., 1989). The first evidence that circulating DNA is, in part, released by tumor cells as it has neoplastic characteristics was provided by Stroun et al. in 1989, furthermore, they reported that normal human lymphocytes can actively release a DNA containing complex in vitro and they recognized the possibility that circulating cancerous DNA may cause cancer metastasis (Stroun et al.; 1989). In the same year Li and Steinman attempted to characterize circulating DNA and reported that it represent a nonrandom selection of human DNA and also contained Alu repeat sequences (Li and Steinman, 1989). In 1994 Vasioukhin et al. and Sorensen et al. (Ziegler et al., 2002)* respectively showed the presence of N-ras and K-ras mutations in circulating DNA, this was followed by detection of microsatellite alterations, loss of heterozygosity, methylation detection and various other mutations present in circulating DNA (Bremnes et al., 2005). Lo et al. (1998) showed the presence of fetal DNA circulating in maternal blood.

* Secondary reference – as original paper is not available or in another language.
Through 80 years of research and after more than 400 papers that have been published on the subject no conclusive evidence for the origin of circulating DNA is available and the function, significance or composition of circulating DNA has not been elucidated yet.

1.1 Problem statement and substantiation

Circulating DNA is fragments of DNA of unknown origin circulating in the blood of humans and other vertebrates.

Stroun and Anker have been working on the topic of circulating DNA for more than 40 years and in 2005 they wrote a paper about the combined findings of their work. They concentrated on a hypothesis describing the release of nucleic acids by living cells (Stroun and Anker, 2005). In the introduction of this paper they mentioned the difficulties they encountered in publishing and finding funds for “this unorthodox research” and despite good results they felt as if they were “talking to the wind”. Despite the fact that they are well known in the community of researchers studying circulating DNA and that they have shown that living cells spontaneously release DNA in vitro this hypothesis is still not accepted. This state of affairs may be attributed to a lack of knowledge about the origin and nature of circulating DNA, especially nucleotide sequence information.

The sequence information of circulating DNA is to a large extent unknown, only in the last few months a few papers on large scale sequencing of circulating DNA was published. However the objective of these studies were not to characterize circulating DNA but rather ultra-deep amplicon sequencing of amplification products of a small number of gene sequences or methylation specific sequencing (Taylor et al., 2007, Bryzgunova et al., 2008, Koishunova et al., 2008). One study used shotgun pyrosequencing for diagnosis of fetal aneuploidy (Fan et al., 2008) and was able to diagnose different trisomy cases very early in pregnancy. A recent study by Beck et al. (2009) described sequencing of total circulating DNA from serum of 51 apparently healthy individuals, a large amount of sequences were obtained and analysis revealed some content information of circulating DNA.

Sequencing of circulating DNA may prove to be beneficial in identification of biomarkers for cancer and other diseases which present with increased circulating DNA in the blood of patients, it may also contribute in determining the origin, function and significance of this nucleic acid molecules.

Ethics approval for this study was obtained from the ethics committee of the North-West University, reference number 05M12.
1.2 Research aims and objectives

The aim of this project was primarily to sequence circulating DNA on a large scale and, secondary, to use this information to determine whether a significant difference exists between circulating DNA obtained from healthy and diseased individuals.

1.3 Structure of thesis

This thesis was compiled in article format and consists of three published papers, two published letters to editors, one accepted manuscript and a pilot study in manuscript format. Each paper, letter or manuscript was inserted in the text exactly as published or submitted and thus it complies with the requirements set by the various journals. Parts of the following papers: "Circulating DNA: its origin and fluctuation" and "A method for characterization of total circulating DNA" describe work done for the M.Sc degree and were included as preliminary manuscripts in the masters dissertation, however, the original manuscripts were extensively revised for publication and inclusion in this thesis.

The question of what circulating DNA is as well as the origin of circulating DNA is discussed in chapter two, chapter three deals with circulating DNA as a biomarker and what is currently being done in an effort to exploit this potential. In chapter four, a method for sequencing circulating DNA and the problems surrounding it is discussed, the pilot study and the results obtained from it is described in chapter five while chapter six deals with the massive sequencing of circulating DNA. In Chapter seven the results and discussion are presented followed by the references. Note that references for each article are given at the end of the article and are not included in the general reference list. The thesis ends with a compilation of supplementary information such as a list of figures, tables and abbreviations (Section 1) as well as materials and methods used accompanied with the results obtained (Section 2), the informed consent form (Section 3), conference proceedings (Section 4), permission to use published papers (Section 5 and 7) as well as supplementary material for the research paper in Chapter 3 (Section 6).
Chapter 2

What is free circulating DNA?

Free circulating DNA is small fragments of DNA that circulate in the blood of humans and probably in all eukaryotes (frogs, mice and also plants, (Stroun and Anker, 2005)). These nucleic acid fragments can be isolated from the plasma or serum of the subject. Circulating DNA can also be isolated from the culture medium of cells for in vitro work.

Different types of circulating DNA can be found in the blood of cancer patients such as genetically altered DNA, microsatellite DNA, methylated DNA, nucleosomal DNA, mitochondrial DNA and viral DNA. Diehl et al. (2005) found the mutant sequences in the plasma to be degraded compared to those derived from nonneoplastic cells while Muller et al. (2008) is of opinion that the shorter fragments of circulating DNA is enriched with tumor DNA.

Fleischhacker and Schmidt (2007) did an extensive review on circulating nucleic acids, which include RNA, DNA and viral molecules. From a total of 443 papers, less than 6% of the reviewed papers contributed to knowledge on the origin, half-life and removal of circulating nucleic acids and less than 10% contributed to the biological meaning of these molecules, the majority of papers deals with the subject of the detection of tumor related changes, quantification, methodology and a few other aspects related to free circulating nucleic acids.

Where does circulating DNA come from? In an attempt to answer this question and to review other issues like the characteristics and occurrence of circulating DNA, the main points of view that exist in the literature that explain the origin of circulating DNA as well as additional sources that might play a role in the origin, but do not enjoy so much attention in the literature, as well as the clearance of these molecules and also other conditions that may give rise to circulating DNA were highlighted in the following published paper. The main thrust of this paper was also presented in a letter to the editor of Clinical Chemistry.

2.1 Paper published in Annals of the New York Academy of Sciences

2.2 Letter to the editor published in Clinical Chemistry
Circulating DNA
Its Origin and Fluctuation
Maniesh van der Vaart and Piet J. Pretorius
School of Biochemistry, North-West University, Potchefstroom Campus, Potchefstroom, South Africa

Circulating DNA is present in the blood of all individuals, but it has been found that cancer patients and patients with a variety of other conditions have increased amounts of these circulating DNA fragments in their blood. Even though more than 30 years of research have been done on this subject, the origin of these nucleic acid molecules is still not clear. We present evidence that does not support the general notion that apoptosis or necrosis is the major source of circulating free DNA. Active release of free circulating DNA by living cells may be a plausible mechanism. Disturbance of the equilibrium between the release of DNA by living cells and the mechanisms used for clearing this DNA may play the main role in the appearance of increased amounts of circulating DNA in the blood of individuals with different ailments. Elucidating the origin and the mechanism that cells use to release free circulating DNA into the blood may enhance the diagnostic and prognostic value of these nucleic acid molecules.

Key words: apoptosis; active release; free circulating DNA; blood plasma

Introduction

It is widely known that higher concentrations of free circulating DNA can be found, in most cases, in the blood of patients with malignant diseases compared to healthy subjects.\textsuperscript{1} Several studies have been performed to establish whether a significant diagnostic and/or prognostic\textsuperscript{1} use could be found for circulating free DNA, both in quantity and quality, because of the noninvasive nature in which it can be obtained. Although much work has been done to determine the mechanism whereby these circulating DNA fragments are released into the blood, a definite conclusion could not be reached.

This chapter will address some aspects of the origin of circulating DNA in normal and in pathologic conditions such as cancer, trauma, stroke, pregnancy, autoimmune disorders, and after solid organ transplant in an effort to shed some light on the origin of this DNA. The putative role of apoptosis and necrosis in their origin will be emphasized, and the release of DNA by living cells as well as clearance of these molecules will be highlighted.

Characteristics and Occurrence of Circulating DNA

The double-stranded nature of free circulating DNA was shown as early as 1975 with hydroxyapatite chromatography and with density gradient centrifugation.\textsuperscript{2,4} It was also in these early days that the low molecular weight and ladder pattern of circulating DNA was revealed by agarose gel electrophoresis.\textsuperscript{4} If this electrophoretic ladder pattern is used as an indication of the size distribution of circulating DNA, it can be accepted that it is in the range of 180 to 10,000 bp.\textsuperscript{5} Furthermore, it is possible that the ends of these molecules are capped. Free circulating DNA may also be
present in the form of nucleosomes or apoptotic bodies.\textsuperscript{6,7}

The amount and composition of circulating DNA varies between patients,\textsuperscript{5} and many reports agree\textsuperscript{8} that cancer patients have much larger amounts of circulating DNA in their blood than do healthy subjects. Early analysis could not detect any DNA in the serum of healthy individuals.\textsuperscript{4,9} However, circulating DNA is not confined to serum or plasma of cancer patients, since elevated amounts of circulating free DNA can also be detected in patients with other pathologic conditions such as systemic lupus erythematosus (SLE), rheumatoid arthritis, glomerulonephritis, pancreatitis, cholelithiasis, inflammatory bowel disease, peptic ulcer disease, hepatitis, esophagitis, pulmonary embolism, ulcerative colitis, and miliary tuberculosis (for review see Anker et al.\textsuperscript{10} and Ziegler et al.\textsuperscript{9}), which are associated with inflammatory processes. Other cases that involve increased cell death include trauma, stroke, myocardial infarction, angina,\textsuperscript{11} sepsis, and septic shock.\textsuperscript{12} Furthermore, overtrained athletes\textsuperscript{13} also show increased amounts of circulating DNA, whereas fetal DNA can be detected during pregnancy.\textsuperscript{14} Can the fact that circulating DNA occurs in increased amounts in so many different conditions point to some correlation between them that may reveal a common mechanism of release or origin?

Many papers report that oncogene mutations and amplifications, microsatellite alterations, and epigenetic changes like DNA methylation (see Refs. 1, 6, 15, and 15 for review) can be found in the circulating DNA similar to that found in tumor tissue of cancer patients. These resemblances suggest that the circulating DNA is most probably derived from the primary tumor or from mature tumor cells.\textsuperscript{3,10,17,18} A direct relationship could not be demonstrated between the amount of plasma DNA and the type or clinical status of the cancer.\textsuperscript{5}

Two main points of view exist in the literature for explaining the origin of circulating free DNA, that is, these DNA fragments enter the bloodstream following cell death or they are actively released by living cells.\textsuperscript{1,4-6,12,19-29}

### Apoptosis and Necrosis: DNA Release after Cell Death

Apoptosis and necrosis are two distinct mechanisms of cell death and represent two extremes of this phenomenon.\textsuperscript{24} During apoptosis, DNA degradation often occurs: chromosomal DNA is first cleaved into large fragments of 50–300 kb and subsequently into multiples of nucleosomal units of 180–200 bp, which is a hallmark of apoptosis.\textsuperscript{23-27} Because this ladder pattern is also visible after electrophoresis of circulating DNA, many believe that apoptosis may be the source of the observed DNA fragments in the plasma.\textsuperscript{5,22,28} Although the mechanisms are not fully understood, the contents of cells dying by apoptosis are rapidly ingested by professional phagocytes (macrophages and dendritic cells) or neighboring cells,\textsuperscript{20} and the DNA is consequently completely digested into nucleotides by DNase II in lysosomes.\textsuperscript{25,27} Thus the possibility exists that DNA fragments released by apoptosis are completely removed before they can appear in the circulation.\textsuperscript{5,22} If this engulfment of apoptotic bodies is impaired or cell death is amplified, tissue injury or autoimmunity will most probably result.\textsuperscript{22,29,30}

Macrophages can be implicated in the generation of circulating DNA as described below. After massive macrophage apoptosis induced by clodronate liposome treatment in mice, a dramatic increase of circulating DNA occurred. The intraperitoneal administration of dead, apoptotic or necrotic, Jurkat cells into mice lacking macrophages caused no further increase in the amount of circulating DNA.\textsuperscript{22} However, the administration of the same dead cells into normal mice caused an increase in circulating DNA in the blood of the mice and the characteristic ladder pattern could be observed after electrophoresis for both necrotic and apoptotic cells, respectively. This may indicate that necrotic human cells are engulfed by.
mouse macrophages or that DNA from necrotic cells is cleaved by the same enzymes functioning in apoptosis, thus causing the same ladder pattern as apoptotic cells. PCR analysis of circulating DNA in the blood of these mice showed the presence of both human and murine sequences. It is possible that macrophage apoptosis is caused by engulfment of high numbers of dead cells or by impaired phagocytic function, thus causing release of murine circulating DNA into the blood of these mice. It would have been interesting to know what the levels of circulating murine DNA were before administration of the dead human cells, and if this level stayed the same or increased after macrophage engulfment of the dead cells.

Most proliferating cells lost the ability to become apoptotic, and several ingenious mechanisms have been identified in which cancer cells become resistant to apoptosis in order to escape the immune system. Various targets for therapeutic intervention in cancer have been explored and many of them are thought to proceed via the preferential induction of apoptosis to eliminate cancer cells without affecting normal cells (for review see Bremer et al.35).

In contrast to apoptosis, necrosis causes random, nonspecific, and incomplete digestion of DNA and thus a smear is observed in an electrophoresis gel. By inducing necrosis in cell cultures, Jahr et al.5 demonstrated that necrotic cells produce DNA fragments larger than ~10,000 bp.

If lysis of circulating cells was to be the origin of circulating DNA, many more dying circulating cells should have been present in the blood since the amount of DNA in the plasma undoubtedly exceeds the amount of such circulating cells, indicating that circulating DNA does not originate from circulating cells dying in the blood. Lysis of T lymphocytes was also examined, but it was shown that T lymphocytes are not the source of circulating DNA. Jahr et al. also tested the possibility that normal circulating DNA might originate from endothelial cells by using the methylation status of the endothelium-specific human gene SELE promoter, which is unmethylated in endothelial cells and hypermethylated in other cells, and they found that only a small contribution, if any, is made by endothelial cells of cancer patients.5

**DNA Release by Living Cells**

The possibility that DNA may be released by living cells was suggested by a number of reports, but convincing evidence does not exist to prove this hypothesis. It is quite astounding that even though Anker et al.2 realized more than 30 years ago that DNA can be actively released by cells, the mechanism of this active release process is still not elucidated.

Four lines of evidence to support the hypothesis that living cells release DNA were highlighted by Chen et al.23: (a) Instead of increased circulating DNA, which is expected if apoptosis is the mechanism of release, Leon et al.9 found circulating DNA to be significantly decreased in response to radiotherapy. This may be because of the inhibitory effect of radiation on the proliferation of the cancer cells and thus less DNA is released. (b) Even with no cells dying in culture, DNA is still observed in the supernatant, and the concentration increases proportional to the proliferation of cancer cells; this unpublished observation of Chen et al. agrees with Anker et al., who observed in 1975 that human blood lymphocytes actively release double-stranded DNA into their culture medium until a certain concentration is reached, no matter how long the incubation lasts, and that newly synthesized DNA is released preferentially. Work by Anker et al.9 on frog auricles corroborated this observation. Anker also stated that the quantities of DNA that were released were similar, irrespective of whether a quarter of the cells or none at all die, except for cancer cells, which can release more DNA than normal cells. This shows that cell death is not responsible for the free DNA observed in the plasma. Furthermore, Stroun et al.20 showed that the characteristic ladder pattern on an electrophoresis
gel can also be observed for actively released DNA. (c) In the early stages of cancer, when seemingly little cell death is occurring, circulating DNA may already be present in higher than normal levels. As the cancer burden increases, so does cell death; however, the amount of proliferating cancer cells and thus the DNA levels increased significantly because of the increased amount of proliferating cells and not because of the amount of cells that die. (d) Lymphocytes are not the only cells that spontaneously release DNA into culture media when stimulated; release may also occur during division of other cell types, which include normal and malignant cells in the body. Additional observations from work on frog auricles are that the transfer of auricles to a new medium caused renewed release to the same concentration as in the previous medium, purified frog DNA did not inhibit release, and damaged auricles did not yield more free DNA. Further evidence for preferential release of DNA by viable cells is given by Stroum et al., when they compared the proportion of Alu repeat sequences to the β-globin gene in serum and lymphocytes.

**Additional Sources of Circulating DNA**

On the basis of the observations made by Raptis et al. in the early 1980s, an exogenous source of free circulating DNA was excluded. However, these observations were shown to be wrong by the presence of viral DNA circulating in the plasma of some patients with cancer associated with viral infection, such as nasopharyngeal carcinoma, where Epstein–Barr virus DNA can be detected in 96% of cases, cervical cancer, where human papilloma virus DNA can be detected in 50% of cases, and hepatocellular carcinoma, where hepatitis B virus DNA can be detected.

Cells that have lost their nuclei but remain functional undergo a process termed dedenudation or terminal differentiation. According to Bischoff et al., this may be another source of circulating DNA, but many tumors do not express enough of the molecular or morphologic markers of the terminally differentiated state to prove the existence of this phenomenon in cancer. Thus terminal differentiation is unlikely to provide a significant contribution to the origin of circulating DNA.

**Clearance of Circulating DNA**

More than one mechanism may be responsible for the clearance of free DNA from plasma since Lo et al. observed that the clearance of fetal DNA after delivery occurs in different phases. First an initial rapid phase is observed, which is followed by a slower second phase. In most of the women in their study, all DNA was cleared 2 hours after delivery.

Regardless of the rapid clearance rate, with a mean half-life estimated to be at only 16.3 minutes, it is known that fetal DNA is present in large amounts in the maternal circulation during pregnancy. This means that fetal DNA must be released in large quantities to maintain the high concentration which is continuously detectable in the maternal circulation during pregnancy. Thus, the rate at which fetal DNA is released exceeds its clearance rate. The concentration of fetal DNA consequently provides an almost real-time picture of the interaction between DNA release and DNA clearance.

Free circulating DNAs found in healthy people, cancer patients, and organ transplant recipients most likely have the same clearance mechanism(s) as in maternal plasma. Therefore DNA will also be rapidly removed and display a real-time picture of release, which may be useful in monitoring disease and transplant efficiency.

Although the mechanism of clearance has not been elucidated, a few possibilities can be proposed. Many reports state that free DNA can be detected in urine and thus it can be expected that the kidneys may play a role in clearance. Animal studies also suggested that the liver, spleen, and kidneys may be responsible for removal of circulating DNA.
Lo et al.\textsuperscript{40} explored the possibility that plasma nuclease s may have a function, but showed that these nuclease s only have a partial role in the removal of circulating fetal DNA. Tamkovich et al.\textsuperscript{42} showed that neutral deoxyribonuclease I activity is inhibited in the blood of cancer patients, thus causing a higher concentration of circulating DNA, while DNase I activity could be detected in healthy patients. DNase I activity, then, may play a significant role in the clearance of circulating DNA.

Chelobanov et al.\textsuperscript{43} summarized the available data on various DNA binding proteins which were detected on the cell surface of many different cells and cell lines. It was suggested that these DNA binding proteins recognize and transport DNA across the plasma membrane into the cell for possible degradation to mononucleotides or transporation into the nucleus. Binding of DNA to the cell-surface receptors is pH- and temperature-dependent and can be inhibited by a number of substances. It was found that serum of SLE patients competitively inhibits binding of DNA, which may be the result of an increased amount of circulating DNA in SLE serum. One study observed between 810 and 2600 molecules of bound DNA per cell and another study observed expression of DNA receptors by 67\% of lymphocytes and 98\% of monocytes.\textsuperscript{43} Thus, depending on the rate of uptake by these cells and the amount of DNA bound to receptors, cells with surface receptors for DNA may contribute and possibly play a major role in the clearance of free circulating DNA.

The mechanism(s) whereby clearance of circulating DNA is achieved is currently still poorly understood,\textsuperscript{44} but binding and uptake by different cells as well as DNase activity in the blood may be the main mechanisms of clearance.

Other Conditions That May Give Rise to Circulating DNA

The concentration of circulating DNA in the plasma of patients after undergoing hemodialysis was shown to be significantly higher than before undergoing hemodialysis or compared to controls, and the typical DNA ladder pattern associated with apoptosis was observed in agarose electrophoresis for the circulating DNA isolated from patients after undergoing hemodialysis.\textsuperscript{28}

Patients with untreated active systemic lupus erythematosus have much higher concentrations of circulating DNA in their plasma than do healthy individuals, but this decreases to normal levels after treatment.\textsuperscript{36} Organ rejection also caused an increase in the amount of circulating DNA, in the urine in this case, and after treatment the amount of free DNA rapidly decreased.\textsuperscript{41} When comparing plasma DNA concentrations in healthy individuals to those in patients having received bone marrow transplants, no significant difference could be found, but when the total circulating DNA was split into two fractions—originating either from the bone marrow or from the rest of the body—it was found that a significantly higher concentration originated from the bone marrow.\textsuperscript{45}

Chang et al.\textsuperscript{46} observed a 10-fold increase relative to controls in the amount of circulating DNA in patients suffering from myocardial infarction. They attribute this to widespread apoptosis followed by necrosis in the infarct. Comparison of the electrophoretic pattern of circulating DNA from healthy individuals, cancer patients, and patients who suffered a myocardial infarction showed a more diffused ladder pattern for the latter.\textsuperscript{46} However, it is possible that this may be an artifact of the study, as all the samples were not analyzed on the same gel.

The presence of fetal DNA in the maternal circulation was demonstrated in 1997 by Lo et al.\textsuperscript{40} Furthermore, it was found that fetal DNA increases with gestational age, and a sharp increase can be observed during the last 8 weeks of pregnancy.\textsuperscript{14,47} In case of complications during pregnancy, such as pre-eclampsia, even more fetal DNA is present, which may be because of impaired clearance or some form of cell injury or placent al breakdown.\textsuperscript{7} An
interesting observation is that circulating DNA in pregnant women has a much wider size distribution than circulating DNA in nonpregnant women. According to Bianchi et al., the majority of fetal DNA during pregnancy originates from the placenta. Free DNA in the fetus may also originate in the same manner as in normal individuals, and this fetal free DNA will cross the placenta to be released in the maternal blood, causing the maternal blood to contain both maternal free DNA as well as fetal free DNA.

It has been shown that exercise overtraining can cause increased amounts of plasma DNA, which can be related to the training load; it can increase 9– to 17.5-fold after long distance running and remain increased even after 96 hours. An increase in oxidative stress was also observed after exercise overtraining. Can the exercise-induced increase in DNA-damaging reactive oxygen species (ROS) be implicated in the production of circulating DNA?

A highly significant difference between the concentration of plasma circulating DNA was found between healthy individuals and those with minor or moderate trauma and major trauma early after injury, and it was found that patients with adverse outcomes, including death, had much higher plasma DNA concentrations than did those who did not develop complications, showing that DNA may be a valuable prognostic marker in trauma patients.

Circulating plasma DNA concentrations were shown to be increased in the first 24 hours after acute stroke, and the amount measured in patients within the first 3 hours after the event was 5-fold higher in those who died than in those who survived. In general it appears that higher circulating DNA concentrations are present in patients with more dramatic clinical presentations, suggesting that it may be a useful indicator for predicting disability and mortality in stroke patients. Although the origin of the circulating DNA in stroke is unknown, ROS are also produced and may be involved in the generation of free circulating DNA, as alluded to above.

Patients with severe sepsis or septic shock had significantly higher circulating plasma DNA concentrations than in normal controls, and even higher concentrations were found in those who did not survive intensive care and those who needed renal or isotropic support within the first 24 hours. When the concentration of circulating DNA was used as a predictor of intensive care survival, a sensitivity of 92% and specificity of 80% was observed, again suggesting that it can be used as a prognostic marker of mortality and sepsis in intensive care patients.

Conclusions

Almost every paper on circulating DNA states that apoptosis and/or necrosis is the source of free circulating DNA in serum and plasma. Many investigators use the fact that a ladder pattern is evident after electrophoresis of free circulating DNA as proof that apoptosis is the source of these fragments. However, this ladder pattern can also be found for the DNA in the culture medium in which lymphocytes grow. Furthermore apoptotic cells are ingested by macrophages and their DNA is digested into nucleotides. If macrophage ingestion fails on a scale large enough to produce the amount of circulating DNA in the blood, inflammation would definitely be a problem, and autoimmunity would occur frequently in cancer and the other conditions mentioned. The fact that many cancer cells are resistant to apoptosis argues against the notion of it as a mechanism for generating circulating DNA. Radiotherapy or irradiation, chemotherapy, and other cancer treatments cause cell death by apoptosis, and the amount of circulating DNA is less in cancer patients under treatment than it is in those patients before treatment, also disproving apoptosis as a source of circulating DNA. Necrosis, on the other hand, produces large DNA fragments, and the ensuing inflammation would
also be a problem if this were to be a source of large amounts of circulating DNA. Thus, we conclude that apoptosis and necrosis are not the main source of circulating DNA in the blood, although it may play a contributing role.

The possibility that circulating DNA may be liberated by living cells was already observed in the late 1970s, and evidence that DNA is released in vitro by human blood lymphocytes was presented in the mid 1970s. To our knowledge it has not been proven that DNA can be released into the circulatory system in vivo by living cells, but we do not anticipate a reason why it is not a possibility. Additionally, even though the mechanism by which clearance of DNA from plasma is achieved is poorly understood, and only a few papers address this issue, the appearance of increasing amounts of circulating DNA in the blood may be because the equilibrium between the release of DNA by living cells and the clearance of DNA is disturbed by an adverse condition. The low concentration of circulating DNA in the blood of normal individuals may thus be due to a lower rate of DNA release by cells, or a rapid removal of DNA by the optimal functioning of clearance mechanisms, and as soon as this equilibrium is disturbed, an increased amount of circulating DNA can be observed in the blood of an individual.

Circulating DNA can be found in a variety of conditions and, even though these conditions are unrelated, the presence of circulating nucleic acids is a common feature and thus some kind of correlation ought to be found that may point to a similar mechanism of origin. Although researchers have been looking for the origin of circulating DNA for more than 30 years, and quite a few possibilities have been explored, the mechanism of release still has to be elucidated. The possibility that more than one mechanism may be involved is feasible, but the factors influencing their relative contribution and the interaction between the mechanisms need to be understood for optimal utilization of this very valuable, noninvasive prediction and prognostic marker.

More work needs to be done to determine the mechanism of clearance and the mechanism cells use to release DNA, as well as to understand the significance of and the effect that these circulating DNA fragments have in the body. In-depth comparative characterization of circulating DNA obtained from patients with various pathologic conditions or compromised cells may also be useful in elucidating the origin of these molecules.

Conflicts of Interest

The authors declare no conflicts of interest.

References


The Origin of Circulating Free DNA

To the Editor:

Almost every report on circulating DNA identifies apoptosis or necrosis or both as the main source of free circulating DNA in serum and plasma. A hallmark of apoptosis is DNA degradation, in which chromosomal DNA is 1st cleaved into large fragments (~50–300 kb) and subsequently into multiples of nucleosomal units (180–200 bp) (1). This ladder pattern is also visible after electrophoresis of circulating DNA and is frequently considered to be evidence that apoptosis may be the source of the observed DNA fragments in plasma (2,3). It has been shown, however, that the characteristic ladder pattern can also be observed for actively released DNA (3).

The contents of apoptotic cells are rapidly ingested by professional phagocytes or neighboring cells through mechanisms that are not fully understood (4), and the DNA is consequently completely digested by DNase II in lysosomes (1). Thus the possibility exists that DNA fragments released by apoptosis are removed before appearing in the circulation. If this engulfment of apoptotic bodies is impaired or cell death is increased enough to produce substantial amounts of circulating DNA, inflammation would definitely be a problem and autoimmunity would occur frequently in cancer and other conditions involving increased circulating DNA (1,4).

Radiotherapy, chemotherapy, and other cancer treatments cause cell death by apoptosis, and less circulating DNA is found in cancer patients after treatment than before treatment, possibly because of the inhibitory effect of treatment on the proliferation of cancer cells. Furthermore, in the early stages of cancer, when little cell death seems to occur, circulating DNA may already be present in higher than normal concentrations. As the cancer burden increases, so does the rate of cell death and the amount of proliferating cancer cells, with a concomitant increase in circulating DNA. In addition, many cancer cells are resistant to apoptosis, and to escape the immune system proliferating cells lose the ability to become apoptotic (2,3), a process that also runs counter to the notion of apoptosis as the main mechanism for generating free DNA. Necrosis, on the other hand, produces large DNA fragments.

Even with no cells dying, the DNA concentration in culture medium increases in proportion to the proliferation of cultured cancer cells. Human lymphocytes have also been observed to actively release double-stranded DNA into culture medium to a certain concentration (5), irrespective of incubation time. A similar observation was made in experiments with frog auriules: DNA was released to the same concentrations during successive transfer of auriules to fresh medium, purified frog DNA did not inhibit release of DNA, and damaged auriules did not yield more DNA into the medium (5). Therefore, quantities of released DNA are similar regardless of the proportion of dying cells (5), with the exception applicable for cancer cells, which can release more DNA than normal cells (2). Evidence for preferential release of DNA was found by comparing the proportion of Alu repeat sequences to β-globin in serum and lymphocytes (5). We thus conclude that apoptosis and necrosis are not the main source of circulating DNA, although they may contribute. DNA released by living cells is a viable alternative.

The mechanism of DNA clearance from plasma is poorly understood. Increased amounts of circulating DNA in the blood of patients may reflect disturbance of the equilibrium between the release of DNA by living cells and the clearance of DNA. The low concentrations of circulating DNA in healthy individuals may reflect a lower rate of DNA release by cells or a rapid removal of DNA by the optimal functioning of clearance mechanisms, and when this equilibrium is disturbed, the amount of circulating DNA increases.

Circulating DNA can be found in a variety of conditions, and although these conditions are unrelated, the presence of circulating nucleic acids is a common feature, and thus some kind of correlation ought to be found that may point to a common mechanism of origin. Although researchers have been studying the origin of circulating DNA for more than 30 years, the mechanism of release still has to be elucidated. A reasonable possibility is that more than one mechanism is involved; if so, the variables influencing the relative contribution and interactions between the mechanisms must be understood for optimal utilization of this very valuable, minimally invasive biomarker. The more work needs to be done to determine the mechanism(s) of release and clearance as well as the significance of circulating DNA in the body.

Grant/funding support: None declared. Financial disclosures: None declared. Acknowledgments: We thank Dr. Phyllis Lebes for proofreading the final manuscript.

References


Maniesh van der Vaart
Piet J. Pretorius*

School of Biochemistry
North-West University
Potchefstroom Campus
Potchefstroom, South Africa

* Address correspondence to this author at School of Biochemistry, North-West University, Potchefstroom Campus, 11 Hoffmann St, Potchefstroom 2531, South Africa. Fax 27(0)18-299-2316; e-mail piet.pretorius@nwu.ac.za.
Chapter 3

Circulating DNA as a biomarker, what is currently being done?

Many scientists have investigated the possibility to exploit free circulating DNA as a diagnostic and/or prognostic tool, however, the majority of the publications in this field report only on the general quantification or detection of tumor related changes in these nucleic acid molecules. Although a number of different genes have been identified in this way, mostly by PCR analysis, it does not provide a solid contribution in elucidating the significance, function or properties of free circulating DNA. Only a very limited number of papers report on research done to elucidate the structure and/or function of free circulating nucleic acids (Diehl et al., 2005, Suzuki et al., 2008).

Several observations regarding circulating free DNA suggest that it has great potential as a source of biomarkers for cancer or other pathological conditions: (i) it can be obtained in a relatively non invasive way, (ii) part of it originates from the tumor tissue and thus show tumor specific changes, (iii) it is cleared rapidly from the blood (Frattini et al., 2008) which also make it applicable as a prognostic and monitoring marker and (iv) current protein based biomarkers for cancer lack the necessary sensitivity and specificity (Board et al., 2007). DNA markers are also qualitative markers, thus they are either absent or present, while protein markers are quantitative markers (Anker and Stroun, 2001b).

Cancer may be characterized by an accumulation of genetic and epigenetic changes (Anker and Stroun, 2001b, Bremnes et al., 2005). Genetic changes are changes on chromosomal level like oncogene or tumor suppressor gene mutations, microsatellite alterations, gene rearrangements, decreased strand stability, insertions, deletions and allelic losses. Epigenetic changes include modification of DNA and chromosomes but not on sequence level such as aberrant promoter hypermethylation (Anker et al., 1999, Bremnes et al., 2005). Furthermore, it has been shown that cancer patients have more free circulating DNA in their blood than normal individuals this observation forms the basis of the efforts to exploit this nucleic acid as diagnostic and prognostic marker (Goebel et al., 2005).

A review paper was compiled with the aim to compare (i) the yield of circulating DNA obtained by a number of studies in which circulating DNA in plasma and/or serum was quantified and (ii) the frequencies in which different mutations and (iii) methylated gene promoters could be detected in circulating DNA of mainly cancer patients and healthy individuals. The wide variety of methods used in these studies was highlighted and the current value of circulating DNA as a
biomarker for disease was evaluated. Although this is not a comprehensive review on the subject, the point we were trying to make is clear, circulating DNA is presently overrated as a prognostic or predictive marker while to little is known about the origin, function or significance of these circulating nucleic acid molecules. We propose total sequencing of circulating DNA as a valuable starting point in the search for much needed biomarkers for cancer, in this way the true potential of circulating DNA as a biomarker can be thoroughly examined.

3.1 Paper accepted for publication in Clinical Biochemistry

Maniesh van der Vaart and Piet J. Pretorius. 2009. Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated? (Manuscript number: CLB-D-09-00153R2)

See supplementary material (Section 6, page 82) for the supplementary table that forms part of this paper.
Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?

Manesh van der Vaart* and Pieter, J. Pretorius

*Biochemistry Division, School for Physical and Chemical Sciences, North-West University, Potchefstroom Campus, South-Africa
*Corresponding author, Tel: +27(0)84 840 0646, E-mail: Manesh.vanderVaart@mwu.ac.za, Fax: +27(0)19 299 2819

Abstract

Background: Circulating DNA is utilized widely as a genetic biomarker in a variety of pathological conditions, mainly in cancerous conditions. Quantification of circulating DNA and identifying the frequencies of a variety of mutations, microsatellite alterations and gene promoter methylation are the main foci of research on circulating DNA.

Content: A compilation of research reports available to us were reviewed to highlight the rather great variety of methods presently used to isolate circulating DNA, the lack of uniformity in presenting and interpreting quantitative research data as well as the virtual absence of information regarding the structure and function of circulating DNA.

Conclusions: The information compelled us to conclude that the application of circulating DNA as an unambiguous biomarker is currently overrated. We therefore emphasize the need for elucidating the prevailing questions regarding the origin, function and significance of these nucleic acid molecules before utilizing circulating DNA as a biomarker.

Key words: circulating DNA, plasma DNA, mutation detection, DNA methylation, DNA sequencing, biomarker, cancer detection

Introduction

Circulating DNA is small fragments of genomic DNA present in the blood of humans and other vertebrates. These nucleic acids are being studied with great expectations as potential biomarkers for cancer and other pathologic conditions. Several observations suggest its potential: firstly, the amount of circulating DNA in plasma or serum of cancer patients and patients with various ailments is more than in healthy individuals. Secondly, alterations that can be detected in primary tumors can also be detected in circulating DNA of a cancer patient. Thirdly, promoter hypermethylation of specific cancer related genes was detected in circulating DNA of cancer patients but not in healthy individuals. Furthermore, circulating DNA is cleared rapidly from blood, providing a real-time picture of the released fragments and finally it can be obtained in a relatively noninvasive manner. Despite these attributes “creating this potential” of circulating DNA as putative biomarkers, we would like to pursue the possibility whether this excitement is not premature, because one is struck by the diversity in the methods of isolating and quantifying circulating DNA and by the substantial void in information about its origin, structure and function.

In this review we only included studies analyzing circulating DNA from plasma or serum even though it can be detected in other body fluids like urine, bronchial lavage fluid, breast milk, sputum, etc. (1). In addition, we do not claim this to be a comprehensive review of the published data but rather an abstract from the literature.

Isolation and quantification of circulating DNA

A variety of methodologies are applied for isolation and quantification of circulating DNA, i.e., magnetic bead-based systems (KINGFisher or Roche), glass-milk-based methods, nucleosip blood kit (Machery-Nagel), phenol-chloroform extraction, PureGene DNA isolation kit (Genta Systems) and various Qiagen kits of which the QIAamp DNA blood mini kit was used most often. A similar variety of DNA quantification methods is utilized, e.g. DipStick (invitrogen), capillary zone electrophoresis, competitive PCR, blood direct PCR, densitometric scanning and video image analysis of agarose and polyacrylamide gels, Nanodrop ND-1000, fourometry with SYBR-green, Hoechst dye or pico-green and real-time PCR. Different targets were quantified with real-time PCR as a measure of the total amount of circulating DNA, for example a part of the AAT gene, Alu sequences, Bcr-abl marker, ERV-3, glyceraldehyde-3-phosphate dehydrogenase gene, human telomerase reverse transcriptase gene (hTert), β-actin and β-globin. The latter is the most popular. A comparison of DNA yields reported reveals a marked inter-experimental variation in terms of reproducibility and sensitivity (Figure 1, see also supplementary material). This is illustrated in Figure 1: Reported yields of circulating DNA in plasma range between 0 and 128 ng/ml (min and max, means: 2.5 and 27) with an average of 15 ng/ml isolated from healthy individuals and between 0 and 4738 ng/ml (min and max, means: 4.8 and 383.5) with an average of 157 ng/ml for cancer patients2. Average values for serum are 104 and 467 ng/ml respectively.

---

*Minimum and maximum values and their medians as combined between the papers reviewed, averages calculated from all averages reported by these papers (see supplementary material for a comprehensive table of all variables used).
Only a few reports compared different extraction methods in an effort to define a robust procedure for isolating circulating DNA (2-5). The most recent of these compared four isolation methods: QIAamp DNA Mini blood kit (Qiagen), Gentford blood and serum genomic DNA isolation kit (Agencourt), QIAamp virus spin kit (Qiagen) and a ChargeSwitch gDNA serum kit (Invitrogen). The QIAamp viral spin kit produced the highest yields of circulating DNA from both plasma and serum (5).

Blood processing methods and storage conditions may also influence the yield of circulating DNA. Prolonged storage of whole plasma as well as isolated plasma DNA leads to substantial DNA degradation, with an annual degradation rate of 30% (6). Other studies found no significant difference in yield whether plasma samples were stored at -20°C for 1 month (7) or -80°C for 2 weeks, as long as the plasma was subjected to a second centrifugation at 16000g for 10 minutes before or after storage. The number of freeze thaw cycles also did not affect the DNA yield, but it did lead to fragmentation of DNA (8). Herrera et al. (9) obtained a much higher yield from banked samples (stored at -80°C) than from fresh samples, they reported no differences in the DNA yield between healthy individuals and cancer patients.

The amount of circulating DNA in serum can be 2 to 24 times higher than in plasma (10-12). When the difference in yield between serum and plasma was evaluated it became apparent that in the case of serum not only consistently higher levels of DNA were detected but also a markedly larger degree of variation in the yield between patients. Plasma samples are less likely to suffer cellular contamination and show less day-to-day variability (6). The higher levels of circulating DNA found in serum of both healthy individuals and cancer patients may be due to cell lysis during the clotting process (5, 8, 13). Comparing DNA yield from serum obtained from blood that was processed either after 2 or 24 hours verified this cell lysis theory since much larger quantities was detected in the 24 hour samples (5).

Two studies reported contradictory results when plasma was processed after 24 hours, in one study the amount of DNA was not significantly altered compared to plasma that was processed promptly (5), while in the other a significant increase in DNA yield was observed (8). The notion that plasma should be processed sooner rather than later, preferably within 6 hours after collection, is supported by a number of studies in which an increase in DNA yield was observed after delayed processing of aliquots of the same sample (8, 11, 13). Differences in the amount of circulating DNA between males and females, normally a higher circulating DNA as biomarker prematurely overrated? yield from males, were reported (14, 15), though not in all cases (15).

Several studies reported a significant difference in the amount of plasma DNA isolated from healthy individuals, benign disease patients and cancer patients (5, 17, 13), while others showed that quantification of circulating DNA is not useful as a diagnostic tool for prostate cancer: the amount of circulating DNA in benign prostate hyperplasia was found to be higher than in prostate cancer (16, 19). Another study observed greater variability of DNA concentration in lung cancer patients than in controls and a higher risk for NSCLC could also be associated with elevated amounts of plasma DNA (20). Although Gornik et al. (21) observed variability in the amount of circulating DNA, they could detect a significant difference in the amount of circulating DNA in the serum of healthy individuals or patients with mild pancreatitis and patients with severe chronic pancreatitis. Furthermore, they concluded that the amount of circulating DNA is a superior early marker for pancreatitis severity than Ranson's score, APACHE II (acute physiology and chronic health evaluation) score and C-reactive protein, the markers currently in use (21).

Frattini et al. (14) observed that the level of plasma DNA can be affected by the location of the primary tumor in CRC patients, they also found that patients with colon cancer had a significantly higher concentration of circulating DNA than patients with rectal cancer. They contribute it to the difference in vascularity between the two cancer types. Furthermore, the amount of circulating DNA in the blood of 77% of patients four months after surgery was lower than at the time of surgery, this amount decreased progressively in relapse free patients until reaching similar levels as in healthy individuals, Sozzi et al. (20) made similar observations. On the other hand, in relapsed patients an initial decrease was observed but the plasma DNA levels rapidly increased after the adverse event (14).

The confusion in the use of fluctuations in the amount of circulating DNA as a biomarker in or between pathological conditions is exemplified by the following two reports. Swisher et al. (22) concluded that simply quantifying plasma DNA cannot predict survival in patients with ovarian cancer, while Frattini et al. (14) concluded that quantification of plasma DNA can confirm the presence of CRC and disease free status or relapse after surgery.

In summary, many factors may have influenced the yield of circulating DNA reported in different studies: the isolation and quantification methods (probably the most important variables), the blood processing methods including the time elapsed till the isolation procedure is started, the centrifugation conditions and whether serum or plasma is used.
Figure 1: Yields of circulating DNA compiled from published studies. Dashed lines (---) indicate healthy individuals, solid lines (---) indicate individuals with cancer, dotted lines (-----) indicate that the individual was not classified as healthy but they also didn’t have cancer (for example pregnancy, benign disease, high volume smoking, etc.). Lines indicate the range between minimum and maximum values (for studies where it was supplied), grey lines indicate standard deviations, circles indicate mean values and triangles indicate median values, empty centres (○ △) indicate that serum was used for analysis. Data was sorted by date.
In the graphical representation (Figure 1) the yields of circulating DNA reported in the literature are illustrated to emphasize the differences and inconsistencies that hinder comparison of different data sets. When data from individual studies are compared (Figure 1) higher DNA yields are generally obtained for cancer patients than for control subjects. However, when comparing data from a collection of studies, these differences become rather more of a trend. Nevertheless, it still seems as if the amount of circulating DNA isolated from healthy individuals tends to be lower than in cancer patients. This may reflect the in vivo situation. However, before this observation can be exploited as a biomarker for cancer a major proviso becomes crucial, i.e. the universal standardization of the methodology and the identification of all the factors that can possibly influence the amount of circulating DNA in the blood, e.g. non malignant diseases, heavy smoking, pregnancy, exercise, heart dysfunction, etc. This is necessary because it can have a major impact on the concentration of circulating DNA and may cause false positive/negative results as is evident for heavy smoking in the study of Sozzi et al. (6) in Figure 1. It is also clear from this compilation of published data, that, when comparing plasma and serum samples, some overlap is apparent between the amount of DNA obtained from plasma of healthy individuals and from cancer patients. This overlap for serum samples is more obvious (see [28, 38] in Figure1) which makes it quite difficult to distinguish between healthy individuals and cancer patients. In addition, it is unfortunate that most studies working with serum did not give the range of observed yields. Considering the variability in serum samples, the possible contamination from cells during the clotting process and also the difficulty to distinguish between healthy and diseased states is taken into consideration, it is quite safe to recommend the use of plasma as source of circulating DNA rather than serum. This pertains to all circulating DNA analyses and in particular for mutation analysis, since the lower levels of background wild-type DNA in plasma would have a beneficial effect on detection of mutated DNA (5). The format in which data is presented should also be uniform in order for data to be reliably compared, some studies just give a mean concentration (see [18, 36] in Figure 1), while others report only the range of yields they detected (see [3] in Figure 1). For this reason we recommend that all data should be reported, i.e. mean, median and range to allow meaningful comparison of data.

Because of the variation in the yields reported for circulating DNA in cancer patients in general, and circulating DNA as biomarker prematurely overrated? even in the same type of cancer, it is virtually impossible to establish a cutoff value for cancer diagnosis. A mere increase in circulating DNA without having knowledge of its origin and nature is too vague to be considered as an informative biomarker. Furthermore, as was alluded to above, there is a number of diverse factors that may influence the amount of circulating DNA in blood, necessitating the study of the kinetics maintaining the amount of circulating DNA (1). Taken together, it is suggested that quantification of circulating DNA per se cannot be used as an informative biomarker for cancer diagnosis or even screening. This situation can be resolved by thorough characterization of circulating DNA, information that may contribute significantly towards the identification of informative diagnostic and prognostic markers.

Circulating DNA with genetic alterations

Malignant cells are characteristically genomically unstable and the accumulation of genetic alterations was indeed observed in a variety of cancers (45). Detection of genetic alterations in circulating DNA isolated from plasma or serum of patients that also exhibit the same alterations in their primary tumor DNA, suggests that at least part of the circulating DNA originates from the tumor. This prompted the question whether circulating DNA could be used as an informative tumor marker. The first mutational analysis on plasma DNA was performed on the ras oncogenes which is the most common genetic alteration found in human malignancies (1, 46). Furthermore, mutations in the p53 and APC genes as well as microsatellite alterations like loss of heterozygosity (LOH) and microsatellite instability (MI) were detected in DNA isolated from plasma and serum (Table 1). These genetic alterations were investigated using a variety of methods, i.e. ligase detection reactions (LDR), PCR followed by sequencing of the amplified region, fluorescent PCR (FPCR), PCR-RFLP (restriction fragment length polymorphisms), digital PCR followed by direct sequencing of PCR products, BEAMing (the name derives from its principal components: beads, emulsion, amplification and magnetic), PCR-SSCP (single-strand conformational polymorphism analysis), mutant-enriched PCR (ME-PCR) and Gclamped PCR followed by temperature gradient gel electrophoresis (TGGE) in various types of cancer, e.g. cancer of the bladder (BladC), ovarian epithelium (ECC), breast (BC), colorectum (CRC), prostate (PCa), gastrointestinal (GI), lung (NSCLC and LC) and also in 129 healthy controls in total (Table 1).
Circulating DNA as biomarker prematurely overrated?

Table 1: Frequency of detection of mutations and microsatellite tumor markers (clinical sensitivity) in circulating and tumor tissue DNA.

<table>
<thead>
<tr>
<th>Genetic Marker</th>
<th>Cancer type (No. Patients)</th>
<th>Method used</th>
<th>Tumor tissue DNA</th>
<th>Plasma (P) / serum (S) DNA frequency of detection</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mutations</strong></td>
<td></td>
<td></td>
<td></td>
<td>Cancer patients</td>
<td>Normal individuals</td>
</tr>
<tr>
<td>K-ras</td>
<td>NSCLC (67)</td>
<td>PCR+seq</td>
<td>-</td>
<td>30% (S)</td>
<td>No controls</td>
</tr>
<tr>
<td>(Proto-oncogene</td>
<td>BladC (27)</td>
<td>PCR-SSCP+seq</td>
<td>7%</td>
<td>11% (P)</td>
<td>No Data MNC</td>
</tr>
<tr>
<td>(45))</td>
<td>NSCLC (35)</td>
<td>ME-PCR</td>
<td>31%</td>
<td>0% (P)</td>
<td>No controls</td>
</tr>
<tr>
<td>Crg18 of 70</td>
<td>BladC (27)</td>
<td>PCR-SSCP+seq</td>
<td>39%</td>
<td>39% (P)</td>
<td>No data</td>
</tr>
<tr>
<td>NSCLC (50)</td>
<td>ME-PCR+seq</td>
<td>18%</td>
<td>24% (S)</td>
<td>No controls</td>
<td>(49)</td>
</tr>
<tr>
<td><strong>PS3</strong></td>
<td>EOC (137)</td>
<td>PCR+seq+Ldr</td>
<td>50%</td>
<td>30% (P,S)*</td>
<td>No controls</td>
</tr>
<tr>
<td>(tumor suppressor (45))</td>
<td>BladC (27)</td>
<td>PCR-SSCP+seq</td>
<td>4%</td>
<td>0% (P)</td>
<td>No Data MNC</td>
</tr>
<tr>
<td>BC (62)</td>
<td>PCR-SSCP+seq</td>
<td>11%</td>
<td>5% (P)</td>
<td>0%</td>
<td>(50)</td>
</tr>
<tr>
<td>BC (142)</td>
<td>PCR-SSCP+seq</td>
<td>No data</td>
<td>7% (P)</td>
<td>0%</td>
<td>(51)</td>
</tr>
<tr>
<td>Glit (25)</td>
<td>PCR-TGGE+seq</td>
<td>87%</td>
<td>79% (P)</td>
<td>37% (P) treatment</td>
<td>(52)</td>
</tr>
<tr>
<td><strong>APC</strong></td>
<td>CRC (33)</td>
<td>PCR+seq + BEAMing</td>
<td>59%</td>
<td>52% (P)</td>
<td>&lt;0.01%</td>
</tr>
<tr>
<td><strong>Microsatellite markers</strong></td>
<td></td>
<td></td>
<td></td>
<td>Cancer patients</td>
<td>Normal individuals</td>
</tr>
<tr>
<td>6 markers</td>
<td>BladC (36)</td>
<td>PCR-RFLP</td>
<td>29%</td>
<td>58% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>5 markers</td>
<td>NSCLC (35)</td>
<td>ME-PCR</td>
<td>31%</td>
<td>31% (P)</td>
<td>No controls</td>
</tr>
<tr>
<td>6 markers</td>
<td>BC (62)</td>
<td>PCR</td>
<td>73%</td>
<td>LOH 59% (P)</td>
<td>No Data MNC</td>
</tr>
<tr>
<td>5 markers</td>
<td>NSCLC (38 of 84)</td>
<td>fPCR</td>
<td>61%</td>
<td>LOH45% (P)*</td>
<td>0%</td>
</tr>
<tr>
<td>12 markers</td>
<td>BC (62)</td>
<td>PCR</td>
<td>61%</td>
<td>LOH61% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>15 markers</td>
<td>LC (34)</td>
<td>fPCR</td>
<td>83%</td>
<td>AL 88% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>56%</td>
<td>LOH 59% (P)</td>
<td>No Data MNC</td>
</tr>
<tr>
<td>10 markers</td>
<td>NSCLC (35)</td>
<td>fPCR</td>
<td>57%</td>
<td>31% (P)</td>
<td>No controls</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>82%</td>
<td>LOH 61% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>5 markers</td>
<td>NSCLC (38 of 84)</td>
<td>fPCR</td>
<td>57%</td>
<td>LOH 59% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>56%</td>
<td>LOH 59% (P)</td>
<td>No Data MNC</td>
</tr>
<tr>
<td>15 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>82%</td>
<td>LOH 61% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>61%</td>
<td>LOH45% (P)*</td>
<td>0%</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>83%</td>
<td>AL 88% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>15 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>56%</td>
<td>LOH 59% (P)</td>
<td>No Data MNC</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>82%</td>
<td>LOH 61% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>15 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>61%</td>
<td>LOH45% (P)*</td>
<td>0%</td>
</tr>
<tr>
<td>12 markers</td>
<td>BladC (27)</td>
<td>PCR</td>
<td>83%</td>
<td>AL 88% (P)</td>
<td>0%</td>
</tr>
</tbody>
</table>

*percentage of patients with mutations in plasma or serum DNA that also have the same mutation in the tumor. 
* Data for patients where the specific mutations could not be found in the tumor DNA were not analyzed.

[1] Controls were not healthy individuals but rather patients with benign prostatic hyperplasia (LOH 12%) or controls were patients that were found to be tumor free, but not healthy. 1/14 (7%) had 1 TP53 mutation.

- Indicates that the analysis were not done.

MNC, Mononuclear cells; LOH, Loss of heterozygosity; MI, Microsatellite instability; AI, Allelic imbalance; other abbreviations in text.

From this compilation of data it is evident that various types of cancer display different degrees of genetic alterations in tumors or plasma DNA (see Table 1) and when the frequencies of detection of a specific marker is considered, results of many studies are rather contradictory (53) and even confusing. The fact that one marker can be detected in plasma DNA of none of the cases for
one cancer and up to 78% of cases in another (Table 1) may indicate that P53 is only a good marker for some cancers. However, when a marker can be detected in the plasma DNA of either none or 30% of cases of the same cancer it may disqualify this marker. Thus, individual markers alone may not be useful for cancer detection. An equally disappointing finding is the observation that even with a panel of 12 microsatellite markers at most 88% of clinically diagnosed patients could be positively confirmed (29).

Another unsatisfactory observation emanating from Table 1 is that scoring of LOH differs markedly between studies. Some cases are based on a reduction of at least 30% in allele intensity, others on a signal reduction of 75% while others calculated the ratio of intensities of the two alleles, which had to be between <0.6 or >1.67 for LOH to be scored. Even reductions as low as >20% or >40% in fluorescence intensity was reported. Thus, consensus in the amount of reduction of allele intensity that is needed before LOH can be scored should be reached, but at this stage it is clear that the results obtained with these differing scoring parameters cannot be compared, which in effect eliminate microsatellite alteration comparison between the different studies. Another factor that complicates microsatellite comparison even more is the use of so many different markers (see Table 1). Each study use their own set of markers in order to find the ideal set that can be used for cancer prognosis. A structured approach has a better chance of success in finding an ideal biomarker panel. Furthermore, the effect of different cancers on the detection rate of the examined markers should also be investigated.

Ramírez et al. (49) detected more mutations in the K-ras gene in serum than in the primary tumor, other studies detected mutations in the K-ras gene in cancer-free patients which lead to the conclusion that K-ras is not an indicator of malignancy (1). Furthermore, the mutations in the tumor and serum or plasma DNA often did not match while other studies (indicated with a< in Table 1 & 2) did not examine the plasma of patients for which genetic alterations were not found in the tumor tissue.

To detect tumor related alterations like mutations in a background of normal molecules is a daunting task which compelled Miyamoto et al. (50) to conclude that it is not useful. Currently the use of mutation detection is held back by the fact that the significance of circulating DNA is unknown, when the origin and function of these molecules is elucidated it can possibly be used on more specific mutations or genetic alterations.

Circulating DNA as biomarker prematurely overrated?

Epigenetic alterations

Great strides have been made in using changes in the levels of methylation of the promoters of several cancer related genes as diagnostic and prognostic biomarkers, unfortunately the same cannot be said for circulating DNA. Between 1999 and 2008 numerous groups detected this epigenetic event in circulating DNA in many different cancer associated genes (Table 2) using a variety of methods, i.e. methylation specific PCR (MSP), quantitative MSP (QMS), fluorescent MSP (f-MSP), methylation specific restriction enzyme digestion (MRE PCR), Methylight, HeavyMethyl or variations of these methods in a variety of cancers. During this time the methylation status of various genes in the circulating DNA were determined in cancer of the nasopharynx (NPC), liver (HCC), lung (NSCLC), lymph (Lymph), cervix (Cerv), renal (RenC), colorectum (CRC), bladder (BladC) and breast (BC) cancer before treatment and in 259 healthy individuals in total (see Table 2).

Markers for methylated circulating DNA could only be found in patients where the corresponding gene was shown to be methylated in the primary tumor (60,61,65); however, the contrary was also seen when methylated DNA was observed in serum of three patients but not of the corresponding tumor DNA (66). This leads to the impression that, in some circumstances, circulating DNA may not be an early tumor marker or else the result would have been the opposite, thus the clinical sensitivity of methylation markers is not satisfying but the analytical sensitivity seems to be adequate (see Table 3 for an explanation). Some individual studies begun by analyzing the primary tumor for DNA methylation or the presence of mutations and when the result was negative or no tumor DNA was available no analysis of plasma or serum DNA of that patient ensued (14,22,34). In addition some studies presented their data according to the result found in the tumor tissue (27,48,60,65). This way of reporting is open to misinterpretation of data, for example from a total of 27 samples in only two plasma DNA samples the p16 gene promoter was methylated, this was presented as 40% (2 out of 5 tumor samples) positive. In fact it should have been stated that only 7% (2 out of 27) of plasma DNA samples were methylated (48). If DNA methylation, mutation detection or microsatellite analysis is to be employed as a tumor or prognostic marker the corresponding genes should preferably be analyzed in circulating DNA of the whole cohort of study subjects to eventually obtain a complete and informative picture.
Circulating DNA as biomarker prematurely overrated?

Table 2: Frequency of detection of methylated tumor DNA markers (clinical sensitivity) in circulating and tumor tissue DNA

<table>
<thead>
<tr>
<th>Methylation Marker</th>
<th>Cancer type (No. Patients)</th>
<th>Method used</th>
<th>Tumor tissue DNA</th>
<th>Plasma (P) / serum (S) DNA frequency of detection</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cancer patients</td>
<td>Normal individuals</td>
</tr>
<tr>
<td>P16 (Tumor suppressor gene)</td>
<td>HCC (22)</td>
<td>MSP</td>
<td>73%</td>
<td>59% (P+S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>NPC (41)</td>
<td>QMSP</td>
<td>-</td>
<td>42% (P)</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (105)</td>
<td>SMSP</td>
<td>75.3%</td>
<td>73% (P)</td>
<td>no controls</td>
</tr>
<tr>
<td></td>
<td>Lymph (37)</td>
<td>MRE PCR</td>
<td>38%</td>
<td>73% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>CerC (40)</td>
<td>MSP</td>
<td>28.2%</td>
<td>10% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>35%</td>
<td>22% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>HCC (45)</td>
<td>MSP</td>
<td>67%</td>
<td>53% (S/P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (22)</td>
<td>MSP</td>
<td>41%</td>
<td>14% (S)</td>
<td>0% nil</td>
</tr>
<tr>
<td></td>
<td>BladC (27)</td>
<td>MSP</td>
<td>18%</td>
<td>7% (P)</td>
<td>0% nmc</td>
</tr>
<tr>
<td></td>
<td>NSCLC (35)</td>
<td>MSP</td>
<td>57%/63%</td>
<td>6%/34% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>BC (43+35)</td>
<td>MRE PCR</td>
<td>23%</td>
<td>14% (P)</td>
<td>0% nmc + P</td>
</tr>
<tr>
<td></td>
<td>CRC (18 of 70)</td>
<td>F-MSP</td>
<td>61%</td>
<td>61% (P)</td>
<td>no data</td>
</tr>
<tr>
<td></td>
<td>NSCLC (91)</td>
<td>MSP</td>
<td>-</td>
<td>15% (S)</td>
<td>?</td>
</tr>
<tr>
<td>DAPK(1) (tumor suppressor (49), apoptosis associated (66))</td>
<td>NPC (41)</td>
<td>QMSP</td>
<td>-</td>
<td>20% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>CerC (40)</td>
<td>MSP</td>
<td>60%</td>
<td>40% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (50)</td>
<td>MSP</td>
<td>45%</td>
<td>40% (S)</td>
<td>No controls</td>
</tr>
<tr>
<td></td>
<td>NSCLC (22)</td>
<td>MSP</td>
<td>23%</td>
<td>18% (S)</td>
<td>0% nil</td>
</tr>
<tr>
<td></td>
<td>NSCLC (91)</td>
<td>MSP</td>
<td>-</td>
<td>11% (S)</td>
<td>?</td>
</tr>
<tr>
<td>MGMT (DNA repair gene (65), (66))</td>
<td>CerC (40)</td>
<td>MSP</td>
<td>18.8%</td>
<td>8% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>6%</td>
<td>9% (S)</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (22)</td>
<td>MSP</td>
<td>27%</td>
<td>18% (S)</td>
<td>0% nil</td>
</tr>
<tr>
<td></td>
<td>BC (33)</td>
<td>QMSP</td>
<td>-</td>
<td>6% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (91)</td>
<td>MSP</td>
<td>-</td>
<td>19% (S)</td>
<td>?</td>
</tr>
<tr>
<td>CDH1 (Involved in tumor invasion (1))</td>
<td>NPC (41)</td>
<td>QMSP</td>
<td>-</td>
<td>46% (P)</td>
<td>9%</td>
</tr>
<tr>
<td></td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>59%</td>
<td>33% (S)</td>
<td>6%</td>
</tr>
<tr>
<td>P15 (RNA repair)</td>
<td>NPC (41)</td>
<td>QMSP</td>
<td>-</td>
<td>20% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>RASSF1A (RNA repair and cell cycle repair (49), apoptosis associated (66))</td>
<td>NPC (41)</td>
<td>QMSP</td>
<td>-</td>
<td>5% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>88%</td>
<td>11% (S)</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (50)</td>
<td>MSP</td>
<td>34%</td>
<td>34% (S)</td>
<td>No controls</td>
</tr>
<tr>
<td></td>
<td>BC (33)</td>
<td>QMSP</td>
<td>-</td>
<td>15% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (91)</td>
<td>MSP</td>
<td>-</td>
<td>12% (S)</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>BC (20)</td>
<td>MSP</td>
<td>-</td>
<td>15% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>BC (50)</td>
<td>MethylLight</td>
<td>-</td>
<td>26% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>MLH1 (DNA repair)</td>
<td>NPC (41)</td>
<td>QMSP</td>
<td>-</td>
<td>0% (P)</td>
<td>0%</td>
</tr>
<tr>
<td>APC</td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>29%</td>
<td>6% (S)</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>BC (33)</td>
<td>QMSP</td>
<td>-</td>
<td>3% (S)</td>
<td>0%</td>
</tr>
<tr>
<td>p14ARF</td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>24%</td>
<td>6% (S)</td>
<td>3%</td>
</tr>
<tr>
<td></td>
<td>BladC (27)</td>
<td>MSP</td>
<td>56%</td>
<td>48% (P)</td>
<td>0% nmc</td>
</tr>
<tr>
<td>TMS1 (Resistance to cytotoxic drugs (49))</td>
<td>NSCLC (50)</td>
<td>MSP</td>
<td>35%</td>
<td>34% (S)</td>
<td>No controls</td>
</tr>
<tr>
<td>GSTP1 (Detoxification gene (65))</td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>12%</td>
<td>6% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>CRC (5)</td>
<td>Heavy/Methyl See*</td>
<td>100% (S)</td>
<td>0%</td>
<td>(69)</td>
</tr>
<tr>
<td></td>
<td>NSCLC (22)</td>
<td>MSP</td>
<td>9%</td>
<td>5% (S)</td>
<td>0% nil</td>
</tr>
<tr>
<td></td>
<td>PCa (12)</td>
<td>MSP</td>
<td>-</td>
<td>73% (P)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>PCa (19) therapy</td>
<td>MSP</td>
<td>-</td>
<td>36% (P)</td>
<td></td>
</tr>
<tr>
<td>RAR-β2 (Tumor suppressor (66))</td>
<td>RenC (18)</td>
<td>FQMSP</td>
<td>53%</td>
<td>6% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>BC (33)</td>
<td>QMSP</td>
<td>-</td>
<td>3% (S)</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>NSCLC (91)</td>
<td>MSP</td>
<td>-</td>
<td>7% (S)</td>
<td>?</td>
</tr>
<tr>
<td></td>
<td>BC (20)</td>
<td>MSP</td>
<td>-</td>
<td>15% (P)</td>
<td>0%</td>
</tr>
</tbody>
</table>
It is noticeable that promoter methylation of the genes of only 50% of P16 and GSTP1 was observed in more than 50% of cases analyzed and in almost none of the controls (see Table 2). According to this data many cancers will be missed if only one DNA methylation marker is used for diagnosis. This lack of sensitivity demonstrates the need for multivariate analysis. However, the highest detection rate in plasma or serum DNA, when combining a number of markers, was still only 77% (18). It is important to keep in mind that methylation of a certain gene promoter is not exclusive to one cancer, but may be methylated in various cancers, thus currently DNA methylation cannot really be used as a cancer specific marker but rather as a marker for cancer in general. Furthermore, normal cells may also have hypermethylated regulatory sequences as methylation is a normal physiological process and an altered methylation pattern is not necessarily indicative of tumor origin, some methylation changes can also be age-dependent. All these factors have to be taken into account and the methylation patterns of circulating DNA in normal individuals will have to be determined before methylation can be used as a tumor specific marker (1).

Discussion

Some studies reported no correlation in the presence of mutant oncogene or tumor suppressor genes in tumor tissue with decreased survival as is the case in the corresponding plasma or serum DNA (49). Also, no correlation was found between methylation of the genes coding for RASSF1A, DAPK and TMS1 in tumor or serum and survival, likewise, no correlation was found between methylation of the genes of p16, DAPK, GSTP1 and MGMT and early recurrence (55). Others found that detectable microsatellite alterations after surgery might indicate metastatic disease or recurrence (34), this was confirmed by the detection of methylation of the K-ras and p16 genes, 4-10 months after surgery. Disease free subjects had no detectable alterations while patients with recurring disease showed alterations in their plasma DNA (14). Dominguez et al. (48) also found that recurrence of disease was associated with p14ARF methylation or LOH only when the alteration could be detected in the plasma of the patient. Sozzi et al. (34) observed that all patients with recurrent disease or metastasis showed reappearance of high levels of circulating DNA.

Others go so far as to say that tumor derived circulating DNA is an independent predictor of poor prognosis (22). It was also indicated that tumors with circulating tumor DNA are more aggressive (22, 51) thus plasma tumor DNA can be considered as a negative prognostic factor (51). Muller et al. (70) showed that the presence of methylated genes of RASSF1A and/or APC in serum could be independently associated with poor outcome.

No correlation between detection of mutant K-ras in serum and disease free or overall survival (47), between tumor burden or stage and either the concentration of APC fragments or percentage of mutant APC fragments (53), between detection of six microsatellite alterations in plasma, serum or urine and tumor staging or tumor progression (54), also no correlation could be found between four microsatellite markers and the occurrence of LOH in serum or between cancer stage and serum DNA concentration (55). Higher levels of circulating DNA and time to progression or overall survival didn’t correlate, but a trend toward detectable plasma DNA alterations and poor survival could be found by Beau-Faller et al. (29). Analysis of 39 methylation markers in breast cancer using Methylation in a preliminary study (26 patients) and then selecting markers that either show no methylation in controls and >10% methylated samples in cancer patients or ≤10% in controls and >20% in cancer patients, only 5 of the 39 markers (ESR1, APC, HSD17B4, HIC1 and RASSF1A) met one of these criteria (70). Frequencies of methylation in the preliminary study was low and results were not reported for the total number (86) of patients, thus it was not included in Table 2, they (70) concluded that ESR1 methylation correlates with APC and RASSF1A methylation and
that RASSF1A and/or APC serum DNA methylation was strongly associated with poor outcome. Muller et al. (58) found that some of the markers they used were tumor specific while others showed LOH in benign disease. Some studies found no correlation between methylation of the genes they analyzed either in tumor or serum and survival (49).

Multivariate analysis in some of the studies compiled in Table 1 and 2 provided the following detection frequencies. Measuring methylation of the respective genes coding for p16, mutations of the p53 gene and six microsatellite markers, Silva et al. (50) could detect 90% cases with at least one molecular event in tumor DNA and 66% with a similar alteration in plasma DNA. Bearzatto et al. (27) obtained similar results by using methylation of the p16 gene, K-ras mutations and 5 microsatellite markers. Wong et al. (19) indicated that they could detect 71% of patients with NPC by detection of DNA methylation in at least one of five genes analyzed with 91% specificity and Yang et al. (62) used three DNA methylation markers and could detect at least one of the three markers in 75% of cases using tumor DNA and 55% of cases using plasma DNA. Hoque et al. (63) found that at least one of nine genes was methylated in 88% when urine and 67% of cases when serum was used and combined they could detect 94% of renal cancer cases which also showed methylation in tumor tissue with high specificity. Esteller et al. (65) could detect DNA methylation in one of five genes analyzed in 68% of tumors and in 50% of serum samples. Fujiwara et al. (66) could identify 51% of stage I lung cancer patients by methylation analysis while serum protein markers were only positive in 11% of these patients. Of the 91 lung cancer patients analyzed they could detect DNA methylation in one out of five genes in 50% of cases with 85% specificity. Beau-Faller et al. (29) used 12 microsatellite markers and was able to identify tumor DNA in the plasma of 88% of the lung cancer patients they analyzed. They also showed that even if no allelic imbalance could be detected in the tumor tissue DNA they could detect alterations in 93% of plasma DNA from the patients with confirmed lung cancer with 100% specificity. Muller et al. (59) detected only 34% of prostate cancer cases using 15 microsatellite markers. Papadopoulou et al. (37) found a combination of DNA concentration and promoter hypermethylation in plasma could detect 88% of prostate cancer cases and 54% of breast cancer cases with 100% specificity. Using all the markers they applied, Ramirez detected 75% of patients showing at least one alteration in tumor and 77% in serum, combined 82% had one or more alteration in tumor and/or serum (49). Taken together, it seems as if quite satisfying results were obtained with these circulating DNA as biomarker prematurely overrated? multivariate analyses, however, tumor material is not always guaranteed and would certainly not be available for screening, therefore detection of mutations in plasma gets paramount importance. Several of the studies report results obtained by analyzing tumor material obtained from selected patients only, which lead to artificially high detection rates and for this reason the results cannot be compared with confidence. Camps et al. (47) commented that many of the studies are also statistically underpowered.

Promising results were reported when plasma as well as total cell-surface-bound DNA were analyzed. Skvortsova et al. (68) analyzed RARB2 and RASSF1A as markers of DNA methylation in MSP and detected methylation in 95% of breast cancer patients, 60% in patients with fibroadenoma and no methylation in healthy controls. Miyamoto et al. (59) made the observation that no control cases to assess the specificity of 'aberrant' DNA methylation were analyzed in many studies and that false positive results may occur in traditional MSP. In the papers we reviewed, we found that quite a few studies did not analyze control patients or used invalid control materials like mononuclear blood cells or DNA from patients with benign disease, the number of controls compared to cancer patients was also significantly less. Furthermore, as mentioned previously, the large variety of methods used for DNA isolation and quantification and also mutation and methylation analysis varied widely, hampering comparison of results and in a way their scientific value.

The need for sensitive and specific prognostic indicators for cancer and other pathological conditions is obvious, ideally a marker must reflect tumor specific alterations in blood. Unfortunately, the arena of circulating DNA is currently a vast field of unknowns, even confusion. Comprehensive sequencing and characterization of circulating DNA would contribute toward alleviating this confusion and to fill the lack of knowledge on the origin and significance of circulating DNA.

Amplon sequencing was performed on four genomic regions (GHSC locus, MGA locus, NF1X locus, unannotated region on Chr 7) which were identified as aberrantly methylated at high frequency in breast cancer. Korshunova et al. (72) observed a great variety of methylated DNA, but no specific methylation pattern exclusive to cancer or normal DNA was observed in the sequenced DNA. They could distinguish between DNA from normal and tumor tissues but not between serum DNA from cancer patients and cancer-free individuals which suggests that tissue data cannot be readily extrapolated to serum (72). Another study was very successful in sequencing the GSTP1 gene promoter using 32P-labeled primers and the
Circulating DNA as biomarker prematurely overrated?

Acknowledgements
We thank Leonard Santana for writing the algorithm used to produce Figure 1 and the National Research Foundation of South Africa for financial support.

References
Circulating DNA as biomarker prematurely overrated?


Chapter 4

Sequencing of Free Circulating DNA

One of the few papers reporting on the sequencing of these DNA molecules in order to determine the composition of circulating DNA was written by Suzuki et al. (2008), however, a vital flaw in the reasoning of this article was pointed out to the editor of Clinica Chimica Acta.

4.1 Letter to the editor published in Clinica Chimica Acta


Letter to the editor

Characterization of circulating DNA in healthy human plasma

Dear editor,

In the article by Suzuki et al. [1] titled “Characterization of circulating DNA in healthy human plasma” they unconvincingly conclude that apoptosis is the main source of the circulating DNA detected. They substantiate their conclusion by their detection of plasma DNA with 5' and 3' ends rich in C and G, and additionally having 5' protruding ends (as indicated by Klenow labelling), using Nagata et al. [2] as a supporting reference. However, Nagata et al., explicitly concluded that caspase-activated DNase (CAD), the enzyme mainly responsible for cell-autonomous DNA degradation during apoptosis, predominantly cleaves at AT rich sites generating mostly blunt ends, indicated by TdT labelling [2]. It is unclear how Suzuki et al. come to the same conclusions, with opposite results. Additional arguments were proposed by us not supporting the apoptosis theory to be the main source of circulating DNA [3]. A salient feature of this reasoning is that apoptotic cells are swiftly engulfed by professional phagocytes, consequently digesting the DNA into nucleotides by DNase II in the lysosomes. Therefore, after complete apoptosis, no DNA fragments derived from apoptosis should be detectable in the blood [2]. Furthermore, the large number of unengulfed apoptotic cells required in order for detectable amounts of DNA to be seen in blood, would result in inflammation in healthy individuals and even autoimmunity, especially in conditions involving increased circulating DNA [4,5]. Hence, the theory of active release of circulating DNA by living cells is for a number of reasons a more viable alternative than apoptosis [3].

References


Maniesh van der Vaart
Piet J. Pretorius*

School for Biochemistry, North-West University, Potchefstroom Campus, Potchefstroom, South Africa
E-mail address: piet.pretorius@nwu.ac.za (P.J. Pretorius).

22 April 2008

2008-8815 – see front matter © 2008 Elsevier B.V. All rights reserved.

Unaware of the work of Suzuki et al., we also designed a method to sequence circulating DNA by using a clone based approach, at first only 65 clones were sequenced and analyzed as a prove of method publication.

4.2 Paper published in Annals of the New York Academy of Sciences

A Method for Characterization of Total Circulating DNA

Maniesh van der Vaart and Piet J. Pretorius

School of Biochemistry, North-West University, Potchefstroom Campus, Potchefstroom, South Africa

Although much work has been done in the field of circulating DNA, no definitive information on sequencing data of total circulating DNA is available. Characterization of total circulating DNA by sequence analysis may give valuable information about the origin and function of these nucleic acid molecules. Circulating DNA was isolated from plasma of one healthy individual and one cancer patient with various methods and was cloned into a blunt-cloning vector. Resulting colonies were sequenced and analyzed. The majority of the DNA that ligated into the vector was about 200 bp in length. Sequence analysis revealed that circulating DNA consists partly of currently uncharacterized human genomic sequences and when human repeats were masked it matched with sequences present in contigs containing known genes situated at various distances from the identified targets. In addition to the presence of large repeats, a variety of Alu repeat sequences were observed. Preliminary results showed that more Alu repeats are present in the plasma of normal individuals than in patient material. None of the gene sequences reported in the literature to be part of circulating DNA (e.g., P53, the Ras family, β-globin, or β-actin) was observed. Cloning and sequencing of free circulating DNA was successful and this first attempt on characterizing sequence data of free circulating DNA not only confirmed previous results, but also revealed a large variety of sequences. Further characterization of circulating DNA may be beneficial in diagnosis and prognosis and may also contribute to determining the source and function of circulating DNA.

Key words: free circulating DNA; cloning; sequencing; plasma.

Introduction

Circulating DNA comprises fragments of nucleic acids that circulate in the blood of healthy individuals, cancer patients, and patients with a number of other diseases, as well as during pregnancy. It is known that the first group has lower levels of these circulating nucleic acids compared to the other groups. Although the existence of free DNA was discovered even before Watson and Crick elucidated the structure of DNA, the origin and function of these molecules are still largely unknown.

Only a few papers have been published on the analysis of the sequences that make up total circulating DNA. PCR and real-time PCR revealed the presence of a number of genes such as P53, the Ras family, Ig Heavy chain, TP53, SRY and DYS1 (for test DNA), lamin B2, leptin, β-Globin, β-actin and Alu repeats, and a number of other polymorphic markers were used to detect alterations that are similar in tumor and plasma DNA. The composition of circulating DNA with regard to the sequence content and the variation in the amount of all or only specific sequences in malignant and other diseases compared to controls is largely unknown. Elucidating the sequence contents of total circulating DNA may contribute to determining the source and function of circulating DNA. This information in turn may be beneficial in...
diagnosis and prognosis of a variety of pathologic conditions.

In this pilot study we have developed a method to clone and sequence total circulating DNA in order to determine the sequence composition of circulating DNA in a healthy individual and a cancer patient.

**Methods**

**Blood Collection and Isolation of DNA**

Two 4-ml EDTA tubes of blood were drawn from one breast cancer patient and one control individual. Plasma was immediately separated from the cellular fraction by centrifugation at 1600 g for 20 min and the resulting supernatant (plasma) was stored at -20°C. Plasma was completely thawed at room temperature and subsequent centrifugation of the supernatant was done at 16,000 g for 10 min, after which circulating DNA isolation from the plasma of the normal individual was performed with a variety of methods: QIAamp DNA Blood Mini kit (Qiagen, Valencia, CA, USA), a salting-out method, phenol/chloroform extraction, and the MagNA Pure Compact system (Roche, Rotkreuz, Switzerland). Circulating DNA from the plasma of the cancer patient was isolated with the phenol/chloroform extraction method.

**Cloning and Sequencing**

Blunting of DNA was performed using T4 DNA Polymerase (Fermentas, Ontario, Canada), 10 μL circulating DNA and Tango buffer (Fermentas) with a final reaction volume of 20 μL. The reaction mixture was incubated for 5 min at 11°C, and dNTPs was then added and incubated for a further 15 min at room temperature, after which the enzyme was inactivated at 70°C for 10 min. DNA was phosphorylated by adding ATP (final concentration 20 pmol) and T4 polynucleotide kinase (Fermentas) followed by incubation at 37°C for 10 min; the enzyme was then heat-inactivated and a chloroform extraction was performed.

The blunt, phosphorylated DNA was ligated into the pEZSeq vector (Lucigen, Middleton, WI, USA) with subsequent heat inactivation of the ligase, after which PCR was performed to estimate the success of the ligation reaction (Fig. 1). Competent SURE (E. coli) cells were transformed with the vector by heat shock and after recovery the cells were plated on YT +Amp agar plates. Colonies containing inserts (Fig. 2) were sequenced by Inqaba Biotechnical Industries (South Africa).

**Sequence Analysis**

Sequences were analyzed with a number of web-based sequence analysis tools like Repbase CENSOR, EMBCSS suite, MEME/MAST (Multiple Em for Motif Elicitation/Motif Alignment & Search Tool), Repeat Masker, Sequence Manipulation Suite and Basic Local Alignment Search Tools (BLAST) from NCBI against various databases.

**Results and Discussion**

As shown in Figure 1, a rather small spectrum of DNA fragments ligated into the cloning vector, and the majority of the DNA fragments
that were cloned into the vector were about 200 bp (length of vector DNA attached to the circulating DNA is 110 bp); smaller amounts of larger and smaller fragments can also be seen. This correlates with results from other groups.3,4

Cloning of total circulating DNA was not as straightforward as was initially anticipated, possibly because the ends of the DNA fragments are capped or something is preventing the blunting enzyme from functioning. Blunting of the circulating DNA fragments was made possible by altering the protocol provided with the enzyme (Fermentas), that is, the incubation times and temperature, as well as the time of dNTP addition.

The different DNA isolation methods used to isolate the control individual’s circulating DNA yielded different numbers of clones (Table 1). In the case of DNA isolated with the QIAamp Blood Mini Kit and the MagNA Pure the lowest number of colonies was obtained, while many colonies were obtained in the case of the phenol/chloroform and salt precipitation extraction methods. The latter two isolation techniques most probably removed some of the above-mentioned blunting obstacles or much more circulating DNA was isolated with these methods.

Colony PCR performed on 65 of the colonies revealed different sizes of inserts, from ~40 to ~600 bp, which indicates that DNA fragments of different lengths were cloned. The results depicted in Figure 2 are a representation of the size distribution of the inserts of the clones that were sequenced.

The sequences obtained from these inserts were analyzed, but characterization of these sequences was not as straightforward as expected. BLAST gave two or more significant (E-value ≤ 1 × 10^-15) matches for most of the sequences analyzed, but the vast majority of these matches were not with known genes, but rather with sequences at various distances from the descriptive sequence(s) in a large contig. This is illustrated in Table 2.

Furthermore, a number of different Alu repeats were matched and some sequences contain other human repeats (Table 3). Interestingly, when the sequences were analyzed with CENSOR, the majority (more than 60%) of

![Figure 2. Products of colony PCR of pEZSeq transformants. Thirty-five colonies from a control and 30 from a cancer patient were analyzed.](image)

<table>
<thead>
<tr>
<th>TABLE 1. Comparison of the Different DNA Isolation Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolation method</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>Qiagen</td>
</tr>
<tr>
<td>MagNA Pure</td>
</tr>
<tr>
<td>Phenol/chloroform</td>
</tr>
<tr>
<td>Salt precipitation</td>
</tr>
</tbody>
</table>

*The number of colonies is a reflection of the DNA yield and cloning efficiency.
### TABLE 2. An Example of BLAST Results

<table>
<thead>
<tr>
<th>Sequence name</th>
<th>Hits</th>
<th>Description</th>
<th>E-value</th>
<th>Features flanking this part of subject sequence</th>
<th>Features in this part of subject sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MS1</td>
<td>2</td>
<td><em>Homo sapiens</em> chromosome 9 genomic contig, reference assembly</td>
<td>$2 \times 10^{-37}$</td>
<td>301409 bp at 5’ side: hypothetical protein 1417011 bp at 3’ side: transducin-like enhancer protein 1</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Homo sapiens</em> chromosome 9 genomic contig, alternate assembly</td>
<td>$2 \times 10^{-37}$</td>
<td>301441 bp at 5’ side: hypothetical protein 1416400 bp at 3’ side: transducin-like enhancer protein 1</td>
<td>None</td>
</tr>
<tr>
<td>MS37</td>
<td>2</td>
<td><em>Homo sapiens</em> chromosome 1 genomic contig, reference assembly</td>
<td>$5 \times 10^{-30}$</td>
<td>None</td>
<td>Hypothetical protein LOC55248</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Homo sapiens</em> chromosome 1 genomic contig, alternate assembly</td>
<td>$5 \times 10^{-30}$</td>
<td>None</td>
<td>Hypothetical protein LOC55248</td>
</tr>
</tbody>
</table>

### TABLE 3. Repeats in Control and Patient Sequences (RepeatMasker results)

<table>
<thead>
<tr>
<th>Sequences</th>
<th>Control sequences</th>
<th>Patient sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sequences</strong></td>
<td>35</td>
<td>30</td>
</tr>
<tr>
<td><strong>Total length</strong></td>
<td>6550 bp</td>
<td>4183 bp</td>
</tr>
<tr>
<td><strong>GC level</strong></td>
<td>45.45%</td>
<td>44.82%</td>
</tr>
<tr>
<td><strong>Bases masked</strong></td>
<td>2943 bp (45.07%)</td>
<td>1458 bp (34.38%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of elements*</th>
<th>Length occupied</th>
<th>Percentage of sequence</th>
<th>Number of elements*</th>
<th>Length occupied</th>
<th>Percentage of sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SINEs</strong></td>
<td>8</td>
<td>875 bp</td>
<td>13.40%</td>
<td>3</td>
<td>358 bp</td>
</tr>
<tr>
<td><strong>ALUs</strong></td>
<td>7</td>
<td>789 bp</td>
<td>12.08%</td>
<td>3</td>
<td>358 bp</td>
</tr>
<tr>
<td><strong>MIRs</strong></td>
<td>1</td>
<td>86 bp</td>
<td>1.92%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>LINEs</strong></td>
<td>10</td>
<td>1583 bp</td>
<td>21.18%</td>
<td>3</td>
<td>252 bp</td>
</tr>
<tr>
<td><strong>LINE1</strong></td>
<td>9</td>
<td>1256 bp</td>
<td>19.23%</td>
<td>2</td>
<td>256 bp</td>
</tr>
<tr>
<td><strong>LINE2</strong></td>
<td>1</td>
<td>127 bp</td>
<td>1.94%</td>
<td>1</td>
<td>87 bp</td>
</tr>
<tr>
<td><strong>LTR elements</strong></td>
<td>2</td>
<td>256 bp</td>
<td>3.92%</td>
<td>3</td>
<td>330 bp</td>
</tr>
<tr>
<td><strong>DNA elements</strong></td>
<td>1</td>
<td>179 bp</td>
<td>2.74%</td>
<td>1</td>
<td>84 bp</td>
</tr>
<tr>
<td><strong>MER2 type</strong></td>
<td>1</td>
<td>179 bp</td>
<td>2.74%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Satellites</strong></td>
<td>1</td>
<td>170 bp</td>
<td>2.60%</td>
<td>3</td>
<td>314 bp</td>
</tr>
<tr>
<td><strong>Low complexity</strong></td>
<td>2</td>
<td>94 bp</td>
<td>1.44%</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td><strong>Total Interpolated repeats</strong></td>
<td>2693 bp</td>
<td>41.24%</td>
<td>1124 bp</td>
<td>25.70%</td>
<td></td>
</tr>
</tbody>
</table>

*Most repeats fragmented by insertions or deletions have been counted as one element. SINEs (short interspersed elements), MIRs (classical tRNA-derived SINEs), LINEs (long interspersed elements), MER (MEdium Reiterated repeats).
the bases in the sequences shown to contain repeats were masked as human repeats.

MEME/MAST analysis showed the presence of various motives, differing in sequence composition and length, in a number of the sequences analyzed. However, the significance of these motives is unknown at present.

According to RepeatMasker, it would be expected that almost 50% of the human genome consists of repetitive DNA, and thus will be masked by the program. A clear difference between the number of masked sequences can be seen for the control (45%) and patient (54%) circulating DNA (Table 3), which may suggest that free circulating DNA release may be altered in cancer patients as more “specific” DNA sequences are released and thus fewer repeats are observed. Differences in the types of repeats can also be observed in Table 3. It was remarkable that no simple repeats were observed.

If it is taken into account that Alu sequences, which are short interspersed elements (SINEs), only make up ~13% of the genome, whereas genes and related sequences constitute ~25% of the human genome, one should expect to see more sequences representing genes. The significance of the presence of so many Alu repeats, especially in the control DNA, is unknown at present.

No apparent difference between the placement of sequences on chromosomes of patients and controls was evident. However, as the circulating DNA was only obtained from females, it is reassuring that, as expected, no sequences matched to the Y-chromosome. It is also worth mentioning that no viral or bacterial contamination was apparent, as all the sequences we obtained matched with the human genome, and no matches to viral or bacterial DNA sequences were found with BLAST analyses.

### Conclusions

Although we have obtained a variety of sequences from two different sources, that is, the control and the cancer patient, no definitive conclusion regarding the composition of the isolated DNA could be drawn from this information. The major feature of our sequencing effort, although only a small number of clones (65) were analyzed, is the predominance of non-repeat sequences since the human genome is reported to contain almost 50 percent repeat sequences. However, these non-repeat sequences could not be allocated to specific genes.

Cloning and sequencing of free circulating DNA was successful and this first attempt at characterizing free circulating DNA by analysis of sequence data not only confirmed previous results (e.g., size and Alu content), but also revealed the large variety of sequences in circulating DNA. Further characterization of circulating DNA may be beneficial in diagnosis and prognosis and may also contribute to determining the source and function of circulating DNA.

### Conflicts of Interest

The authors declare no conflicts of interest.

### References


Chapter 5

What is the significance of free circulating DNA? More results from a bigger pilot study

From the small study described in chapter 4 it became clear that the number of individuals involved as well as the number of sequences obtained were not sufficient to derive any significant conclusions about the composition, origin or significance of circulating DNA. Thus we decided to conduct a larger pilot study with more patients and controls, unfortunately we were not able to sequence all the colonies that grew, but were limited to a small number per individual because of logistical constraints.

5.1 Prepared manuscript

This manuscript was prepared to present the data obtained for the sake of completeness of this study, however, it will not be submitted as such for publication.

Sequencing circulating DNA – a pilot study

Maniesh van der Vaart and Piet J. Pretorius

Abstract

Circulating DNA was first described before the double helix structure of DNA was known, however research on this subject were quenched by Watson and Crick's findings and by the lack of sensitive methods to study these very small molecules. Despite this, large amounts of research have been done in an attempt to use circulating DNA as biomarkers for cancer or other ailments. The current methods used is lacking the potential to characterize total circulating DNA, we make use of a clone based sequencing method to exploit this gap and obtained 126 sequences from twelve individuals. In conclusion we suggest a high throughput method like 454 sequencing technology to generate a large amount of sequences for meaningful analysis of circulating DNA.

Introduction

Many attempts are being made to develop non-invasive methods of tumor detection by means of circulating DNA analysis, however most of these involve some form of PCR for specific gene targets even though the composition of circulating DNA is unknown, as very little de novo sequencing attempts have been made. According to the results from PCR detection of certain gene targets, one would expect to find many different genes and also some repeating sequences comprising, or at least adding to, the composition of the circulating DNA.

In this pilot study we used a previously described method (Van der Vaart and Pretorius, 2008) to clone and sequence free circulating DNA from a larger cohort of patients, in an attempt to differentiate between the free DNA from prostate cancer patients and healthy individuals. We attempt to describe the size distribution, chromosomal distribution, repeat content and composition of sequences compared to the human genome.

Methods

Blood collection and DNA isolation

Blood from six patients with prostate cancer and six age matched healthy controls were collected in EDTA tubes and centrifuged, plasma was stored at -20°C. Before DNA isolation,
plasma was thawed and centrifuged again. Free circulating DNA was isolated with a phenol-chloroform method from 1.5 ml of plasma which was incubated overnight with a SDS and proteinase K solution. After two extractions, free circulating DNA was precipitated overnight with the aid of glycogen, ammonium acetate and ethanol at -20°C. (Van der Vaart and Pretorius, 2008)

**Cloning and sequencing**
Free circulating DNA was blunted and phosphorylated after which it was ligated to a blunt cloning vector and transformed into competent cells. Twelve clones from each patient and healthy control were sequenced and analyzed. (Van der Vaart and Pretorius, 2008)

**Results**
A total of 126 sequences were obtained, 65 from prostate cancer patients, with a median length of 172 bp (range: 85-466 bp) and 61 from healthy individuals, with a median length of 174 bp (range: 29-743 bp) (see Fig 1a for size distribution). No apparent difference could be observed when circulating DNA isolated from patients was compared with circulating DNA from controls. DNA from both groups showed random distribution over all chromosomes except for chromosome 22 with no matches with sequences (see Fig 1b). Guanine and cytosine composition of the sequences were 43% for cancer patients and 41% for controls. Repeat content differed widely between cancer patients and controls (see Table 1). Blast analysis (Altschul et al., 1997) of the obtained sequences against the reference RNA sequence revealed that 36% of the sequences match with mRNA transcripts.

![Figure 1](image)

(a) Size distribution of sequences
(b) Chromosomal distribution of sequences determined with NCBI Megablast

**Table 1: Repeat analysis with Repeatmasker**

<table>
<thead>
<tr>
<th></th>
<th>Cancer patients</th>
<th></th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of elements*</td>
<td>Length occupied</td>
<td>Percentage of sequence</td>
</tr>
<tr>
<td>SINEs:</td>
<td>14</td>
<td>1915 bp</td>
<td>15.51%</td>
</tr>
<tr>
<td>LINEs:</td>
<td>16</td>
<td>2357 bp</td>
<td>19.09%</td>
</tr>
<tr>
<td>LTR elements:</td>
<td>8</td>
<td>1372 bp</td>
<td>11.11%</td>
</tr>
<tr>
<td>DNA elements:</td>
<td>1</td>
<td>205 bp</td>
<td>1.66%</td>
</tr>
<tr>
<td>Total interspersed repeats:</td>
<td>5849 bp</td>
<td>47.38%</td>
<td></td>
</tr>
<tr>
<td>Small RNA:</td>
<td>0</td>
<td>0 bp</td>
<td>0.00%</td>
</tr>
<tr>
<td>Satellites:</td>
<td>6</td>
<td>867 bp</td>
<td>7.02%</td>
</tr>
<tr>
<td>Simple repeats:</td>
<td>4</td>
<td>281 bp</td>
<td>2.28%</td>
</tr>
<tr>
<td>Low complexity:</td>
<td>2</td>
<td>189 bp</td>
<td>1.53%</td>
</tr>
</tbody>
</table>

* most repeats fragmented by insertions or deletions have been counted as one element
The query species was assumed to be homo, default mode run with blastp version 2.0MP-WashU (Smit et al., 2009)
Conclusions

The size and chromosomal distribution of the obtained sequences compares well with results from other studies (Suzuki et al., 2008). GC composition compares well with that in the human genome (41%) (Scherer, 2008) but repeat composition differs largely: SINEs 16%, 8% and 13%, LINEs 19%, 17% and 21%, LTR elements 11%, 16% and 8% for cancer patients, controls and the human genome respectively (Scherer, 2008). The human genome contain 37.8% of genes (including introns and exons) (Venter et al., 2001), when this is compared to the 36% of sequences that matched mRNA sequences the circulating DNA we analyzed are in line with that observed in the genome. From the large amount of genes and mutations detected in circulating DNA in a range of different studies, one would expect to find a large variety of circulating DNA in different individuals. From these observations the following question arises: Does circulating DNA cover the whole genome or just certain parts?

From the results obtained we conclude that circulating DNA contains a similar proportion of coding sequences and a varying amount of repeat sequences, however, this could be because of sequencing bias based on the small number of sequences involved hindering an accurate analysis of the composition of circulating DNA. This clone-based sequencing method was rather quick and easy to execute but too little data was generated. Therefore we envisioned using 454 sequencing technology to generate a larger number of sequences which may allow a more meaningful analysis.

References


Chapter 6

454 Sequencing™ GS FLX – Massive sequencing

As was alluded to in the previous chapters, sequencing of circulating DNA is crucial in determining the origin, function or significance of these DNA fragments. A logical consequence to the pilot studies is large scale sequencing of circulating DNA in order to generate a vast amount of sequence data. As 454 sequencing technology was locally available to us, we made use of this new innovative method of sequencing in combination with a method to tag DNA from different individuals in order for it to be sequenced simultaneously.

Circulating DNA was isolated from 22 individuals, 12 of which were prostate cancer patients and 10 were age matched healthy controls. Circulating DNA were blunted, phosphorylated and ligated to a blunt cloning vector. High fidelity PCR was performed in order to increase the DNA concentration after which circulating DNA from each individual was separately tagged. Tagged DNA were pooled and subsequently sequenced on one 16th of a picotiter plate on a GS FLX sequencer (454 life sciences, Roche). Obtained sequences were processed and analyzed. A paper describing the sequencing method and results was prepared and accepted for publication.

6.1 Paper published in Clinica Chimica Acta

This paper was prepared in collaboration with Dr. Dmitry V. Semenov from the Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, Russia (see supplementary material, Section 7 for permission from co-authors). It was accepted for publication in Clinica Chimica Acta, International Journal of Clinical Chemistry and Diagnostic Laboratory Medicine, on the 4th of August 2009, with digital object identifier: 10.1016/j.cca.2009.08.011.
**Characterisation of circulating DNA by parallel tagged sequencing on the 454 platform**

Manlesh van der Vaart\(^{1*}\), Dmitry V. Semenov\(^{2}\), Elena V. Kulagina\(^{2}\), Vladmir A. Richter\(^{2}\), Piet J. Pretorius\(^{1}\)

\(^{1}\)Biochemistry Division, School for Physical and Chemical Sciences, North-West University, Potchefstroom Campus, South Africa, \(^{2}\)Institute of Chemical Biology and Fundamental Medicine SB RAS, Lavrentiev Ave 9, Novosibirsk, 630090, Russia.

*Corresponding author. Tel: +27(0)18 871 2396, E-mail: Manlesh.vanderVaart@mwu.ac.za, Fax: +27(0)18 219 3316*

**Abstract**

Background: Fragments of genomic DNA that can be isolated from the blood and body fluids of vertebrates is also known as circulating DNA. This DNA has widely been investigated as a biomarker for cancer and other diseases but the origin and significance of circulating DNA has not been elucidated to date.

Methods: We used a parallel tagged sequencing method to sequence circulating DNA obtained from control individuals as well as cancer patients on the GSFLX sequencer (454 life sciences).

Results: Circulating DNA sequenced on one 16\(^{th}\) of a picoliter plate produced \(\sim 3600\) unique circulating DNA sequences which were distributed over the human genome and a higher frequency of mutations were observed in cancer patients compared to healthy controls.

Conclusion: Circulating DNA represent genomic DNA in the blood of an individual, some sequence related differences might be evident between circulating DNA from cancer individuals and controls but distribution over the genome is similar.

Key words: Human genome, 454 Sequencing, circulating DNA

**Introduction**

Circulating DNA has become an important source of DNA biomarkers for mutation and epigenetic assays related to various forms of human cancers. These DNA markers are, however, focused on specific genes and gene promoters and are limited in their scope of revealing information on the characteristics and function of circulating DNA. In addition, some controversy exists in the literature on the origin of these nucleic acids in the blood of vertebrates [1-5]. Consequently the need for a broader characterization of circulating DNA has become a priority since little is known about their origin, function and significance even though research in this area has been done for almost 60 years.

Thus far the immense data mining power of 454 sequencing technology have not really been utilized in an attempt to characterize circulating DNA. A few reports utilizing 454 sequencing technology on circulating DNA has been published, but the goal of these studies were not to characterize total circulating DNA but rather ultra-deep amplicon sequencing of amplification products of a small number of gene sequences or methylation specific sequencing [6-8]. One study used shotgun pyrosequencing for diagnosis of fetal aneuploidy [1] and was able to diagnose different trisomy cases very early in pregnancy. A recent study by Beck [9] described sequencing of total circulating DNA from serum of 51 apparently healthy individuals, a large number of sequences were obtained and analysis revealed some content information of circulating DNA.

We previously reported [10] on a pilot study of sequencing circulating DNA using clonet-based conventional sequencing. In this study we observed that from the small number of sequences we obtained, circulating DNA cannot be characterized nor can the significance be determined, this prompted the need for large-scale sequencing. We utilised a parallel tagged sequencing method introduced by Meyer et al. [11] which enabled us to sequence circulating DNA of 22 individuals in a single GS FLX run.

**Methods**

*Isolation, cloning and sequencing of circulating DNA*

Ethics approval from the North-West University ethics committee was obtained (05M12) and this study was performed in accordance with the ethical guidelines for human genome research. Blood was collected with informed consent from 12 prostate cancer patients and 10 age-matched healthy males. Plasma DNA was isolated using a phenol-chloroform based method as previously reported [10]. Circulating DNA from each individual was blunt, phosphorylated and ligated to a blunt cloning vector after which the cloned DNA was amplified using high fidelity PCR. In subsequent reactions the amplified DNA from each individual
was blunted, repaired and phosphorylated separately after which sample specific barcoding adapters were ligated to both ends. After each reaction DNA was cleaned with either SureClean, Ampure magnetic beads or the Qiagen min elute kit [12]. Each barcoding adapter comprises of a single self-hybridized palindromic oligonucleotide containing a SrfI restriction site flanked by complimentary sequence tags [11]. Tagged samples were quantified using the NanoDrop® ND-3300 fluorospectrometer and pooled in equimolar ratios. The resulting sample pool were dephosphorylated and subsequently digested with the SrfI restriction enzyme. This resulted in blunt DNA fragments bound to a PCR primer from the vector and a unique barcoding tag for each individual’s sample on both ends of the DNA fragments. The pool of barcoded samples was sequenced using the 454 sequencing platform (454 life sciences, Roche). After preparing a GS FLX library starting with the adapter ligation step, one 16th of a lane of a GS FLX picotiter plate was used for the sequencing reaction.

**Bioinformatic methods**

Raw 454 sequence reads were separated using the primer and barcoding tag sequences, subsequently the sequences were cleaned from vector and tag ‘contamination’ and exact duplicates were identified. The generated array of circulating DNA sequences was compared to known sequences including the human reference genome using the BLAST (Basic local alignment search tool) search engine [13]. Three consecutive searches were performed using different parameters and databases. The first BLASTn search was against the NCBI nr database with the following non-default parameters: expect = 0.001, filter = false. The two best matches were collected and analysed. This allowed sequences that still contained imperfect vector tags and non-human sequences to be removed. The second blast search was performed against the current human genome RefSeq database (NCBI Build 36.1) with the same non-default parameters, again the two best matches for each sequence were collected. The third BLASTn search was performed using a local copy of the human_ref_contig database with locally installed program blastall (ftp://ftp.ncbi.nlm.nih.gov/blast/). The local blastn search was performed with the “touch down expect” approach first with expect value of 1e-40, all unresolved sequences with expect value of 1e-20 then 1e-10 and still unresolved sequences with expect value of 1, other parameters as above. Positions of matches in RefSeq human contigs were converted to human chromosome locations using the contig location map from the UCSC (http://genome.ucsc.edu). The chromosome locations reported by the two BLAST searches performed against the human RefSeq database were compared and sequences were separated into three groups: sequences with exactly reproducible locations, those with shifted locations within the length of the sequence, and sequences that could be attributed to completely different chromosome locations. The latter represented human genome repeat sequences. About 65% of doubtfully located sequences were human satellite repeats. The perfect matches of the latter were found by BLASTn in either the human HTGS or in the human WGS traces databases (NCBI).

Strongly overlapping sequences or not-exact duplicates in individual donors were identified using the second BLAST search results. Fragment length and G+C content estimation were performed on the generated unique sequences using direct string analysis with Microsoft Office Excel 2007, repeat content was determined using RepeatMasker [14]. Overlaps with known human genes were evaluated with the use of the RefSeq Genes track of the UCSC Table Browser [15, 16]. In order to characterize mutations we composed a Microsoft Office Excel 2007 database of the xml formatted BLASTn output results, and generated a complete list of mismatches observed in the experimental DNA sequences relative to the human reference sequences. To avoid errors that could be introduced by sequences with ambiguous genome locations, only reproducibly located sequences in the RefSeq human genome experimental sequences were included in the search of mismatches. All mismatches (12109 mismatches within a total of 805170 bp in the search = 1.5%) were compared with others within the set of donor-specific records to calculate reproducibility and quasi-heterozygosity. Mismatches that were twice and more frequently reproduced and were not "heterozygous" (have no correct counterpart) were regarded as mutations. All mismatches were also compared with known SNPs using the SNF129 track of the genome database (ucsc.genome.edu).

**Results**

Various analyses were performed on the generated sequences in an effort to characterize the sequenced circulating DNA from 12 prostate cancer patients and 10 healthy male controls. A total of ~8600 sequences were obtained, sequences from cancer patients and controls...
were identified by tags. Vector contamination and tags were removed as well as exact duplicates after which ∼4150 sequences remained, from those a further 518 sequences were identified as not-exact duplicates, thus a total of ∼3600 unique sequences remained 49% of which were from cancer patients and 51% from control individuals.

Figure 2 it is evident that the G+C content of the sequences were not normally distributed as each group shows the presence of two peaks, one at 40-41% and the other at either 46-47% or 48-49% for the cancer and control group respectively. BLAST searches for chromosome distribution revealed that from a total of ∼3600 unique sequences, 86% showed high probability matches with the human reference genome, this value applies equally to the cancer patients and the controls (14% of sequences gave uncertain locations in the human genome because of repeat content or closely related sequence content). When the number of sequences from the cancer patients and controls were compared (Figure 3), no distinct difference in representation was apparent, however, in some cases more sequences were from control individuals while in other cases more were from cancer patients. The numbers of sequences that matched to each chromosome (see Figure 3) show a correlation with the size of the chromosomes (except for chromosome X and Y). Mean distances (Figure 3, insert) for unique locations were 1.92±0.32 Mb/seq and 1.86±0.36 Mb/seq for cancer and healthy autosomes, respectively (R²=0.90 and 0.87, respectively). For chromosome X a higher distance variant (or lower representation variant) were observed (∼3.23 Mb/seq) for both the cancer and healthy groups of sequences. Chromosome Y of the healthy individuals was represented with 4.1 Mb/seq while that for cancer patients were underrepresented with 14 Mb/seq.

Figure 1: Size distribution of sequenced DNA fragments.

The size distribution of sequences (see Figure 1) showed that 393 and 413 sequences from patients and controls respectively were between 161 and 170bp in length and 80% of all the sequences were between 140 and 200bp in length. No significant difference could be observed between the length of sequenced circulating DNA from cancer patients and controls. Taking into account that the upper limit for the size of fragments sequenced with the GS FLX is ∼250bp. We did not assemble the sequences since nebulization was not performed before sequencing as the majority of the isolated DNA fragments in the sample pool were ∼300bp (including vector and tag sequences). Thus some sequences only contained the sample tag on one end of the sequence because their entire length was not sequenced. It is possible that this could have introduced some bias in the length distribution shown in Figure 1. This bias would most probably be in the size range ∼160-210bp with an average of 184bp as the tag and vector sequence that was removed was ∼66bp in length.

The G+C content of the sequences was calculated (see Figure 2) and almost equal representation of sequences from controls and cancer patients were observed. The average and median G+C content was ∼45%, with a minimum of ∼21% and a maximum of ∼78% for both patients and controls. However, according to

Figure 2: G+C content of the sequenced DNA fragments

Thus, a tendency of underrepresentation of chromosome Y-derived sequences in circulating DNA of cancer patients can be observed, but the total number of Y derived sequences was only 18, which do not allow us to make a conclusion about the diagnostically significance of the fact.
RepeatMasker [14] are grouped in the four main repeat classes found in the genome (Table 1) which can be compared to the human genome. The majority of repeat sequences in circulating DNA from patients as well as controls are either LINE1 or Alu repeats since they represent almost half of the total repeat count. A substantial number of satellite sequences were also detected in both groups. An interesting feature of the results given in Table 1 is that the sequenced DNA from cancer patients comprises of 42% repeats and the controls 46%. This difference was not due to overrepresentation of a single repeat element since all the repeat classes in Table 1 had slightly higher representation in the control group than in the cancer group.

In Figure 4 it is noticeable that circulating DNA from cancer patients as well as from control individuals are scattered over the whole genome, with no apparent pattern of distribution. Although this figure might give the impression that almost the whole genome is covered by the sequenced circulating DNA this is not the case as only ~572 kb of unique sequence data versus the ~5080.4 Mb of the human genome [17] were obtained, thus the sequenced circulating DNA covered only 0.01% of the human genome.

Repeat analyses were performed in order to determine the number and type of repeats present in the circulating DNA. The results obtained with

Figure 5: Chromosome ideograms with positions of unique sequences of circulating DNA fragments. Sequenced circulating DNA is indicated as horizontal lines next to the ideograms to indicate the positions of circulating DNA fragments in the human genome. DNA from cancer patients are represented by red lines on the left of each chromosome while DNA from controls is represented by blue lines on the right of each chromosome. In order for the lines to be visible on this figure their widths are not drawn to scale.

(Figure composed from NCBI map viewer ideograms: http://www.ncbi.nlm.nih.gov/projects/mapview/map_search.cgi?taxid=9606)
The mutation rate of sequenced circulating DNA was higher in cancer patients than in control individuals (see Figure 5). It was also notable that transitions from A to G and T to C and the reverse of these, thus G to A and C to T (A→ T>G; C and G:C→A:T), were increased significantly if compared to other types of mutations in both cancer patients as well as in control individuals. No direct correlation could be found between donor age and mutation rate observed in sequenced circulating DNA (unpublished data).

Table 1: Repeat analysis of unique sequences by RepeatMasker

<table>
<thead>
<tr>
<th></th>
<th>Cancer Patients</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of sequences:</td>
<td>1758</td>
<td>1836</td>
</tr>
<tr>
<td>Total length:</td>
<td>280389 bp</td>
<td>291807 bp</td>
</tr>
<tr>
<td>GC level:</td>
<td>45.05 %</td>
<td>44.92 %</td>
</tr>
<tr>
<td>Bases masked:</td>
<td>118574 bp (42.29 %)</td>
<td>134371 bp (40.05 %)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Number of elements</th>
<th>Length occupied</th>
<th>Percentage of sequence</th>
<th>Number of elements</th>
<th>Length occupied</th>
<th>Percentage of sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINEs</td>
<td>270</td>
<td>28709 bp</td>
<td>10.24 %</td>
<td>302</td>
<td>33755 bp</td>
</tr>
<tr>
<td>ALUs</td>
<td>222</td>
<td>24007 bp</td>
<td>8.59 %</td>
<td>258</td>
<td>29321 bp</td>
</tr>
<tr>
<td>MIRs</td>
<td>47</td>
<td>4555 bp</td>
<td>1.62 %</td>
<td>42</td>
<td>4206 bp</td>
</tr>
<tr>
<td>LINEs</td>
<td>284</td>
<td>34466 bp</td>
<td>12.29 %</td>
<td>280</td>
<td>36540 bp</td>
</tr>
<tr>
<td>LINE1</td>
<td>223</td>
<td>30295 bp</td>
<td>10.79 %</td>
<td>224</td>
<td>30354 bp</td>
</tr>
<tr>
<td>LINE2</td>
<td>36</td>
<td>3978 bp</td>
<td>1.32 %</td>
<td>52</td>
<td>5316 bp</td>
</tr>
<tr>
<td>L3CR1</td>
<td>3</td>
<td>235 bp</td>
<td>0.08 %</td>
<td>4</td>
<td>398 bp</td>
</tr>
<tr>
<td>LTR elements</td>
<td>214</td>
<td>28212 bp</td>
<td>10.05 %</td>
<td>233</td>
<td>31423 bp</td>
</tr>
<tr>
<td>ERVL</td>
<td>46</td>
<td>5731 bp</td>
<td>2.04 %</td>
<td>53</td>
<td>7495 bp</td>
</tr>
<tr>
<td>ERVL-MaLRs</td>
<td>101</td>
<td>13655 bp</td>
<td>4.87 %</td>
<td>91</td>
<td>11530 bp</td>
</tr>
<tr>
<td>ERV_class1</td>
<td>61</td>
<td>7586 bp</td>
<td>2.65 %</td>
<td>80</td>
<td>10846 bp</td>
</tr>
<tr>
<td>ERV_class2</td>
<td>8</td>
<td>747 bp</td>
<td>0.27 %</td>
<td>7</td>
<td>1072 bp</td>
</tr>
<tr>
<td>DNA elements:</td>
<td>48</td>
<td>4701 bp</td>
<td>1.68 %</td>
<td>57</td>
<td>6226 bp</td>
</tr>
<tr>
<td>hAT-Charlie</td>
<td>26</td>
<td>2620 bp</td>
<td>0.93 %</td>
<td>32</td>
<td>3250 bp</td>
</tr>
<tr>
<td>TcMar-Tigger</td>
<td>12</td>
<td>1343 bp</td>
<td>0.46 %</td>
<td>15</td>
<td>1506 bp</td>
</tr>
<tr>
<td>Unclassified</td>
<td>3</td>
<td>600 bp</td>
<td>0.21 %</td>
<td>3</td>
<td>469 bp</td>
</tr>
<tr>
<td>Total interspersed repeats:</td>
<td>96597 bp</td>
<td>34.45 %</td>
<td>108232 bp</td>
<td>37.09 %</td>
<td></td>
</tr>
<tr>
<td>Small RNA:</td>
<td>2</td>
<td>263 bp</td>
<td>0.09 %</td>
<td>6</td>
<td>563 bp</td>
</tr>
<tr>
<td>Satellites:</td>
<td>121</td>
<td>18380 bp</td>
<td>6.56 %</td>
<td>140</td>
<td>21132 bp</td>
</tr>
<tr>
<td>Simple repeats:</td>
<td>31</td>
<td>2700 bp</td>
<td>0.96 %</td>
<td>47</td>
<td>4145 bp</td>
</tr>
<tr>
<td>Low complexity:</td>
<td>11</td>
<td>634 bp</td>
<td>0.23 %</td>
<td>12</td>
<td>444 bp</td>
</tr>
</tbody>
</table>

* most repeats fragmented by insertions or deletions have been counted as one element
The query species was assumed to be Homo default mode run with blastp version 2.0MP-WashU (1)

Additional analyses
Estimation of overlaps of genome locations of circulating DNA with the RefSeq genes revealed that an average of 40.3% and 36.8% of DNA from cancer patients and controls respectively (32.3%-51.5%, min-max) coincided with human genes. An average of 2.6% and 2.4% of total basepair counts from cancer patients and controls respectively (0.4%-4.4%, min-max) overlapped with exons in the human genome.

BLAST analysis of the sequenced circulating DNA also indicated that some sequences could not be matched to the current human reference sequence. Non-human DNA that was detected included one of each of the following: Phileum pratense, Agrobacterium tumefaciens, Chaetomium globosum, Escherichia coli, Glaucoma hebraicum and Salmonella enterica and also nine sequences that matched to uncultured bacterium clones.

We did not perform analysis of sequence termini as it would not provide accurate end point information of circulating DNA because of the various blunting reactions performed on the DNA.

Discussion
We utilized a parallel tagged sequencing method [11] to tag the double stranded circulating DNA fragments from 22 individuals and sequenced it simultaneously on one 16th of a plate in a GS FLX run. After processing the raw data a total of ~3600 unique sequences were obtained and comparison between circulating DNA and the human genome was made. This comparison was done to determine whether circulating DNA is comprised of random fragments distributed over the whole genome map or if certain parts of
the genome is in abundance or represented to a
greater extent. Even though the circulating DNA
sequences we have sequenced only covered
0.01% of the genome, a chromosomal distribution
pattern was observed (Figure 3) which is in good
agreement with the sizes of the different
chromosomes except for chromosome Y. Beck et
al. (9) observed the same correlation, but they
found chromosome 19 to be under-represented.
Positioning of these sequences on the human
ideogram map (Figure 4) showed a distribution of
sequences over the whole genome, again with
under-representation of chromosome Y. The low
number of sequences obtained for chromosome Y
does not allow us to make a conclusion about the
diagnostically significance of this under-
representation. No significant difference could be
observed between circulating DNA from cancer
patients or controls with regards to chromosomal
location and distribution. The observation that
circulating DNA may represent the whole genome
supports its use as an easily accessible source of
material for DNA based analysis. The size
distribution of sequences (Figure 1) showed that
most sequences were between 161 and 170bp in
length and 80% of all the sequences are between
140 and 200bp in length, a similar pattern was
observed in other studies [1, 3].

The average GC content of the circulating
DNA was found to be slightly (~4%) higher if
compared with that of the human genome (41%)
[17], Fan et al. [1] observed a 10% higher GC
content in circulating DNA sequences than in the
genome which they attributed to sequencing bias.
According to Scherer [17] gene density is
positively related with guanine and cytosine
(G+C) base composition and a figure produced by
Lander et al. [18] in which they compare the G+C
distribution of the genome versus that only in
genes produce a similar bimodal distribution for
genes than that which can be observed in Figure
2 for the G+C content of circulating DNA from
either cancer patients or control individuals. This
may indicate that circulating DNA is slightly
enriched with gene rich sequences as it has a
similar GC distribution as that found in genes.

The repeat content of circulating DNA from
controls and cancer patients were 46.1% and
42.3% respectively, the sequenced fragments
being comprised of repeats were thus slightly
higher for controls and slightly lower for cancer
patients when compared to the repeat content of
the human genome (see Table 1) which is ~45%
[17]. Furthermore the distribution of the repeat
classes individually were for all but LTR elements
lower in the sequenced circulating DNA than in
the human genome: 10.2% and 11.6% SINE’s vs
13%, 12.3% and 12.5% LINE’s vs 21%, 10.0% and
10.7% LTR elements vs 8% and 1.7% and
2.1% DNA transposons vs 3% were identified for
cancer patients and controls respectively
compared to that of the human genome [17].
Only the amount of LINEs which comprise 21% of
the human genome differed significantly from that
in the sequenced fragments. This large
underrepresentation of LINEs may be explained
by the fact that LINEs are situated in gene-poor
AT-rich regions of the human genome [19]. The
small difference in the amount of SINEs in control
patients is caused by a relatively smaller amount
of mammalian interspersed repeats (MIR) while
the amount of Alu sequences correlate with that
of the draft genome sequence [18], however, in
cancer patients both Alu sequences and MIRs are
underrepresented. The fact that the repeat
content of the sequenced circulating DNA differ
from that found in the human genome may probably
be because of the mechanism of release of
circulating DNA, but also because of sequence
bias [9]. When the repeat content of circulating
DNA obtained from cancer patients was
compared with that of control individuals a slightly
lower repeat content was observed in cancer
patients.

Mutation analysis of the sequenced
circulating DNA revealed a significant increase in
transitions, particularly A>G and T>C and its
reverse G>A and C>T. These two groups of
mutations were also shown to be increased in
interspersed repeats in the analysis of the initial
human genome sequence [18]. Furthermore, in a
study of the effect of aging and proliferation rate
on mutation frequency, it was shown that the rate
of G>A and C>T mutations increase significantly
with age and proliferation rate of cells in mouse
studies [20]. An overall increase in mutation rate
was observed in circulating DNA from cancer
patients when compared to controls.

While most of the sequenced circulating DNA
could be identified as human sequences, a few
sequences matched better with non-human and
primate sequences of the nr database. The 14
non-human sequences which matched to
bacterial genomes can be explained by a
presence of infection agents’ DNA. The two
sequences that matched to fish and plant
genomes could possibly be explained by
presence in circulation of food derived DNA. The
220 sequences of circulating DNA that produced
better matches to non-human primate records in
the nr database than to RefSeq human
sequences, were found to be attributed to human
satellite repeats, which are not yet represented in
the human Ref_seq genome database but

Characterisation of circulating DNA by parallel tagged sequencing  p6
represented in the human HTGS or in the human WGS traces databases.

The transcribed DNA content of the sequenced circulating DNA was 40.3% and 36.8% in cancer patients and controls respectively, compared to the 37.8% of genes (including introns and exons) in the human genome [21]. A slight increase can be observed in the gene content of sequenced circulating DNA from cancer patients. The number of basepairs that matched to exons was 2.6% and 2.4% of the total basepairs from cancer patients and controls respectively. This is slightly elevated if compared to the 1.4% to 1.8% in the human genome [21, 22]. This increase in coding sequences observed in both cancer patients and control individuals support the conclusion of gene enrichment of circulating DNA as alluded to above.

Circulating DNA may represent the whole genome in spite of some differences in repeat and GC content, basically the same composition and chromosomal distribution as in the genomic DNA was observed for sequenced circulating DNA. However, a possibility for gene enrichment of circulating DNA is evident.

Except for: a higher frequency of mutations, obvious differences could not be observed between circulating DNA from cancer patients and control individuals, this is a somewhat disappointing observation, but all findings are in favor of the notion that circulating DNA present a minimally invasive source of genetic material which can be used for DNA based molecular analysis.

Acknowledgements

The work was partially supported by The National Research Foundation of South Africa (MvdV) and with the SB RAS integration projects grant # 18, 2009-2011 (DVS).

References
6.2 Some technical aspects

- Quality assurance of tagged sequences is essential, we found that analysis of circulating DNA on the bioanalyzer (Agilent) has proved to be valuable and is recommended for future use. Comparison between untagged, tagged and pooled and restriction digested DNA is a useful way to determine if tagging and restriction digestion was successful (see Figure 6 of Section 2 in the supplementary material).

- It is possible that less circulating DNA of sample 2 was present in the sample pool resulting in less sequences being produced for sample 2. This could be the result of repeated library preparation because of delays in processing of the sample pool by Inqaba Biotech. This proved the need for immediate preparation of the sequencing library once the sample pool is prepared as degradation may occur if stored longer than one week.

- More pEZSeq cloning vector DNA was present in the circulating DNA from cancer patients, initially 52% and 48% of the sequences were identified by the tags as DNA from cancer patients and controls respectively. After all duplicates and vector sequences were removed these numbers were reversed. This was an interesting observation as there were more cancer patients (n=12) than control individuals (n=10) participating in the study, thus it is expected that more cancer DNA should be obtained as circulating DNA was added in equal molar ratios to the sequence sample pool. It was determined that this shift in the amount of DNA contributed by each group was caused by a much larger vector content in the circulating DNA isolated and cloned from cancer patients (61% of total vector sequence count were from sequences tagged as from the cancer group). This could indicate that the efficiency of cloning of circulating DNA from cancer patients is less than for the control group as more vector self-ligated without a circulating DNA insert even though all reaction conditions were identical.

6.3 Suggested improvements to the sequencing method

- Elimination of PCR to increase DNA concentration would be beneficial as PCR could produce sequencing bias and may influence copy number representation of individual sequences.

- Combining A B adapter ligation (included in preparation of a 454 sequencing library) with tag ligation could improve sequencing efficiency by eliminating two steps of blunt ligation.

- If a whole picotiter plate on the GS FLX 454 sequencer instead of one 16th of a plate can be used for sequencing, a much more thorough sequencing result would be obtained, however, this was not possible because of logistical constraints.

- Making use of the new GS FLX Titanium series reagents will make it possible to obtain sequences of up to 400-500 bp in length.
Chapter 7

Summary and conclusions

Short DNA fragments known as circulating DNA can be isolated from the blood of humans and vertebrates, it has been shown that patients with cancer and a variety of other conditions have increased concentrations of these circulating DNA molecules. The main question regarding circulating DNA concerns their origin and function or significance. Even though a number of groups have done extensive research on the origin of circulating DNA an explicit conclusion have not been reached regarding their origin and significance (see chapter 2 for references). Leading hypotheses are apoptosis, necrosis and active release by living cells.

In order to establish which mechanism of origin is the most plausible, available research findings on this topic were systematically compiled and close evaluation of these reports revealed that the majority of papers on circulating DNA, even though not on the origin, make a general statement that apoptosis is the main source of circulating DNA, based on the fact that circulating DNA produce a ladder like pattern after electrophoresis (see chapter 2 for references). However, it was shown that this ladder pattern can also be observed after electrophoresis of DNA isolated from cell culture medium where no cells were undergoing apoptosis. Furthermore, as a rule apoptotic cells are ingested by phagocytes and DNA are consequently completely digested. Most proliferating cells lose the ability to become apoptotic, thus many cancer cells are resistant to apoptosis. Necrosis on the other hand produces large, randomly sheared DNA fragments, which is in contrast to that observed for the main part of circulating DNA. These arguments exclude apoptosis and necrosis as the main source of circulating DNA. In support of an active release mechanism, the amount of circulating DNA decreased after radiotherapy, while an increase is expected if cell death were involved in their origin. It was observed that circulating DNA is released until a certain concentration is reached, with extended incubation the concentration is unchanged, but when cells are transferred to new medium DNA is released to the same level, this observation was independent of dying cells and newly synthesized DNA were released preferentially. Preferential release were also evident when the proportion of Alu repeat sequences where compared to β-globin in serum and lymphocytes. Thus we concluded that apoptosis and necrosis are not the main mechanism of origin for circulating DNA and that active release of DNA by living cells is to a great extent a more plausible proposition. Another prominent fact is that cancer patients and patients with a variety of other conditions have more circulating DNA in their blood compared to healthy individuals. Increased levels of circulating DNA in the blood of these patients may be because of a disturbance in the equilibrium between DNA release by living cells and DNA clearance mechanisms by adverse conditions. Even though many of the conditions where increased
circulating DNA levels can be observed in the blood are unrelated, the presence of circulating DNA is a common feature, thus some correlation ought to be found which might point to a similar mechanism of origin. More work should be done in order to determine the particular mechanism of DNA release from cells.

A large number of research reports are available about utilizing the potential of circulating DNA as biomarker(s) for a variety of pathological conditions (see Chapter 3 for references). A range of methods are used to isolate and quantify circulating DNA and also to detect tumor related changes in the circulating DNA. Even though higher concentrations of circulating DNA can be detected for cancer patients in individual studies, data from a collection of studies show this only as a trend and standardization of the methodology and identification of the factors that may influence the amount of circulating DNA in the blood is crucial before quantification of DNA can be exploited as a biomarker for cancer. Detection of genetic alterations like mutations, microsatellite alterations or DNA methylation can be rather contradictory and even confusing as the same marker can have different detection patterns in one as well as for different cancers and panels of markers may miss previously clinically diagnosed patients. Thus, a mere increase in DNA concentration without knowing the origin or nature of circulating DNA and detection of genetic alterations without knowledge about the significance of circulating DNA is too vague to use as informative markers, from this we concluded that the application for circulating DNA as an unambiguous biomarker is currently overrated.

The unresolved question about the origin and significance of circulating DNA led to the hypothesis that elucidating the sequence composition of circulating DNA might contribute to understanding the origin and possibly the significance of these DNA fragments in the human body. A method to sequence isolated circulating DNA on small scale was standardized in order to obtain preliminary sequence information.

A paper was published about a clone based sequencing method where circulating DNA were isolated, blunted and phosphorylated after which it was ligated into a blunt cloning vector. The resulting plasmids were transformed into competent cells and colonies were screened for inserts which were subsequently sequenced. In the first study a total of 65 colonies from two individuals were sequenced, this was followed by a bigger pilot study where a total of 128 colonies from 12 individuals were sequenced. A large variety of sequences were obtained without a significant number of duplicates and no apparent difference could be observed when circulating DNA from cancer patients were compared to that of controls. Size and chromosomal composition compared with that found in other studies, GC composition and gene content compared with that of the human genome. Repeat composition which differed largely from that
of the genome could be attributed to sequence bias or as a result of the small amount of sequences obtained which may not be representative of circulating DNA as a whole. Although the method for sequencing circulating DNA was successful, it was concluded that the sequencing information obtained with this small scale clone based sequencing method was insufficient to perform meaningful analysis or to derive significant conclusions from. It was therefore suggested that large scale sequencing on the 454 sequencing platform should be performed.

Circulating DNA isolated from 12 cancer patients and 10 age matched healthy individuals were blunted, phosphorylated and tagged in order to be sequenced on one 16th of a picotiter plate in a GS FLX sequencer (454 Life sciences, Roche). The large scale sequencing was successful and a great amount of data was generated in the process since almost 5500 usable sequences was obtained which were subsequently analyzed. While this study was in progress, a paper was published also reporting on large scale sequencing of circulating DNA, but from the serum of only healthy individuals (Beck et al, 2009). This group sequenced DNA isolated from serum, while we sequenced DNA isolated from plasma. We believe that plasma is a more appropriate source of circulating DNA than serum because of the cellular breakdown during the blood clotting process, thus plasma samples are less likely to suffer cellular contamination and it shows less day-to-day variability. Furthermore, in our study circulating DNA obtained from cancer patients and healthy individuals were analyzed and compared in order to determine if the sequence composition between these two groups varies and whether any such differences would be significant enough to be utilized as biomarker(s) for cancer.

Sequence analysis of the sequences showed that even though the obtained circulating DNA covered only 0.01% of the genome, a chromosomal distribution pattern which is in correlation with the sizes of the different chromosomes, except for chromosome Y, was observed. Positioning of these sequences on the human ideogram map showed a distribution of sequences over the whole genome, again with underrepresentation for chromosome Y. No significant difference could be observed between circulating DNA from cancer patients or controls with regards to chromosomal location and distribution. The average GC content of the circulating DNA for both groups was found to be slightly elevated when compared to the human genome and GC distribution showed a similar pattern as that observed for genes by Lander et al. (2001) which may indicate the possibility for gene enrichment in circulating DNA. Repeat content showed slight differences on the one hand, between circulating DNA from cancer patients and controls and, on the other hand, when compared to that of the human genome. Mutation analysis of the sequenced circulating DNA revealed a significant increase in transitions, particularly A>G and T>C and its reverse G>A and C>T. These two groups of
mutations are also elevated in interspersed repeats of the human genome (Lander et al. 2001). An overall increase in mutation rate was observed in circulating DNA from cancer patients when compared to controls. Sequenced circulating DNA that produced better matches with non-human sequences could be explained by presence of infectious agents’ DNA or food derived DNA in the circulation and better matches to primate DNA were found to be attributed to human satellite repeats, which are not yet represented in the human genome reference sequence database. A slight increase could be observed in the gene (introns and exons) content of sequenced circulating DNA from cancer patients and increased exon content in circulating DNA from both cancer patients and control individuals support the possibility of gene enrichment of circulating DNA as alluded to above.

The question arises whether circulating DNA truly represents the whole human genome. If this is the case the role of apoptosis as the source of circulating DNA may not be underrated. On the other hand, if circulating DNA contains only a portion of the genome or an increase in certain portions, a mechanism must exist to perform this selection. This mechanism can be a specific release mechanism or sifting process utilized by cells to excrete a specific proportion of DNA fragments into the circulating DNA pool. If so, is this proportion always the same and which portion of the genome would it be?

If circulating DNA is in fact of apoptotic origin, why and how did it bypass macrophage engulfment and DNA digestion? And how does circulating DNA bypass the immune response to DNA? Why would cells release their content into the blood? Does circulating DNA have a function? Increased circulating DNA can be observed in athletes after high volume exercise (Fatouros et al., 2006), if apoptosis is the origin of this DNA a very large number of cells must undergo apoptosis during exercise?

Why is circulating DNA released?

- Methylation: The importance of maintenance of methylation patterns is emphasised in diseases like cancer where the normal methylation patterns are lost, this is thought to be among the earliest molecular alterations during tumorigenesis (Korshunova 2008). However it is not known how methylation is changed in already divided and differentiated cells. Maintenance methylation is carried out as the cell divides and DNA is copied to the new cell, DNA from the parent cell acts as a template for methylation and the new cell thus has the same methylation pattern as the parent cell. But methylation changes in adult cells occur via de novo methylation of which the mechanism is not fully understood at present. Is it possible that circulating DNA play a role in methylation by possibly facilitating de novo methylation? If so, circulating DNA may act as an
"instigator" to change methylation patterns in other (possibly distant) cells where circulating DNA may be absorbed.

- Metastasis: Could circulating DNA that contain cancer specific mutations be absorbed by other cells where the mutation can possibly be copied into the acceptor cell’s DNA causing the cancer mutation and thus cancer?

- Signaling repair or synthesis processes in the cell: Could circulating DNA be released as a signal of DNA sequences that was repaired by a specific cell as indication for the repair machinery of other cells that may be in need of repair in the same sequences?

- Could circulating DNA play a role in maintenance of the genome, where one cell release circulating DNA which could have a maintenance effect in other cells or tissues?

- If any of the above mentioned speculative theories are true, circulating DNA may have a prominent effect in blood transfusions: A patient receiving blood will in the process receive circulating DNA from the blood donor which may have a number of effects on the receiver, including the possibility of cancer contagiousness via metastasis.

Suggestions for further research would include:

- Sequencing circulating DNA from individuals at different time intervals and under different conditions (e.g. exercise), to determine if the composition of circulating DNA change as a result of different conditions.

- Microarray or fish analysis of circulating DNA of cancer patients versus healthy individuals, as well as normal individuals exposed to a number of different conditions.

- Cell culture or in vitro work with different changes in the environment or challenges to the cells like heat shock or concentration changes in the medium – what effect could be observed on the circulating DNA, do these changes alter the nature and composition of circulating DNA?

The success of this study lies in its contribution to fill the gap in the knowledge pertaining to the isolation and characterization of circulating DNA from the plasma of healthy as well as diseased individuals. Isolation, sequencing and analysis of circulating DNA sequences revealed that a significant difference could not be found between the location and distribution of circulating DNA from cancer patients and control individuals. However, a higher frequency of mutations has been observed in circulating DNA from cancer patients. Another interesting observation was the slightly elevated exon content of circulating DNA and similarities of GC content distribution to that of genes in the human genome which might indicate gene enrichment of circulating DNA. Except for the observation that distribution of circulating DNA over the human genome may point to whole genome presence in the circulating DNA, characterization of circulating DNA, thus far, did not reveal the origin or significance of circulating DNA, it does, however, support the use of circulating DNA as minimally invasive source of material for DNA based assays.
References


Supplementary material

Section 1: List of figures, tables and abbreviations

List of figures

Chapter 3
   Figure 1: Yields of circulating DNA compiled from published studies 19

Chapter 4
   Figure 1: Products of PCR on circulating DNA... 32
   Figure 2: Products of colony PCR of pEZSeq transformants... 33

Chapter 5
   Figure 1: Size and chromosomal distribution of sequences 38

Chapter 6
   Figure 1: Size distribution of sequenced DNA fragments 43
   Figure 2: G+C content of the sequenced DNA fragments 43
   Figure 3: Chromosomal distribution of sequences... 44
   Figure 4: Chromosome ideograms with positions of unique sequences... 44
   Figure 5: Rate (%) and type of mutations detected in sequenced circulating DNA 45

Supplementary material
   Figure 1: Flow diagram of experimental work 58
   Figure 2: Bioanalyzer gel image of ligated circulating DNA 61
   Figure 3: 12 Colony PCR products for sample 17 62
   Figure 4: KapaFidelity PCR products on agarose gel electrophoresis 63
   Figure 5: DNA quantification with the NanoDrop Fluorospectrometer 66
   Figure 6: Bioanalyzer gel to confirm tagging of circulating DNA 68
   Figure 7: Electropherograms of the samples in the bioanalyzer gel in Fig 6 69

List of tables

Chapter 3
   Table 1: Frequency of detection of mutations and microsatellite tumor markers 21
   Table 2: Frequency of detection of methylated tumor DNA markers 23
   Supplementary table: Quantification of circulating DNA 87

Chapter 4
   Table 1: Comparison of the different DNA Isolation Methods 33
   Table 2: An Example of BLAST Results 34
   Table 3: Repeats in Control and Patient Sequences 34
   Table 4: Table of Placement of Sequences on Chromosomes 35

Chapter 5
   Table 1: Repeat analysis with Repeatmasker 38

Chapter 6
   Table 1: Repeat analysis of unique sequences by RepeatMasker 45

Supplementary material
   Table 1: Sequences of barcoding adapters 65
   Table 2: Concentrations of tagged free circulating DNA 67
Abbreviations

AMP: Ampicillin
APC: Adenomatous polyposis coli
ATP: Adenosine triphosphate
BLAST: Basic Local Alignment Search Tool(s)
Bp: Base pair(s)
CDH1: E-cadherin
DAPK1: Death-associated protein kinase
DNA: Deoxyribonucleic acid
dNTP: Deoxyribonucleotide triphosphate
F-MSP: Fluorescent methylation specific PCR
FQ-MSP: Fluorogenic (taqman) real-time methylation specific PCR
GSTP1: Glutathione S-transferase P1
Kb: Kilo base pairs
LINEs: Long Interspersed elements
MER: Medium Reiterated repeats
MGMT: O6-methylguanine-DNA-methyltransferase
MiRs: Classic, tRNA-derived SINEs
MLH1: mutL homolog 1
Mnc: Mononuclear cells
MRE PCR: Methylation-sensitive restriction enzyme-related PCR
MSP: Methylation specific PCR
NCBI: National Center for Biotechnology Information
Nlt: Normal lung tissue
PCR: Polymerase Chain Reaction
QMSP: Quantitative methylation specific PCR
RAR-β2: Retinoic acid receptor β2
RASSF1A: RAS association domain family protein 1 A
ROS: Reactive oxygen species
SINE: Short Interspersed Elements
SMSP: Modified semi-nested methylation specific PCR
TIMP3: Tissue inhibitor of metalloproteinase-3
TMS1: Target of methylation-induced silencing
Section 2: Materials, methods and supporting results

Blood collection and processing

Isolation of circulating DNA

Cloning and sequencing for pilot study

Blunting and phosphorylation of isolated DNA fragments

Vector ligation

Ligation PCR

Transformation into competent cells

Colony PCR

DNA sequencing

Preparation of DNA for 454 GS FLX Sequencing

Blunting and phosphorylation of isolated DNA fragments

Vector ligation

High fidelity ligation PCR

Blunting and phosphorylation

SureClean cleanup

Ligation of barcoding adapters

SPRI cleanup

Adapter fill-in

Quantifying and pooling of tagged samples (according to Table 2)

Dephosphorylation and restriction digestion

Verification of tagging and restriction digestion efficiency

GS FLX sequencing

Figure 1: Flow diagram of experimental work
i. **Blood collection and processing**

Ethics approval for this study was obtained from the Ethics Committee of the North-West University, reference number 05M12. AMPATH, PathCare or Lancet drew blood by venipuncture in three EDTA tubes (4ml BD Vacutainer EDTA tubes from Transgen, REF# 368861) for each prostate cancer patient before treatment and age matched control after informed consent was received (see supplementary material: Part 3 for an example of the informed consent form), for each control one coagulation tube for PSA count was also drawn at the time of EDTA blood collection, determination of the PSA count was performed by AMPATH. Blood was subject to a single centrifugation step at 3200 rpm (1600 x g) for 10 minutes after which plasma was carefully transferred to 2 ml microcentrifuge tubes and frozen at -20°C until isolation of circulating DNA. Blood was kept at 4°C and processed within two hours of collection.

ii. **Isolation of circulating DNA**

Circulating DNA was isolated using a phenol-chloroform method (Anker and Stroun, 2001a). Plasma from patients and controls were thawed and centrifuged at 16000 x g for 10 minutes, 1.5 ml of the top clear part of the plasma was mixed with 1.5 ml of a SDS (1%), proteinase K (2 mg/ml) and TE (pH9) solution in a 15 ml tube and incubated overnight in a waterbath at 55°C for protein digestion to take place. Following incubation, 1.5 ml of 1:1 mixture of chloroform and fresh water saturated phenol was added, the mixture was vortexed for 30 seconds and centrifuged at 2500 rpm for 10 minutes. The top water phase was transferred to a clean 15 ml tube and another phenol-chloroform extraction was performed. After the second extraction the interphase was usually transparent and DNA precipitation could be performed. The aqueous phase was transferred to a new tube and 3 μl glycogen (20 mg/ml) was added as co-precipitant, 1 ml ammonium acetate (7.5 M) to adjust for salt concentration and 8 ml ice cold ethanol (100%) was added in this order. This mixture was carefully mixed and DNA was allowed to precipitate at -20°C overnight. This was followed by centrifugation at 6000 rpm (5017 x g) in a fixed angle rotor. The supernatant was decanted and fresh 70% ethanol was added to the pellet and mixed carefully followed by another centrifugation at 6000 rpm for 12 minutes. The supernatant was decanted again and the pellet was allowed to air-dry, then 100 μl of TE buffer was added. The isolated DNA was kept overnight at 4°C to resuspend after which it was transferred to a 1.5 ml microcentrifuge tube. Isolated circulating DNA was stored at 4°C until further use.
iii. Cloning and sequencing in the pilot study

Addition of primer binding sites to the isolated sequence fragments of circulating DNA was done as described previously (van der Vaart, 2005). Since the method was also used in this study it is given below.

Blunting and phosphorylation of isolated DNA fragments

The ends of circulating DNA are most likely ragged and of random configuration, thus it needs to be blunted in order for it to be ligated into a blunt vector. Blunting was performed by using T4 DNA polymerase (Fermentas, #EP0061) with a protocol slightly altered from that of the supplier: 2 μl 10x Tango buffer (Fermentas), 10 μl circulating DNA, 7.3 μl water and 0.5 μl T4 DNA polymerase was mixed in an microcentrifuge tube and incubated at 11°C for 5 minutes. This was followed by the addition of 0.2 μl of nucleotide solution mix (10 mM ea. dNTP, New England BioLabs® Inc. #N0447S) after which the mixture was incubated for a further 15 minutes at 21°C (room temperature). The enzyme was heat inactivated by keeping it at 70°C for 10 minutes. The blunted DNA was kept on ice until 0.55 μl ATP (40 μM) and 1 μl T4 Polynucleotide kinase (Fermentas, #EK0031) were added, after which it was vortexed briefly, spun down and incubated at 37°C for 30 minutes. Phosphorylation of the 5'-ends of the blunt double stranded circulating DNA had to be done in order for it to be able to ligate to the blunt dephosphorylated ends of the vector. The reaction was cleaned with a single chloroform extraction.

Vector ligation

Blunted and phosphorylated circulating DNA was ligated to the pEZSeq™ vector according to the protocol of the pEZSeq™ Blue/White Cloning kit from Lucigen® Corporation (# 40464-2). Briefly, 6.5 μl chloroform extracted DNA was mixed with 2.5 μl 4x pEZSeq™ vector premix and 1 μl CloneSmart DNA ligase supplied with the kit. This mixture was incubated at room temperature (21°C) for 2 hours after which the enzyme was deactivated at 70°C for 15 minutes. The mixture was then cooled to room temperature for 15 seconds, followed by cooling on ice for 15 seconds after which it was centrifuged at 13500 x g (12000 rpm) for 1 minute in a table top microcentrifuge centrifuge. The ligated product was stored at 4°C until further use. This vector was chosen because it produces 99% recombinant colonies and is supplied with blunt, dephosphorylated ends. This vector contributes 127 bp (including the sequences of the M13 primers) to the sequenced product.

Ligation/coiiony PCR

Amplification of the ligated circulating DNA was performed to verify that ligation took place and to estimate the size of the ligated products. Colony PCR was subsequently performed to verify
that the colonies contain inserts and to estimate the size of the respective inserts. For both a master mix of the following was made: 5 µl 5x colorless GoTaq Flexi buffer (Promega), 6.7 µl MgCl₂ (25mM, Promega), 0.5 µl dNTP mix (10 mM ea., New England BioLabs® Inc. #N0447S), 1 µl of both forward and reverse M13 primers (10 µM, manufactured by Inqaba Biotech) and 0.25 µl GoTaq® polymerase (Promega #M8305). For ligation PCR a further 8.55 µl HPLC water (BDH #152736D) and 2 µl ligated DNA was added before amplification for a total reaction volume of 25 µl. For colony PCR 10.55 µl water was added to each PCR tube and a single colony was touched with a pipette tip which was subsequently rinsed in the water after which 14.45 µl master mix was added, the reaction was mixed and amplified according to the following reaction conditions:

3 min at 94°C
25 cycles of:
   30 sec at 94°C
   30 sec at 55°C
   1 min at 72°C
10 min at 72°C

The resulting PCR products were analyzed using agarose gel electrophoresis or in an Agilent bioanalyzer 2100 with the DNA 1000 kit (#5067-1504), see Figure 2.

![Figure 2: Bioanalyzer gel image of ligated circulating DNA](image)

The dark band at ~280bp show that the majority of ligated DNA is more or less 150bp, the faint band at 127bp represent empty vector, thus vector that closed on itself. This figure is representative of all the samples analyzed.
Transformation of competent cells

Chemically competent SURE cells where thawed on wet ice, 120 μl cells were pipetted into a chilled 15 ml tube and 5 μl ligated DNA was added, it was stirred with the pipette tip and kept on ice for 30 minutes, after which it was heat shocked by keeping it 45 seconds in a water bath at 42°C, after the heat shock it was kept on ice for 2 minutes and then 200 μl SOC at room temperature were added. The cells were allowed to recover for 1 hour at 37°C in a shaking incubator at 200 rpm. Transformed cells were plated out on YT +AMP agar plates, 155 μl cells per plate and 2 plates per individual. Plates were incubated overnight at 37°C. Colonies were transferred to master plates and colony PCR were performed as described above. Frozen stocks of the colonies were made by inoculating 3-5 ml of LB +Amp culture medium with a single colony, inoculated media was incubated overnight at 37°C after which 700 μl of the culture medium and 300 μl 50% glycerol were mixed and quick frozen for each colony.

![Image](image.png)

Figure 3: 12 Colony PCR products for sample 17

In this figure, the different sizes of cloned circulating DNA can be observed

DNA sequencing

Master plates of the colonies were sent to Inqaba Biotec for sequencing, where high fidelity colony PCR with Pfu Polymerase (Fermentas, #EP0501) was performed on 12 colonies per individual (only sample 7 to 18 – half of which were cancer patients and the other half age matched healthy controls, were used in the pilot study). PCR products were sequenced on a ABI 3130XL sequencer using M13 sequencing primers, a total of 126 sequences were obtained and analyzed.

iv. Preparation of circulating DNA for 454 sequencing™ (GS FLX)

Circulating DNA of 12 prostate cancer patients and 10 age matched controls were tagged in order for it to be pooled and sequenced on one region of a 454 GS FLX picotitter plate. Blunting, phosphorylation and vector ligation were performed as described above and the ligated DNA fragments were amplified using high fidelity ligation PCR to increase the amount of circulating DNA in order to have enough DNA to be able to have a pooled yield of at least 2 μg.
of tagged DNA. Tagging of circulating DNA was done according to the protocol of Meyer et al. (2008) with a few minor alterations, mainly to the cleanup procedures in order to increase the yield.

High fidelity PCR of ligated DNA fragments
Amplification of the ligated circulating DNA was performed with KapaFidelity polymerase (KapaTaq Trial Kit #CT04000065 Code: KK1013) according to the following protocol: a master mix of 35.5 µl HPLC water, 5 µl 10x reaction buffer, 1.5 µl dNTP mix, 3 µl MgCl (all supplied with KapaFidelity polymerase kit), 1 µl of both forward and reverse M13 primers (10 µM) and 1 µl KapaFidelity polymerase were mixed, 48 µl of the master mix was combined with 2 µl of ligated circulating DNA and amplified as follows:

2 min at 95°C
25 cycles of:
15 sec at 98°C
15 sec at 55°C
30 sec at 72°C
1 min at 72°C

The resulting PCR products were analyzed using agarose gel electrophoresis.

![Figure 4: KapaFidelity PCR products on agarose gel electrophoresis](image)

This figure shows that most of the ligated circulating DNA is in the size range between 250-350 bp, some of the products show a smaller band at 127bp which represents empty vector.

KapaFidelity PCR were repeated twice for each sample (see Figure 4) and the two products for each sample (~100ul in total) were concentrated in a Thermo Speedvac at medium heat for ~45 minutes after which 19 µl TE buffer were added to each sample.
Blunting and phosphorylation
A master mix was prepared according to the protocol of Meyer et al. (2008): 7.8 μl HPLC water, 4 μl 10x Buffer tango (Fermentas), 1 μl Bovine serum albumin (10 mg/ml, Boehringer Mannheim # 591987), 0.4 μl dNTPs (10 mM ea.), 4 μl ATP (10 mM), 2 μl T4 Polynucleotide kinase (10 U/μl, Fermentas, #EK031) and 0.8 μl T4 Polymerase (5 U/μl, Fermentas, #EP0061) was mixed together and 20 μl was mixed with 20 μl of each KapaFidelity PCR product. The reaction was incubated in a thermal cycler for 15 minutes at 12°C followed by 15 minutes at 25°C. Samples were cleaned with SureClean Plus from Bioline.

SureClean Cleanup
SureClean cleanup proved to be a quick and efficient cleanup procedure which gave a much better yield than the SPRI cleanup suggested by Meyer et al. (2008). In short, 6 μl of Pink Co-Precipitant (Bioline, #BIO-37076) was added to each sample and mixed after which 46 μl of SureClean solution (Bioline, #BIO-37042) was added and mixed thoroughly by vortexing, samples was incubated for 10 minutes at room temperature followed by centrifugation at 14000 x g for 15 minutes, the supernatant was aspirated and discarded, 92 μl of fresh 70% ethanol was added to the pellet and it was vortexed for 10 seconds. The sample was centrifuged again at 14000 x g for 10 minutes and the supernatant was discarded, the pellet was allowed to dry and it was resuspended in 20 μl of EB buffer.

Oligonucleotides
In order to convert the single stranded oligonucleotides listed in Table 1 to double stranded tags (barcoding adaptors), the following protocol was followed. Oligos were dissolved in TE buffer to obtain stock solutions of 500 μM. Barcoding adapters were created by mixing each oligo (400μM) with 1x T4 DNA ligase buffer in a PCR tube and incubating it for 10 seconds at 95°C followed by a ramp to 25°C at a rate of 0.1°C per second. Tags were centrifuged and frozen at -20°C until further use.

Ligation of barcoding adapters (tags)
Circulating DNA of each patient and control were ligated to a different barcoding adapter (see Table 1) in order to be able to identify the circulating DNA of each individual after sequencing, this was done by first adding 2 μl of a different barcoding adapter to each sample and vortexing it to ensure that it is properly mixed. A master mix was prepared by mixing 9 μl HPLC water, 4 μl T4 ligase buffer (Fermentas), 4 μl PEG-4000 (50%, Fermentas) and 1 μl T4 ligase (5 U/μl, Fermentas, #EL0331). Samples mixed with barcoding adapters in 500 μl Lo-Bind Eppendorf tubes (Eppendorf #022431005) were each mixed with 18 μl of master mix and incubated for one hour at 22°C in a thermal cycler.
Table 1: Sequences of barcoding adapters

<table>
<thead>
<tr>
<th>Tag name</th>
<th>Tag Sequence*</th>
</tr>
</thead>
<tbody>
<tr>
<td>454 Tag 1 (sample 1)</td>
<td>CACGTAGCCGCGGCTGACGCTG</td>
</tr>
<tr>
<td>454 Tag 2 (sample 2)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 3 (sample 3)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 4 (sample 4)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 5 (sample 5)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 6 (sample 6)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 7 (sample 7)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 8 (sample 8)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 9 (sample 9)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 10 (sample 10)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 11 (sample 11)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 12 (sample 12)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 13 (sample 13)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 14 (sample 14)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 15 (sample 15)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 16 (sample 16)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 17 (sample 17)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 18 (sample 18)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 19 (sample 19)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 20 (sample 20)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 21 (sample 21)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 22 (sample 22)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 23 (100bp control fragment)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
<tr>
<td>454 Tag 24 (Extra)</td>
<td>CAGCTGTGCCCGGACCAGCTG</td>
</tr>
</tbody>
</table>

*Sequences for the barcoding adapters were kindly supplied to us by Mathias Meyer and single stranded oligo’s were synthesized by Inqaba Biotech.

**SPRI purification**

Samples were cleaned with the AMPure SPRI PCR Purification kit (Agencourt, #000130) to eliminate unbound barcoding adapters according to the altered protocol provided by Meyer et al. (2008). After resuspension of SPRI beads, 72 µl beads were added to each sample and mixed by vortexing it several seconds. Samples were left to stand for 5 minutes at room temperature and then centrifuged to collect all liquid at the bottom of the tube. The tubes were transferred onto a 96 well ring magnetic plate (Agencourt #000219), and beads were collected onto the walls of the tubes for 2 minutes, all of the supernatant was aspirated and discarded. Beads were then washed while the tubes were still in the magnetic plate by adding 150 µl fresh 70% ethanol and aspirating it after 1 minute incubation at room temperature. Washing was repeated once more. Beads were left to air-dry for 30 to 60 minutes after which 22 µl of EB buffer was added to elute DNA, tubes were vortexed to resuspend the beads and centrifuged to collect all liquid at the bottom of the tube, after which it was reinserted into the magnetic ring plate. After 1 minute all liquid was aspirated without disturbing the beads and the aspirated samples were collected in clean tubes.
Adapter fill-in

A mastermix was prepared for the adapter fill-in reaction by mixing 12.5 μl water, 4 μl 10x Thermopol buffer (New England Biolabs), 1.25 μl dNTPs (10 mM ea.) and 1 μl Bst Polymerase (8 U/μl, New England Biolabs, #M0275S), 18 μl of the master mix was added to each 22 μl sample and incubated at 37°C for 20 minutes in a thermal cycler. SureClean cleanup was done once more and the DNA pellets were resuspended in TE buffer and stored at -20°C.

---

Figure 5: DNA quantification with the NanoDrop Flurospectrometer

PicoGreen standard curve (bottom) and an example of a sample read on the NanoDrop® ND-3300 Flurospectrometer

Quantifying and pooling of tagged samples

Sample concentrations were measured on a NanoDrop® ND-3300 Fluorospectrometer at Inqaba Biotec with the PicoGreen® Assay for dsDNA. Samples were diluted 20 or 40 times in order for it to be measured in the linear range of the standard curve (between 0.1 and 1000 ng/ml, see Fig 5). An average between three readings for each sample was calculated and is reported in Table 2. Tagged samples were pooled in mass ratios in order to get a sample pool with a total concentration of ~3 μg, all samples had similar mean fragment sizes as shown by ligation PCR. The sample pool had a total volume of 110.37 μl and was concentrated using the MinElute PCR Purification Kit (Qiagen, #28006), two alterations was made to the kit
protocol, a second dry spin was added after turning the tube 180° in the rotor and spinning it for an additional 30 seconds after which the DNA was eluted in 20 µl of EB buffer.

<table>
<thead>
<tr>
<th>Sample nr</th>
<th>Concentration (ng/µl)</th>
<th>Volume in µl added to DNA pool [3080ng]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample 1 (C)</td>
<td>35.75</td>
<td>3.92</td>
</tr>
<tr>
<td>Sample 2 (C)</td>
<td>20.56</td>
<td>6.81</td>
</tr>
<tr>
<td>Sample 3 (C)</td>
<td>22.93</td>
<td>6.10</td>
</tr>
<tr>
<td>Sample 4 (C)</td>
<td>20.75</td>
<td>6.75</td>
</tr>
<tr>
<td>Sample 5 (C)</td>
<td>25.08</td>
<td>5.58</td>
</tr>
<tr>
<td>Sample 6 (C)</td>
<td>27.64</td>
<td>5.07</td>
</tr>
<tr>
<td>Sample 7 (C)</td>
<td>28.16</td>
<td>4.97</td>
</tr>
<tr>
<td>Sample 8 (C)</td>
<td>21.84</td>
<td>6.41</td>
</tr>
<tr>
<td>Sample 9 (C)</td>
<td>24.46</td>
<td>5.72</td>
</tr>
<tr>
<td>Sample 10 (C)</td>
<td>29.18</td>
<td>4.80</td>
</tr>
<tr>
<td>Sample 11 (C)</td>
<td>30.26</td>
<td>4.68</td>
</tr>
<tr>
<td>Sample 12 (C)</td>
<td>36.98</td>
<td>3.79</td>
</tr>
<tr>
<td>Sample 13 (H)</td>
<td>35.45</td>
<td>3.95</td>
</tr>
<tr>
<td>Sample 14 (H)</td>
<td>32.35</td>
<td>4.33</td>
</tr>
<tr>
<td>Sample 15 (H)</td>
<td>27.49</td>
<td>5.09</td>
</tr>
<tr>
<td>Sample 16 (H)</td>
<td>33.98</td>
<td>4.14</td>
</tr>
<tr>
<td>Sample 17 (H)</td>
<td>33.64</td>
<td>4.24</td>
</tr>
<tr>
<td>Sample 18 (H)</td>
<td>30.95</td>
<td>4.52</td>
</tr>
<tr>
<td>Sample 19 (H)</td>
<td>30.44</td>
<td>4.60</td>
</tr>
<tr>
<td>Sample 20 (H)</td>
<td>28.36</td>
<td>4.94</td>
</tr>
<tr>
<td>Sample 21 (H)</td>
<td>27.81</td>
<td>5.09</td>
</tr>
<tr>
<td>Sample 22 (H)</td>
<td>28.38</td>
<td>4.99</td>
</tr>
<tr>
<td>100 bp control fragment</td>
<td>22.06</td>
<td>no value</td>
</tr>
</tbody>
</table>

The first 12 samples were prostate cancer patients (C), and samples 13 to 22 were age matched healthy controls (H).

**Dephosphorylation and restriction digestion**

In order to dephosphorylate the ends of the tagged circulating DNA the following reaction mix was prepared: 14 µl HPLC water, 4 µl 10x NEBuffer 3 (New England Biolabs), 20 µl sample pool and 2 µl calf intestinal alkaline phosphatase (CIP, 10 U/µl, New England Biolabs, #M0290S) was mixed gently and incubated at 37°C for 30 minutes after which an extra 1 µl CIP was added and the mixture was incubated for a further 30 minutes at 37°C. The dephosphorylated sample pool was cleaned with the MinElute PCR Purification kit as described above. Restriction digestion of the sample pool was performed by mixing 15.5 µl HPLC water, 4 µl 10x Universal buffer (Stratagene), 20 µl sample pool and 1 µl Srf I (10 U/µl, Stratagene, #501064) and incubation of the mix at 37°C for 1 hour after which another 1 µl of Srf I was added and it was incubated for another hour at 37°C. MinElute PCR Purification was repeated once more.
Dephosphorylation of the DNA was to prevent any DNA that was not cut during the Srf I digestion from binding to the A or B adapters during subsequent GS FLX sequencing library preparation. DNA that was cut by Srf I now have only half of the original tag sequence which is phosphorylated and thus able to be ligated to the dephosphorylated A and B adapters.

Verifying tagging and restriction digestion efficiency

Circulating DNA was analyzed using the Agilent Bioanalyzer with the Agilent DNA 1000 kit (#5067-1504) according to the manufacturer's protocol. Untagged and tagged circulating DNA samples were run side by side to the Srf I cut sample pool to verify the tagging and restriction digestion efficiency (see Figure 6 and 7).

Before commencing with the GS FLX sequencing a cloning and conventional sequencing test run was performed to ensure correct binding of the tags. This was successful and 454 sequencing was subsequently performed.

![Figure 6: Bioanalyzer gel to confirm tagging of circulating DNA](image)

This figure illustrates that tagging of circulating DNA was successful, when the size of free DNA for the tagged and untagged samples are compared it is clear that there is a size difference of ~10 bp while the sample pool of Srf I restriction digested circulating DNA has a size between that of the tagged and untagged samples which confirms that half of the sample tag (~20 bp) were removed by restriction digestion.
Figure 7: Electropherograms of the samples in the bioanalyzer gel in Fig 6

This is another way of presenting bioanalyzer results, DNA concentrations can be measured using the area under the peak for each sample and extra peaks that are not always visible on the gel can be visualized in this way, however sample comparison is difficult using this view. The first and last peak in each electropherogram is the marker peaks, used as internal standards to align all samples to the size marker or ladder.

454 Sequencing

The pooled Srf I cut sample was transported to Inqaba Biotec on dry ice and a 454 sequencing library was prepared within one week. The library preparation protocol was followed from the adaptor annealing step.

Sequencing results was sifted for sequences containing the sequence tags, a total of 8607 sequences were obtained of which 52% were from cancer patients and 48% from controls. These sequences were screened for vector sequences and after removal of empty vector 5825 sequences remained. These sequences were analyzed with a variety of available bioinformatics tools. Results are described in Chapter 6.
PART 1: General Project Information

The part below provides you with additional information about the nature of this project and the legal agreements that apply. If you have any questions or concerns, please contact the project organizers.

1. Title of the Project:
   Characterization of circulating free DNA in healthy and diseased individuals

2. Institution / Subject group / Institute:
   North-West University, School for Biochemistry

3. Nature & contact details of Project Staff:
   (Please note: This is not a full list of staff involved. Additional members are not mentioned here. For more information, please contact the project organizers.)

4. Name & contact details of Independent Person:
   (These persons are usually your second line of contact. They are not involved in the project itself but are available to assist you in case of any problems.)

5. Conflicts of Interest:
   (Please indicate if there are any potential conflicts of interest that might influence the results of the project.)

6. Consent:
   (Please read and sign the consent form before participating in the project.)

FORM FOR INFORMED CONSENT
for participation in scientific projects

Version 1.01 (Jun 2007)

CONFIDENTIALITY/VEREISTELING

This document contains confidential information and superseded by the agreement. The Ethical Committee of the North-West University and the defendant have no obligation to consult the patient or any other person involved in the project. The information provided is for the project only and may not be used for any other purpose except as stated in the consent form.

Every human participant in any project is the subject of research or education and, therefore, the authorized parties (i.e., the patient, the research team, and the authorized representatives of the patient) must be fully informed about the project and sign a form of informed consent stating their participation or withdrawal. The content of this document is not to be disclosed to any other person except as stated in the consent form.

The informed consent form consists of three parts:
   Part 1: General project information
   Part 2: Confidentiality
   Part 3: Information about you and the project
   Part 4: Consent

You are being asked to participate in a clinical research study where circulating DNA will be isolated from blood samples for characterization. The characteristics of circulating DNA might be specific to certain genomes or even specific tissues that might change during processing. It might also have prognostic or predictive values in determining outcome or treatment.

Your participation in this study is completely voluntary, so you will participate only if you agree. If you have any questions or concerns, please contact the ethical committee. The information contained herein is provided for your personal use and may not be disclosed to any other person except as stated in the consent form.

All we need from you for this study is three tubes (10 ml or approximately one tablespoon) of blood. You don't have to participate in any tests or claims, any medication for purposes of this project.
5.1 You are requested to take part in the project, and may have the following questions:

5.1.1 What are the key requirements that persons must meet to be able to take part in the project? Why and how was I chosen?

Persons in the patient group have to be diagnosed with cancer, but are excluded from the study if any treatment has been received. Persons in the control group have to be healthy and in no medication.

5.2 What is the purpose of the project?

The aim of this study is to determine whether circulating DNA can be selected as a biomarker(s) for specific malignancies.

5.3 What will be expected of me as a participant in this intervention/programmes will I have to take part? What exactly will it involve?

Three EDTA tubes (10 ml or approximately one tablespoon) of blood will be collected from you. You won't be required to do or take anything else.

5.4 What are the potential benefits and potential dangers and/or potential permanent consequences (however negligible) that may arise in this project lacking?

You may experience some side effects of blood collection which may include pain, swelling, bruising around the vein that is used to draw the blood sample, discomfort and/or anxiety may also occur.

5.5 What precautions have been taken to protect me as participant?

A qualified nurse will be collecting the blood.

5.6 How long am I expected to be involved in the project (e.g. number and duration of visits)

Blood will only be collected once and no further participation is required from you.

5.7 What direct benefits can I expect from the project? What renumerations (monetary or services) I expect for my participation?

You will not be paid for your participation in this study and no direct benefit can be expected.

5.8 What potential general benefits are there for the broader community which may result from this project?

You and society may benefit from any experimental treatment, although there may not be direct medical benefit to you. Others may benefit from the information learned in this study. This study may help to develop a new therapy for other mild viral conditions.

5.9 How will the findings of this project (general results, as well as individual results) be made available or can be used?

You may receive information about the results of the trial at any time but individual results in this group study is meaningless, combining the results may lead to discovery of a biomarker which may be used in diagnostic or prophylaxis.

5.10 What measures have been taken to handle and store my data confidentiality?

Your research records are confidential unless disclosure is required by law. Your name or other personal identifying information will not be used in any reports or publications resulting from this study. Data from this study will be reported in a way that ensures statistical analysis and recorded as such by the North-West University.

IMAU Ethics Committee: Form for informed consent

CONFIDENTIAL
PART 2: General Principles

To the signatory of the consent contained in Part 4 of this document:

You are invited to take part in the research project as described in Part 1 of this informed consent form. It is important that you read and understand the following general principles, which are applicable to all participants in our research projects:

1. Participation in the project is completely voluntary and no pressure, however subtle, may be placed upon you to take part.

2. It is possible that you may not derive any benefit personally from your participation in the project, although the knowledge that may be gained by means of the project, may benefit others or the community. In exceptional cases where you do receive personal financial benefit, these are usually, in the normal course of participation and for personal assistance (e.g., minor) during your participation. You may not be asked to participate.

3. You are free to withdraw from the project at any time, without stating reasons, and you will be recorded as having withdrawn your consent.

4. By agreeing to take part in the project, you are also giving consent for the data that will be generated to be used by the researchers for scientific purposes as they see fit, with the caveat that it will be confidential and that your name will not be linked to any of the data without your consent.

5. The Ethics Committee, Murdoch University, Department of Health and the Court of Law may request access to information to ensure the confidentiality of all identities, for the sake of participants and the public.

6. You will be given access to your own data upon request, unless the Ethics Committee has provided temporary non-disclosure (in the latter case, the reasons in Part 1 will be explained to you).

7. A summary of the nature of the project, the potential risks, factors that may cause you possible inconvenience or discomfort, the benefits that can be expected and the known or unforeseeable consequences that your participation in this project may have for you as participant, are set out for you in Part 1 thereof.

8. You are encouraged to ask the Project Leader or co-workers any questions you may have regarding the project and the related procedures at any stage. They will gladly answer your questions. They will also discuss the project with you in detail.

9. If you are a minor, the written consent of your parent or legal guardian is required before you participate in this project, as well as (in writing if possible) your voluntary assent to take part—no consent may be placed upon you.

10. The project operations are always necessary to your well-being and actions taken will always place your interests before those of the project.

11. No project may be commenced before it is approved by the Ethics Committee. Furthermore, the Project Leader must report any detrimental effects experienced during the implementation of the project to the Ethics Committee if any unforeseen serious detrimental effects are observed during the project, it may be necessary to terminate the project immediately.

INQUIE Ethics Committee Form for Informed Consent

CONFIDENTIAL
PART 3: Information about you as participant

This part of this form is voluntary and you don't have to answer anything you don't want to. However, you are likely requested to answer all the questions as information you provide here may be critical in the analysis of the subsequent data and thus the reliability of the project.

1. Gender:
2. Ethnicity and race:
3. Weight:
4. Race:
5. Do you smoke?
6. Are you pregnant?
7. Were you diagnosed with any chronic condition? If so, please specify:
8. Do you use any medication? If so, please specify:
9. Are you currently in good health?
10. Any other remarks that might be of importance?

PART 4: Consent

Title of the Project:
Characterization of circulating free DNA in healthy and diseased individuals

I, the undersigned

Full name & Surname

have read the preceding premises in connection with the project, as discussed in Part 1 and Part 2 of this informed consent form. I have also heard the explanation thereon, and I declare that I understand it. I have also initialed every page of Part 1 and Part 2. I have given the opportunity to discuss relevant aspects of the project with the Project Head and I hereby declare that I am taking part in the project voluntarily.

Signature of Participant

Signed at

Place of Signature

WITNESSES

Signature of Witness 1

(translator, if applicable)

Signed at

Place of Signature

Signature of Witness 2

Signed at

Place of Signature

NWU Ethics Committee Form for informed consent

CONFIDENTIAL
Credits

Consulted on request of the NWU Ethics Committee by Prof. Christian E. Bink (PhD)

Advisory panel: Prof. Hester Klappe, Dr. Alan MacLeod, Prof. Niel Malan, Dr. David van der West, Dr. Francois van der Westhuizen, Mr. Michiel Vlok and Prof. M. H. Wehrly

Other credits: Many individuals contributed in various ways to formulate, develop and complete previous ethics application forms of the University, of which excerpts were used in formulating some of the contents of the current ethics application form.

©Copyright 2017, North-West University, all rights reserved.
Section 4: Conference proceedings

Abstracts and posters were prepared for "The Circulating Nucleic acids in plasma or serum V (CNAPS-V) Symposium" proceeding from the 23rd – 27th of August 2007 in Moscow, Russia. Posters as well as a talk were presented by the first author. These posters were also presented at the Bio-08 (South African Society of Biochemistry and Molecular Biology) conference proceeding form the 21st – 25th of January 2008 in Grahamstown, South-Africa. Posters were presented by the first author.

CIRCULATING DNA: ITS ORIGIN AND FLUCTUATION

M. van der Vaart and P. J. Pretorius

School for Biochemistry, North-West University, Potchefstroom Campus,
Private Bag X6001, South Africa, 2520
E-mail: manlesh.vandervaart@nwu.ac.za, pjet.pretorius@nwu.ac.za

Background: Elevated amounts of circulating DNA can be found in the blood of cancer patients compared to healthy individuals and a variety of diseases has been shown to present with increased amounts of circulating DNA.

Methods: Relevant literature were considered and compared in an effort to explain the origin and the reason for the fluctuation of circulating DNA.

Results: The general notion that apoptosis is the major source of circulating DNA is unjustifiable because several observations, like macrophage engulfment of apoptotic cells, loss of apoptotic ability by proliferating cells and DNA release by living cells, demonstrate that apoptosis may only have a minor contribution to circulating DNA in the blood of healthy as well as most diseased individuals.

Conclusion: We propose that DNA release by living cells is the main source of circulating DNA in the blood. The occurrence of increased amounts of circulating DNA may be the result of a disturbance in the equilibrium between the release of DNA by living cells and the mechanisms involved in the clearance of these molecules.

A METHOD FOR CHARACTERIZATION OF CIRCULATING DNA

M. van der Vaart and P. J. Pretorius

School for Biochemistry, North-West University, Potchefstroom Campus,
Private Bag X6001, South Africa, 2520
E-mail: manlesh.vandervaart@nwu.ac.za, pjet.pretorius@nwu.ac.za

Background: Although much work has been done in the field of circulating DNA, no definitive information on sequencing data of total circulating DNA is available. Characterization of total circulating DNA by sequence analysis may give valuable information about the origin and function of these nucleic acid molecules.

Methods: Circulating DNA was isolated from plasma of healthy individuals as well as cancer patients with various methods and was cloned into a blunted cloning vector. Resulting colonies were sequenced and analyzed.

Results: It was found that the majority of the DNA that ligated into the vector was about 200 bp in length. Sequence analysis revealed that the cloned circulating DNA consists partly of currently uncharacterized human genomic sequences or when human repeat sequences were masked it matched with sequences present in contigs containing known genes situated at various distances from the identified targets. In addition to the presence of large repeats a variety of Alu repeat sequences were observed. Preliminary results showed that more Alu repeats are present in the plasma of normal individuals than in patient material. None of the gene sequences reported in the literature to be part of circulating DNA, e.g. P53, the Ras family, β-Globin or β-actin was observed in our analysis.

Conclusion: Cloning and sequencing of free circulating DNA was successful and this first attempt on characterizing free circulating DNA by analysis of sequence data not only confirmed previous results, e.g. size and Alu content, but also revealed the large variety of sequences in circulating DNA. Further characterization of circulating DNA may be beneficial in diagnosis and prognosis and may also contribute to determining the source and function of circulating DNA.
CIRCULATING DNA: ITS ORIGIN AND FLUCTUATION

Maniesh van der Vaart and Pieter J. Pretorius

School of Biochemistry, North-West University, Private Bag X6001, Potchefstroom, 2520, SOUTH AFRICA.

Abstract

Background

Estimated amounts of circulating DNA can be found in the blood of cancer patients compared to healthy individuals and a variety of diseases has been shown to present with increased amounts of circulating DNA.

Methods

Changes in extracellular DNA were considered and compared in an effort to explain the origin and the reason for the fluctuation of circulating DNA.

Results

The general notion that apoptosis is the major source of circulating DNA is incorrect because the evidence for this is based on a lack of understanding of the role of circulating DNA. The role of apoptosis in the release of circulating DNA in the bloodstream of healthy as well as diseased individuals is not confirmed by the results presented in this study.

Conclusion

We propose that DNA release by dying cells is the main source of circulating DNA in the blood. The increase of increased amounts of circulating DNA may be the result of a disturbance in the equilibrium between the release of DNA by dying cells and the mechanism involved in the clearance of these molecules.

Rationale

Almost every paper on circulating DNA state that apoptosis and necrosis are the main source of free circulating DNA in serum and plasma. A large number of apoptosis DNA degradation pathways which degrades circulating DNA to oligodeoxynucleotides (ODN) and subsequently to nucleotides. The DNA degradation is catalyzed by DNase I and II. However, this view has been challenged by Moore et al. who have shown that apoptotic bodies can be observed for actively released DNA (8).

Although the mechanisms are not fully understood, the contents of cells dying by apoptosis are rapidly degraded by professional phagocytes or neighboring cells (9) and the DNA is consequently digested by Chae II in lymphocytes (10). Thus, the possibility exists that DNA fragments released by apoptosis are removed before the production of circulating DNA (11). If this is the case, then the total amount of circulating DNA in the blood, in the absence of DNA degradation, would be the number of DNA fragments released by dying cells. However, the amount of DNA fragments released by dying cells is not specified in the literature. In the absence of DNA degradation, the total amount of circulating DNA in the blood, in the absence of DNA degradation, would be the number of DNA fragments released by dying cells. However, the amount of DNA fragments released by dying cells is not specified in the literature.

Radiation, chemotherapy and other cancer treatments cause cell death by apoptosis (12) and the amount of circulating DNA is seen in cancer patients under treatment before treatment, possibly due to the inhibition of apoptosis in the cancer cells. Furthermore, in the early stages of cancer where seemingly little cell death occurs, circulating DNA may already be present in normal patients as well. As the cancer burden increases, so does cell death and an increase in the amount of circulating DNA is expected. In the absence of DNA degradation, the total amount of circulating DNA in the blood, in the absence of DNA degradation, would be the number of DNA fragments released by dying cells. However, the amount of DNA fragments released by dying cells is not specified in the literature.

Figure 1: Representation of release and clearance of free circulating DNA.

Conclusion

The increased amount of circulating DNA in the blood of patients may be the disease of the release of DNA by dying cells and the clearance of DNA is disturbed. This low concentration of circulating DNA in the blood of normal individuals may thus be due to the lower rate of DNA release by cells or a rapid removal of DNA by the nucleic acid degradation pathway. In the absence of DNA degradation, the total amount of circulating DNA in the blood, in the absence of DNA degradation, would be the number of DNA fragments released by dying cells. However, the amount of DNA fragments released by dying cells is not specified in the literature.

Circulating DNA can be found in a variety of conditions and even though these conditions are unrelated, the presence of circulating nucleic acids is a common feature and thus more research is required in order to find a common mechanism of origin. Even though research has been made regarding the origin of circulating DNA for more than 30 years, the mechanism of release still eludes us. The possibility that more than one mechanism may be involved in the release of circulatory DNA is not excluded in the present study (13).

References

A METHOD FOR CHARACTERIZATION OF CIRCULATING DNA

Maniesh van der Vaart and Pieter J. Pretorius

School of Biochemistry, North-West University, Private Bag X6001, Potchefstroom, 2520, SOUTH AFRICA.

ABSTRACT

Background

Although much work has been done in the field of circulating DNA, no definitive information on the nature of circulating DNA is available. Characterization of circulating DNA by sequence analysis may give valuable information about the origin and function of these nucleic acids and molecules.

Methods

Circulating DNA was isolated from plasma of healthy individuals as well as cancer patients with various methods and was cloned into a blunted cloning vector. Resulting colonies were sequenced and analyzed.

Results

It was found that the majority of the DNA isolated into plasmid vectors was about 200 bp in length. Sequence analysis revealed that the cloned circulating DNA was made up of currently uncharacterized human genomic sequences or when human repeats were applied. To match sequences present in the collection, known genes isolated at various distances from the identical target. In addition, in the presence of large repeats of a variety of Alu repeat sequences were observed. Preliminary results showed that more Alu repeats are present in the plasma of healthy individuals than in adherent material.

Methods for cloning of circulating DNA were each analyzed by using Basic Local Alignment Search Tools (BLAST, CENSOR), and the results were analyzed using the Multiple Em by Local Alignment Search Tools (MEME/MAST) [1], etc.

RESULTS

The cloned sequences were analyzed and it was found that it was not easily characterized. BLAST gave hits of more significant (E-value ≤ 1×10^-10) matches for most of the sequences, but these matches were not current genes but rather genomic regions that contained features such as primer ends, etc. These sequences could be characterized by the presence of telomere or features in parts of the subject sequence. Furthermore, the number of different Alu repeats in the sequences was determined, some of which contained target sequences for the analysis of potential targets for the sequences. The significance of the presence of these motives is unknown at present.

CONCLUSION

Circulating DNA in the plasma of healthy individuals was found to vary in length from 200 bp to several kilobases. The majority of the DNA isolated into plasmid vectors was about 200 bp in length. Sequence analysis revealed that the cloned circulating DNA was made up of currently uncharacterized human genomic sequences or when human repeats were applied to match sequences present in the collection, known genes isolated at various distances from the identical target. In addition, in the presence of large repeats of a variety of Alu repeat sequences were observed. Preliminary results showed that more Alu repeats are present in the plasma of healthy individuals than in adherent material.

REFERENCES


Figure 1. Flow diagram of experimental workflow.

Figure 2. Products of PCR amplifying DNA fragments with the primers shown.

Figure 3. Products of colony PCR of DNA cloned in pCDNA3 and pCDNA3.1 vector. The vector plasmid DNA (open circle) and the plasmid DNA (filled circle) were sequenced.

Figure 4. Validation of the primers used in the experiment. The number of fragments detected in the sequencing reaction is shown as a histogram. The data was obtained by colony PCR and sequencing.

No significant difference between the treatment of sequences on chromosomes of patients and controls was evident.
Section 5: Permission from editors to use published papers

AUTHORS' ASSURANCES AND ASSIGNMENT OF COPYRIGHT

Clinical Chemistry, International Journal of Molecular Diagnostics and Laboratory Medicine

Date: 19 February 2007

Ms. No. CLINICH-A/2007/007130

Title: THE ORIGIN AND CLEARANCE OF CIRCULATING FREE DNA

The following Authors' Assurances and Assignment of Copyright ("License") must be signed and returned to the Clinical Chemistry Editorial Office, 1224 Jefferson Park Ave. Suite 600, Charlottesville, VA 22903, Fax (434) 979-3599, before a manuscript can be accepted for publication. If the contribution to Clinical Chemistry, described above, is a joint work, the American Association for Clinical Chemistry (the "Publisher") requires that each joint Author sign this License. If the copyright in the contribution is owned by the Author's employer, the employer or an authorized representative must sign this form. In the event that the Publisher decides not to publish the Work in Clinical Chemistry, this License shall be null and void.

1. In consideration of publication in Clinical Chemistry of the above-titled work (the "Work") the undersigned (the "Author") transfers, to the extent that there is copyright to be considered, the exclusive copyright in the above-referenced work to the Publisher, including the irrevocable right to publish, reproduce, distribute, transmit, display, store, translate, create derivative works from and otherwise use the Work in any form, manner, format, or medium, whether now known or hereafter developed, throughout the world and in any language, for the entire duration of any such right and any renewal or extension thereof, and to permit sublicensees to do any or all of the foregoing as well. This Assignment shall be governed by and construed under the laws of the District of Columbia.

2. Author retains the non-exclusive right to use the Work in the following ways without further permission but only after publication of the Work in Clinical Chemistry and subject to the requirement that credit be given to its first publication in the appropriate issue of Clinical Chemistry: 1) Reproduce the Work in print collections of Author's own writings; 2) Reproduce the Work in print format for inclusion in a thesis or dissertation that the Author writes; 3) Present the Work orally; 4) Reproduce the Work for use in courses the Author is teaching (If the Author is employed by an academic institution, that institution may also reproduce the Work for Course Teaching); 5) Distribute photocopies of the Work to colleagues for non-commercial purposes only (providing that recipients are informed that they may not further distribute or copy the Work); 6) Route figures and tables created by the Author in future works the Author writes; 7) If the Work is prepared as a work made for hire, the Author's employer may make photocopies for internal use only; 8) Author may use or authorize use of Supplemental Online Material associated with the Work for any purpose and in any format.

3. In addition to the rights retained by the Author in 2 above, for Works that are both (A) created under a grant from NIH and (B) accepted for publication in Clinical Chemistry on or after May 1, 2005: the Author may cause the posting, no sooner than 12 months after final publication of the Work in Clinical Chemistry, of the "Accepted Version" of the Work in PubMed Central (PMC). The "Accepted Version" shall be defined as the version of the paper accepted for publication in Clinical Chemistry including changes resulting from peer review but prior to Clinical Chemistry's copy editing and production.

4. In order that we may preserve the commercial value of our publication, the Author agrees to obtain authorization from AACC in advance for any uses not expressly described in paragraphs 2 or 3.

5. Each Author warrants that his/her manuscript is an original work not published wholly or partly elsewhere, except in the form of an abstract of 400 words or fewer, and except as noted immediately below; that he/she has full power to make this assignment; that he/she has disclosed to the Publisher all contributors to the Manuscript and the underlying work; and that the Manuscript contains nothing untoward, libelous, invading the right of privacy, or infringing a proprietary right. If portions have been published previously, permission has been obtained for publication in Clinical Chemistry, and Author has submitted a copy of the permission release and copy for credit lines with this Manuscript. The Publisher can edit, adapt, or change this Manuscript, at his/her sole discretion.

6. Each Author acknowledges that he/she has participated in the work in a substantive way and is prepared to take public responsibility for its content. Each listed Author must have (A) participated in conception, design, analysis, or interpretation; (B) drafted or critically reviewed the Manuscript and (C) read and approved the final submitted manuscript and its revision.

Please sign and return this form to the Editorial Office. For U.S. Government employees, your signature represents that you wrote this manuscript as part of your official duties and cannot assign the copyright.

<table>
<thead>
<tr>
<th>PRINT NAME</th>
<th>ALL AUTHORS MUST SIGN ON BOTH SPACES BELOW</th>
<th>SIGNATURE</th>
<th>U.S. GOV'T EMPLOYED?</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. M. VAN DER VAART</td>
<td></td>
<td></td>
<td>NO</td>
<td>19 FEB 2007</td>
</tr>
<tr>
<td>2. J. PRETTO RUILES</td>
<td></td>
<td></td>
<td>YES</td>
<td>19 FEB 2007</td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please photocopy this form as necessary for additional signatures.

CONFLICT OF INTEREST DISCLOSURE

Each Author is responsible for disclosing to the Publisher all possible conflicts of interest regarding this Manuscript (including sources of support, monetary or other, monetary interests in products studied, consultancies, etc.) and whether the Author regards them to be actual conflicts of interest. Please sign below to indicate that the necessary disclosure has been completed.

<table>
<thead>
<tr>
<th>PRINT NAME</th>
<th>SIGNATURE</th>
<th>DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. M. VAN DER VAART</td>
<td></td>
<td>19 FEB 2007</td>
</tr>
<tr>
<td>2. J. PRETTO RUILES</td>
<td></td>
<td>19 FEB 2007</td>
</tr>
<tr>
<td>3.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Author Services

Terms & Conditions

Annals of the New York Academy of Sciences

e-Offprint Terms & Conditions

Please note the following Terms and Conditions regarding the use of this e-Offprint file. Use of this e-Offprint file indicates consent to these Terms and Conditions.

Provided that you give appropriate acknowledgement to the Journal, Association/Society and Blackwell Publishing, and give full bibliographic reference for the Article, and as long as you do not sell or reproduce the Article or any part of it for commercial purposes (i.e., for monetary gain on your own account or on that of a third party, or for indirect financial gain by a commercial entity) you may use the PDF, in the following ways:

- you may share print or electronic copies of the Article with colleagues;
- you may use all or part of the Article and abstract, without revision or modification, in personal compilations or other publications of your own work;
- you may use the Article within your employer's institution or company for educational or research purposes, including use in course packs.

Please note that you are not permitted to post the Blackwell Publishing PDF version of the Article online unless you have paid the OnlineOpen fee.

For additional information regarding any use of this PDF file not covered by these Terms and Conditions, please refer to the Blackwell Rights Department at: journalsrights@oxon.blackwellpublishing.com.

I accept the Terms and Conditions ☐
I do not accept the Terms and Conditions ☐
What rights do I retain as a journal author?*

As a journal author, you retain rights for a large number of author uses, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier. These include:

- the right to make copies (print or electric) of the journal article for their own personal use, including for their own classroom teaching use;
- the right to make copies and distribute copies (including via e-mail) of the journal article to research colleagues, for personal use by such colleagues (but not for Commercial Purposes**), as listed below);
- the right to post a pre-print version of the journal article on Internet web sites including electronic pre-print servers, and to retain indefinitely such version on such servers or sites (see also our Information on Electronic Preprints for a more detailed discussion on these points);
- the right to post a revised personal version of the text of the final journal article (to reflect changes made in the peer review process) on the author's personal or institutional web site or server, incorporating the complete citation and with a link to the Digital Object Identifier (DOI) of the article;
- the right to present the journal article at a meeting or conference and to distribute copies of such paper or article to the delegates attending the meeting;
- for the author's employer, if the journal article is a 'work for hire', made within the scope of the author's employment, the right to use all or part of the information in (any version of) the journal article for other intra-company use (e.g., training), including by posting the article on secure, internal corporate Intranets;
- patent and trademark rights and rights to any process or procedure described in the journal article;
- the right to include the journal article, in full or in part, in a thesis or dissertation;
- the right to use the journal article or any part thereof in a printed compilation of works of the author, such as collected writings or lecture notes (subsequent to publication of the article in the Journal); and
- the right to prepare other derivative works, to extend the journal article into book-length form, or to otherwise re-use portions or excerpts in other works, with full acknowledgement of its original publication in the Journal.

*Please Note: The rights listed above apply to journal authors only. For information regarding book author rights, please contact the Global Rights Department.

Elsevier Global Rights Department
phone (+44) 1865 843 930
fax (+44) 1865 853 333
email: permissions@elsevier.com

Other uses by authors should be authorized by Elsevier through the Global Rights Department, and journal authors are encouraged to let Elsevier know of any particular needs or requirements.

**Commercial Purposes includes the use or posting of articles for commercial gain including the posting by companies or their employee-authored works for use by customers of such companies (e.g., pharmaceutical companies and physician-prescribers); commercial exploitation such as directly associating advertising with such postings; the charging of fees for document delivery or access; or the systematic distribution to others via e-mail lists or file servers (to parties other than known colleagues), whether for a fee or for free.

http://www.elsevier.com/wps/find/authorsview.authors/copyright

23/05/2009
<table>
<thead>
<tr>
<th>Number of patients, controls and other</th>
<th>Tubes and centrifugation</th>
<th>Plasma/serum and starting volume for isolation</th>
<th>Time before plasma collection and storage temp.</th>
<th>Isolation method</th>
<th>Alterations before and during kit protocol</th>
<th>Quantification method</th>
<th>Quantity of circulating DNA Average (range) ng/ml</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other: 10 ml EDTA or plain, centrifuged at 3000g and supernatant re-centrifuged at 3000g, times not available</td>
<td>10 ml EDTA or plain, centrifuged at 3000g and supernatant re-centrifuged at 3000g, times not available</td>
<td>QIAamp blood kit (Qiagen) according to blood and body fluid protocol, elution volume 50 µl</td>
<td>-20°C</td>
<td>Either 7.5 and 12 ml of plasma were heated at 99°C for 30 min, then centrifuged at 14000 rpm for 30 min, proteinase K and buffer Al was added to the supernatant (1/10) and incubated overnight at 55°C</td>
<td>Real-time PCR for B-globin with Taqman probe</td>
<td>Plasma: 6.5 (2.4-12.9) **</td>
<td>22.9 (2.4-215.6)</td>
<td>(21)</td>
</tr>
<tr>
<td>Patient: 35 with breast cancer</td>
<td>plasma</td>
<td>QIAamp Blood kit (Qiagen)</td>
<td></td>
<td></td>
<td></td>
<td>Healthy: (0-45)</td>
<td>Patients: 110 (20-178)</td>
<td>(23)</td>
</tr>
<tr>
<td>Control: 17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other: 64 with acute blunt traumatic injury</td>
<td>64 with acute blunt traumatic injury</td>
<td>QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol</td>
<td>-30 or -20°C</td>
<td>Real-time PCR for 3'non-coding with Taqman probe</td>
<td></td>
<td>Healthy: 20.6</td>
<td>Minor/moderate trauma: 91.2</td>
<td>(24)</td>
</tr>
<tr>
<td>Control: 23</td>
<td>10 ml EDTA, divided in 2 parts: 1. Initially centrifuged 10 min at 16000g followed by enrichment a: nothing b: filtration (0.2 µm filter) c) 10 min at 16000g d) 10 min at 16000g and filtered 2. Diluted and centrifuged at 12000 g on a discontinuous percol gradient followed by either a, b or c 400-800 µl plasma</td>
<td>QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol</td>
<td></td>
<td></td>
<td></td>
<td>1a: 16.5 (4.7-7.1) ** 1b: 8.7 (2.5-8.6) ** 1c: 6.7 (2.3-18.6) ** 2a: 117.8 (25-333) ** 2b: 8.7 (2.3-18.6) ** 2c: 17.4 (4-63) **</td>
<td>(25)</td>
<td></td>
</tr>
<tr>
<td>Other: 34 pregnant woman (gestational age 13-35 weeks)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient: 39 with cancer</td>
<td>EDTA, centrifugation for 20 min at 3000g</td>
<td>QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol</td>
<td></td>
<td>Competitive PCR using lamin B2 locus as single copy gene and ddPCR using sequential dilutions of elution: bromide-stained gel after PAGE and Real-time PCR for B-globin with LCR-Red 600</td>
<td></td>
<td>Healthy: 3.7 (2-15) Cancer: 216.0 (101-1200) A difference of ~15% between the two methods was observed</td>
<td>(26)</td>
<td></td>
</tr>
<tr>
<td>Control: 20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Supplementary table for "Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?" (Chapter 3)
<table>
<thead>
<tr>
<th>Patients: 56 NSCLC patients Controls: 16</th>
<th>7.5 ml EDTA two times centrifugation for 10 min at 2500 rpm</th>
<th>1000 ul plasma</th>
<th>Immediately</th>
<th>QIAamp DNA mini kit (QIagen)</th>
<th>DNA DipStick kit (Invitrogen)</th>
<th>Healthy: No data Patients: (5-500) (27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other: 22 bone marrow transplantation recipients Controls: 14</td>
<td>EDTA or plain; Centrifuged first at 3000 g for 10 min followed by 10 min at 16000 g</td>
<td>400-600 ul Plasma and serum (from only 7 patients)</td>
<td>Within 6 hours</td>
<td>QIAamp blood kit (QIagen) according to blood and body fluid protocol</td>
<td>Real-time PCR for B-globin with Tagman probe,</td>
<td>Healthy: 8.3% Patients: Plasma: 7.7%, Serum: 107,9% (13)</td>
</tr>
<tr>
<td>Patients: 186 with breast cancer 161 with ovarian cancer 74 with prostate cancer Controls: 100</td>
<td>1 ml of serum</td>
<td>-7C° for ~1 year</td>
<td>QIAamp 96 Spin blood DNA extraction kit (QIagen) according to the blood and body fluid protocol, eluted in 200 ul sterile water or Tris-EDTA</td>
<td>Proteins were digested with proteinase K before DNA extraction and three consecutive washes was performed during isolation</td>
<td>PicoGreen Double strand DNA detection kit (Molecular probe) and fluorescence measured by spectro-fluorometer</td>
<td>Healthy: 57.0±30 Breast cancer; CA15-3&lt;100U/ml:74±49** CA15-3&lt;500: 111±104** CA15-3&lt;500: 307±601 Ovarian cancer: CA125&lt;100U/ml: 67±30** CA125&lt;500: 49±27** CA125&lt;500: 118±125 Prostate cancer: PSA &lt;4ng/ml: 458±790** PSA &gt;20: 490±471 (28)</td>
</tr>
<tr>
<td>Patients: 34 with lung cancer Controls: 20</td>
<td>5 ml EDTA Two times centrifugation at 800 g for 10 min</td>
<td>2000 ul plasma</td>
<td>Immediately, plasma was stored at -80°C</td>
<td>QIAamp DNA mini kit (QIagen)</td>
<td>Fluorometry: PicoGreen dsDNA quantification reagent (Molecular Probes)</td>
<td>Healthy: 78 (23-128) Patients: 157(14-1054) (29)</td>
</tr>
<tr>
<td>Patients: 37 with lymphoma Controls: 20</td>
<td>EDTA</td>
<td>500 ul plasma</td>
<td>NucleoSpin Blood DNA purification kit (Machery-Nagel)</td>
<td></td>
<td>Video gel image analysis (SuI DNA were spotted on a 1% agarose gel containing 5ug/ml EBBr and allowed to dry. UV intensities were analyzed and concentrations calculated</td>
<td>Healthy: 44* (all lower than 50) Patients: 256* (65-2265) (30)</td>
</tr>
<tr>
<td>Patients: 100 with NSCLC Controls: 100 matched by sex, age and smoking habits</td>
<td>7.5 ml EDTA Two times centrifugation for 10 min at 2500 rpm</td>
<td>1 ml plasma</td>
<td>Blood stored at -140°C before plasma collection</td>
<td>QIAamp DNA mini blood kit (QIagen) according to blood and body fluid protocol, eluted in 50 ml water</td>
<td>Real time PCR for hTERT with Taqman probe</td>
<td>Healthy: 46 (0.1-27) Median: 3.1* Patients: 75.0 (0.5-3010) Median: 24.3* (20)</td>
</tr>
</tbody>
</table>

Supplementary table for "Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?" (Chapter 3)
<table>
<thead>
<tr>
<th>Number of patients, controls and other*</th>
<th>Tubes and centrifugation</th>
<th>Plasma/serum and starting volume for isolation</th>
<th>Time before plasma collection and storage temp.</th>
<th>Isolation method</th>
<th>Alterations before and during kit protocol</th>
<th>Quantification method</th>
<th>Quantity of circulating DNA Average (range) ng/ml</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other: 88 with a stroke-like syndrome Controls: 24</td>
<td>10 ml EDTA, centrifuged once for 5 min at 1500g</td>
<td>200 μl plasma</td>
<td>-80°C</td>
<td>QIAamp blood kit (Qiagen) according to the blood and body fluid protocol, eluted in 200 μl sterile water</td>
<td>Real-time PCR for B-globin with Taqman probe</td>
<td>Healthy; no data available Fallouts: 9.5* (1.6 – 54.6)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Other: 96 myocardial Infarction (M) patients Controls: 274 | 200 μl serum | | | QIAamp 96 Spin blood DNA extraction kit (Qiagen) according to the blood and body fluid protocol, eluted in 200 μl sterile water | PicoGreen DNA kit (Molecular probe) and fluorescence measured by spectro-fluorometer | Healthy | 36.3 ± 23.6
Ml: 510.6 ± 398 |
| Other: 31 with heart, liver or renal transplants Controls: 10 | EDTA, two step centrifugation, first at 1600g for 10 min, second at 16000g for 10 min | 400 μl plasma | | QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol, elution volume 50 μl | Real-time PCR for B-globin with Taqman probe | Healthy: 6.0* Patients: 8.6* |
| Patients: 84 with NSCLC (before surgery) Controls: 43 | Lithium-heparin, two times centrifugation for 10 min at 2500 rpm | 1000 μl plasma | Immediately, 80°C | QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, elution volume 50 μl | DipStick TM Kit (Invitrogen) | Healthy: 18.0
Patients: 318.9* Measurable amounts of DNA could not be found in 3 of the patients and 11 of the controls |
| Patients: 19 with lung cancer 4 with colon cancer (patients was under treatment) Controls: 20 | 5 ml EDTA tubes, centrifuged 10 min at 800g, supernatant centrifuged 10 min at 1500g, and the resulting plasma was incubated with proteolytic buffer and concentrated by further centrifugation in Amicon Ultra-15 filtration devices (Millipore). | Up to 300 μl concentrated plasma | Unknown, -20°C | KingFisher allocate magnetic beads and a KingFisher ML robotic magnetic particle processor (ThermoLifeSciences) And QIAamp DNA Midi kit (Qiagen) according to blood and body fluid protocol | Fluorometry (Fluoroscan with PicoGreen reagent (Molecular Probes)) | Healthy: (3-22)
Cancer: (13-127)
No significant difference in yield was observed if DNA was isolated from concentrated plasma with either the KingFisher method or the Qiagen kit |
| Patients: 41 with NPC Controls: 14 EBV+ (healthy) 29 EBV- (healthy) | Plasma stored at -70°C until use | 400 μl plasma | | QIAamp blood and tissue kit (Qiagen) | Real-time PCR for B-globin | Healthy: EBV+:18.43**
EBV-:15.67**
All:16.57
Fallouts: 28.79 |

Supplementary table for “Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?” (Chapter 3)
<table>
<thead>
<tr>
<th>Number of patients, controls and other</th>
<th>Tubes and centrifugation</th>
<th>Plasma/serum and starting volume for isolation</th>
<th>Time before plasma collection and storage temp.</th>
<th>Isolation method</th>
<th>Alterations before and during kit protocol</th>
<th>Quantification method</th>
<th>Quantity of circulating DNA Average (range) ng/ml</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 with prostate cancer</td>
<td>5 ml lithium heparin. Two times centrifuged at 1000g for 15min</td>
<td>2 ml plasma</td>
<td>within 2 h of collection, plasma stored at -70°C until isolation</td>
<td>QIAamp DNA blood midi kit</td>
<td>Washes with buffer AW2 was done 3 times to remove PCR inhibitors, DNA was eluted in Tris-HCl</td>
<td>Real-time PCR for Glycolaldehyde-3-phosphate dehydrogenase gene with Taqman probe</td>
<td>Healthy: Plasma: 1.8 Serum: 12.6 Patients: Plasma: 3.7 (0.1-93.5) Serum: 35.6 (0.3-641.3)</td>
<td>(10)</td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>10 ml EDTA, two consecutive centrifugations of 1500g for 10 min 10 ml plain tubes for serum</td>
<td>1 ml Plasma and 1 ml serum</td>
<td>-70°C</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, Elution volume 70 µl</td>
<td>The difference in DNA concentration of stored samples was tested. Isolated DNA was analyzed and stored and reanalyzed and plasma was used for immediate DNA isolation and then stored for later isolation. (stored 9 months: smokers or 41 months: healthy and patients) DNA eluted in distilled water</td>
<td>Real-time PCR for hTERT with Taqman probe</td>
<td>Isolated DNA: Healthy: 2.4* (0.3-18.6) Patients: 23.4* (1.3-823) Smokers: 10.3* (8.1-117.9)</td>
<td>(6)</td>
</tr>
<tr>
<td><strong>Other:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34 with benign prostate hyperplasia (BPH)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Lithium-heparin, two times centrifugation for 10 min at 2500rpm</td>
<td>1000 µl plasma</td>
<td>-60°C</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, Elution volume 50 µl</td>
<td></td>
<td>Real-time PCR for hTERT with Taqman probe</td>
<td>Healthy: 1.8* (0.1-4) Patients: 5.8* (1.3-24.8) Smokers: 6.1* (0.5-94.9) Stressed isolated DNA: Healthy: 0.8* (0-4.3) Patients: 6.4* (0.8-150.9) Smokers: 7.9* (1.6-57.2)</td>
<td>(6)</td>
</tr>
<tr>
<td><strong>Other:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>2 ml plasma</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Healthy:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Men:</td>
<td>21.9 ±9.7 (9.4-43.6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(18)</td>
</tr>
<tr>
<td>Women:</td>
<td>20.5 ±7.4 (6.7-37.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Pentol:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No data: concentration did not differ significantly between healthy and cancer, but it did for BPH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(17)</td>
</tr>
</tbody>
</table>

**Supplementary table for “Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?”**  
Chapter 3  
P4
<table>
<thead>
<tr>
<th>Number of patients, controls and other*</th>
<th>Tubes and centrifugation</th>
<th>Plasma/serum and starting volume for isolation</th>
<th>Time before plasma collection and storage temp.</th>
<th>Isolation method</th>
<th>Alterations before and during kit protocol</th>
<th>Quantification method</th>
<th>Quantity of circulating DNA Average (range) ng/ml</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Controls: 40 men</td>
<td>Plasma: 7 ml EDTA Either single or double centrifugation at 1600g for 10 min Serum: clotted at room temp for 30 min—single centrifugation</td>
<td>600 ul Plasma or serum</td>
<td>Processed within 1 h of collection and plasma was stored at -70°C until isolation</td>
<td>QIAamp DNA Blood Mini kit</td>
<td>Real-time PCR</td>
<td>One spin: 2.9* (0.0-6.9)** Two spin: 2.8* (0.0-1.8) Serum: 27.1* (2.6-242.7)</td>
<td>(19)</td>
<td></td>
</tr>
<tr>
<td>Patients: 9 with lung cancer</td>
<td>EDTA, Initially centrifuged 10 min at 350g followed by 15 min at 2000g</td>
<td>1 ml plasma</td>
<td>-80°C</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol And Salting-out method</td>
<td>Real-time PCR for ERV-3 with Taqman probe</td>
<td>Lung cancer: Qiagen: 4.0 Salting-out: 7.9** Benign: Qiagen: 3.0 Salting-out: 15.8**</td>
<td>(36)</td>
<td></td>
</tr>
<tr>
<td>Other: 7 with benign lung disease</td>
<td>10 ml EDTA, Banked samples: centrifuged once for 5 min at 1600g, Fresh samples: Centrifuged twice for 10 min at 1600g</td>
<td>200 ul plasma</td>
<td>1 hour, -80°C</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol</td>
<td>Added 400 ul water and 400 ul buffer ALF to sample, to increase DNA yield</td>
<td>Real-time PCR for β-globin with Taqman probe</td>
<td></td>
<td>(9)</td>
</tr>
<tr>
<td>Patients: 68 with breast cancer</td>
<td>Two times at 2000rpm for 10 min Plasma</td>
<td></td>
<td></td>
<td>QIAamp DNA mini blood kit (Qiagen)</td>
<td></td>
<td>Real-time PCR for β-globin</td>
<td>Healthy: 4.7 ±4 (men) 11.4 ± 9.6 (women) Breast cancer: 45.8 ± 6.5 New prostate cancer: 20.2 ± 18.7 Treated prostate cancer: 9.7 ± 14.4</td>
<td>(37)</td>
</tr>
<tr>
<td>12 with prostate cancer</td>
<td>Stored at 4°C, processed within 4 h of collection, Plasma stored at -20°C until DNA isolation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15 with prostate cancer under therapy</td>
<td>Controls: 13 (probably men) 54 (probably women) Patients: 9 with colon cancer 9 with stomach cancer Controls: 10 men 10 women</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of patients, controls and other</td>
<td>Tubes and centrifugation</td>
<td>Plasma/serum and starting volume for isolation</td>
<td>Time before plasma collection and storage temp.</td>
<td>Isolation method</td>
<td>Alterations before and during kit protocol</td>
<td>Quantification method</td>
<td>Quantity of circulating DNA Average (range) [ng/ml]</td>
<td>Reference</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>--------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------</td>
<td>-----------------</td>
<td>-----------------------------------------</td>
<td>---------------------</td>
<td>-----------------------------------------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Patients:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32 with colorectal cancer (CRC)</td>
<td>1 ml CORVAC serum separator tubes, centrifugation for 15 min at 10000g and filtered (3 mm serum filter; Fisher/Scientific)</td>
<td>Serum</td>
<td>Within 8 hours, -80°C</td>
<td>No isolation, only proteinase K digestion and further centrifugation for 5 min at 10000g</td>
<td>Real-time PCR for ALU with SYBR Gold (Molecular Probe)</td>
<td></td>
<td>Healthy: $540 \pm 250$</td>
<td></td>
</tr>
<tr>
<td>19 with periampullary cancer (PACs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stage III: $1630 \pm 330$</td>
<td>(30)</td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stage III/IV: $1730 \pm 450$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PACs: $840 \pm 350$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Stage III/IV: $680 \pm 620$</td>
<td></td>
</tr>
<tr>
<td><strong>Other:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>62. In a general intensive care unit (ICU)</td>
<td>EDTA, initially centrifuged 6 min at 3000 rpm followed by 10 min at 14000 rpm</td>
<td>200 µl plasma</td>
<td>Within 3 hours, -20°C</td>
<td>High Pure PCR template preparation kit (Roche)</td>
<td>Real-time PCR for B-globin with Taqman probe</td>
<td></td>
<td>Healthy: $17.0$ (14-19)</td>
<td>(39)</td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Patients: $90.0$ (46-260)</td>
<td></td>
</tr>
<tr>
<td><strong>Other:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17 recreationally trained men (joining four 3-week training periods with 5 day rest in-between, T1- T4 of which T1 and T4 included low-volume training, T2 high-volume training and T3 very high-volume training)</td>
<td>Type of tubes not available, centrifugation for 10 min at 8000g followed by 10 min at 16000g</td>
<td>400 µl plasma</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 50 µl sterile water</td>
<td>Real-time PCR for B-globin with SYBR Green I</td>
<td>Baseline: $0.2$ (0.1-0.4)</td>
<td></td>
<td>T1: $0.8$ (0.5-2.5)</td>
<td>(40)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T2: $1.8$ (0.2-2.5)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T3: $4.0$ (1.7-6.3)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>T4: $0.5$ (0.2-0.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Other:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 Hemodialysis (HD) patients</td>
<td>EDTA, centrifugation time and speed not available</td>
<td>500 µl plasma</td>
<td>-20°C</td>
<td>DNA isolation reagent set (Roche) according to protocol, elution volume 50 µl</td>
<td>Real-time PCR with Vista Green</td>
<td></td>
<td>Healthy: $12.6 \pm 4.7$</td>
<td>(41)</td>
</tr>
<tr>
<td>Controls:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Before HD: $13.9 \pm 5.2$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>After 20 min of HD: $20.3 \pm 10.3^*$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>End of HD: $129.4 \pm 83.4$</td>
<td></td>
</tr>
<tr>
<td>Controls:</td>
<td>Plain tubes, centrifuged 10 min at 3000g</td>
<td>200 µl Serum</td>
<td>12 samples: 2 hours</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 100 µl sterile water</td>
<td>DNA was either extracted (CZE) or digested with proteinase K (CZE) for capillary zone electrophoresis</td>
<td>Capillary zone electrophoresis (CZE) with laser-induced fluorescence detection and Real-time PCR for B-globin with Taqman probe</td>
<td>2h PCR: $87.7$ (23-122)</td>
<td>(42)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>36 samples: 8 hours, -20°C</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 100 µl sterile water</td>
<td>DNA was either extracted (CZE) or digested with proteinase K (CZE) for capillary zone electrophoresis</td>
<td>Capillary zone electrophoresis (CZE) with laser-induced fluorescence detection and Real-time PCR for B-globin with Taqman probe</td>
<td>2h CZE: $70.8$ (27-146)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2h CZE: $70.8$ (27-146)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5h CZE: $103.8$ (24-362)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5h CZE: $103.8$ (24-362)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6h CZE: $192.8$ (48-317)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6h CZE: $192.8$ (48-317)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8h CZE: $195.7$ (50-355)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The free methods compared very well although CZE give slightly higher results</td>
<td></td>
</tr>
</tbody>
</table>

Supplementary table for "Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?" (Chapter 3) p6
<table>
<thead>
<tr>
<th>Number of patients, controls and other*</th>
<th>Tubes and centrifugation</th>
<th>Plasma/serum and starting volume for isolation</th>
<th>Time before plasma collection and storage temp.</th>
<th>Isolation method</th>
<th>Alterations before and during kit protocol</th>
<th>Quantification method</th>
<th>Quantity of circulating DNA Average (range) ng/ml</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other: 58 with chest pain suggestive of cardiac ischemia Controls: 21</td>
<td>4 ml EDTA, divided in 2 portions and centrifuged once for 10 min at 1600g, plasma from 1 portion was filtered (0.22 µm pore) and used for quantification</td>
<td>400-800 µl plasma</td>
<td>-80°C</td>
<td>QIAamp DNA blood kit (Qiagen) according to manual and protocol</td>
<td>Real-time PCR for Beta-actin with Taqman probe</td>
<td>Healthy: 2.3³, Minor cardiac injury: 3.1³, ST elevation angiography: 3.5⁴, ST elevation myocardial infarction: 4.4⁴, Died: 7.7⁴</td>
<td>(43)</td>
<td></td>
</tr>
<tr>
<td>Patients: 61 breast cancer Other: 38 non-neoplastic breast diseases Controls: 27 woman</td>
<td>5 ml EDTA tubes, centrifuged 10 min at 2000g, supernatant centrifuged 10 min at 12000g</td>
<td>200 µl Plasma</td>
<td>Within 2 hours, -80°C</td>
<td>QIAamp DNA mini blood kit (Qiagen) according to manual and protocol</td>
<td>Elution volume 50 µl</td>
<td>Real-time PCR for HBV with Taqman probe</td>
<td>Healthy: 12.9 (3-73), Benign: 22.3 (6-106), Malignant: 64.6 (9-585)</td>
<td>(44)</td>
</tr>
<tr>
<td>Patients: 10 with small cell lung cancer Controls: 10</td>
<td>Plasma: EDTA, two times centrifugation at 2000g for 10min. Serum tubes were left for 30-60 or 24h at room temp and centrifuged at 2000g for 10min</td>
<td>Plasma and serum was stored at -80°C</td>
<td>Test four kits, QIAamp Virus Spin Kit (Qiagen) give highest yield</td>
<td>Real-time PCR of a 77bp fragment of the AAT gene</td>
<td>Healthy: Plasma:5.07 (2.5-7.5), Serum: 24.65 (5-64)</td>
<td></td>
<td>(5)</td>
<td></td>
</tr>
<tr>
<td>Patients: 70 with colorectal cancer Controls: 20</td>
<td>Lithium-heparin</td>
<td>1000 µl plasma</td>
<td>Within 3h of collection</td>
<td>QIAamp Tissue extraction kit (Qiagen)</td>
<td>Five passages of purification on the same column, final elution in 50uL water</td>
<td>DNA DipStick kit (Invitrogen)</td>
<td>Healthy: 5° (5-50), Patients: 437° (125-1750)</td>
<td>(14)</td>
</tr>
<tr>
<td>Other: 30 with pancreatitis of which 9 had severe acute pancreatitis Controls: 18</td>
<td>Serum</td>
<td>160 µl serum</td>
<td>Immediately, sera stored at -80°C</td>
<td>DNasey Blood and Tissue Kit (Qiagen)</td>
<td>Real-time PCR with Quantifier® Human DNA quantification kit (Applied Biosystems)</td>
<td>Healthy: 70° (10-231), MILD: 59° (7-170), Severe: 270° (128-672)</td>
<td>(21)</td>
<td></td>
</tr>
</tbody>
</table>

* Individuals was in the "other" group when they could not be classified as healthy but they also did not have cancer.

** The median values.

*** Indicate which values were not included in figure 1

†Value in ng/ml calculated from genome equivalents

‡Value estimated from box plots

Supplementary table for "Is the role of Circulating DNA as a biomarker of cancer being prematurely overrated?" (Chapter 3)
Section 7: Permission from co-authors

RUSSIAN ACADEMY OF SCIENCES
SIBERIAN BRANCH
Institute of Chemical Biology and
Fundamental Medicine
(ICBFM SB RAS)

Lavrentyev Ave. 8
Novosibirsk, 630090, Russia
tel.: +7 383 330-72-41, fax: +7 383 333-36-77
e-mail: niboch@niboch.nsc.ru
www.niboch.nsc.ru

To whom it may concern

Hereby we, Dr. Dmitry V. Semenov PhD, Dr. Elena V. Kuligina PhD and Dr. Vladimir A. Richter PhD, as co-authors of the paper titled "Characterisation of circulating DNA by parallel tagged sequencing on the 454 platform" give permission from that the manuscript can be submitted for degree purposes by Maniesh van der Vaart.

The contribution of the co-authors was as follows: Dr. Dmitry V. Semenov did extensive sequence analysis including application of personally developed algorithms, interpretation of the results and some graphical presentation; Dr. Elena V. Kuligina did database assembling and general analysis of the results, Dr. Vladimir A. Richter provided scientific supervision of Kuligina and Semenov and financial support of their work.

Dr. Dmitry V. Semenov, PhD Senior Researcher, ICBFM SB RAS

Dr. Elena V. Kuligina, PhD Researcher, ICBFM SB RAS

Dr. Vladimir A. Richter PhD Deputy Director, ICBFM SB RAS