Characterization of circulating DNA as a biomarker for genetic aberrations in humans

Maniesh van der Vaart Hons. B.Sc

Dissertation submitted in partial fulfilment of the requirements for the degree *Master of Science* at the North-West University

Supervisor: Prof. P.J. Pretorius

Potchefstroom Campus
2006

Every written masterpiece begins with a single word on a page.
Every long journey begins with a single step.
...Anonymous

Acknowledgements

My greatest appreciation goes to our Father in heaven without Whom none of this would have been possible.

I would like to thank the following people:

- Prof. Piet Pretorius for his guidance and support as promoter, for having patients and being understanding, for always having time for me and also for being a great mentor on personal as well as professional level.
- Dr. Marinda Oosthuizen at the center for paraclinical research of the University of Pretoria for making their equipment available to me.
- Prof. Fourie Joubert at the Bioinformatics and Computational Biology unit of the University of Pretoria for help with sequence analysis.
- Dr. Samual Fourie for providing clinical material of breast cancer patients.
- Dr. Oliver Preisig for informative discussions, late work and not holding back on protocols for methods and advice, also to your team at Inqaba Biotec for always responding promptly on questions and for friendly service.
- The Roche team: Anna Kenyon, Austen Cohen and Liegh Berrie for always being friendly and for help with a smile in crisis times.
- My parents, Henk and Martie van der Vaart, who granted me a university education, always believe in me and hold me in their prayers every day.
- My sister, Nadia van der Vaart, who are always there for me and know just what to say to make me feel better.
- Ruan Steyn, who kept me thinking straight and love me no matter what.
- My friends: Althea Veldman, Anchen van der Walt and Carina van der Walt, who always listens; for support and encouragement when it was needed most and for your valuable friendship.
- Everyone at Biochemistry for advice, friendship and just for being there.

It is much appreciated!

Abstract

Circulating DNA is fragments of DNA which can be found in the blood of healthy as well as diseased individuals. Higher levels of these nucleic acid molecules can be found in diseased and pregnant individuals in contrast to healthy controls. The origin of circulating DNA has not been elucidated, but release of DNA after apoptosis or necrosis or active release by living cells has been hypothesized. It was concluded in this study that apoptosis or necrosis may only be a minor source of circulating DNA and that release of DNA by living cells might play a major role in the origin, while disturbance of the equilibrium between release by living cells and clearance mechanisms may cause the rise in the levels of circulating DNA observed in different conditions.

Before circulating DNA can be analyzed, it has to be isolated from the blood. A number of different preanalytical conditions can have an impact on the quantity and quality of circulating DNA that can be isolated. Furthermore, the choice of isolation and quantification method may also influence the results obtained. Quantitative analysis of circulating DNA was done by real-time PCR analysis of the ß-*Globin* gene and the DNA levels obtained for healthy controls and cancer patients correlated with levels reported in the literature.

Characterization of total circulating DNA may be beneficial in diagnosis and prognosis and may also contribute to determining the source and function of circulating DNA. In order for characterization to take place a method to clone total circulating DNA was developed and standardized and thirty-five clones were obtained and analyzed. It was found that the sequences contain a large amount of Alu repeats and the significance of this has not been determined yet. This is a first step towards future studies.

Keywords: Circulating free DNA, cancer, plasma, characterization.

// Opsomming

Sirkulerende DNA is kort fragmente DNA teenwoordig in die bloed van siek asook gesonde mense. Hoër vlakke van hierdie nukliënsuur molekules kan by siek mense en swanger vroue gevind word, in teenstelling met gesonde mense. Die oorsprong van sirkulerende DNA is nog nie bevestig nie, maar 'n paar hipoteses is gevormuleer naamlik vrystelling van DNA na apoptose of nekrose, of aktiewe vrystelling van DNA deur lewende selle. Tydens hierdie sudie is bevind dat apoptose en nekrose slegs 'n klein aandeel in die oorsprong van sirkulerende DNA het en dat vrystelling van DNA deur lewende selle 'n groot rol kan speel in die oorsprong van hierdie molekules. Versteuring van die balans tussen vrystelling en verwyderings meganismes kan tot gevolg hê dat sirkulerende DNA vlakke tydens verskillende toestande verhoog.

Voordat sirkulerende DNA geanaliseer kan word, moet dit eers uit die bloed geïsoleer word. 'n Verskeidenheid verkillende toestande kan 'n invloed hê op die hoeveelheid en kwaliteit van die geïsoleerde DNA voordat DNA geïsoleer word. Die keuse van isolerings metode kan ook 'n invloed op die resultate hê. Kwantitatiewe analise van sirkulerende DNA is deur "Real-time PCR" analise van die β-*globien* geen gedoen en die resultate wat verkry is vir beide gesonde en siek persone vergelyk met die gepubliseerde resultate van ander groepe.

Karakterisering van totale vry DNA kan voordelig wees vir diagnose sovel as prognose, dit mag verder ook bydra om die oorsprong en funksie van sirkulerende DNA op te klaar. Om karakterisering moontlik te maak is 'n metode om totale sirkulerende DNA te kloon, ontwikkel en gestandardiseer. Vyf en dertig klone is geanaliseer en die volgordes bevat 'n groot hoeveelheid Alu herhalings, maar die betekenis hiervan is nog nie bekend nie. Die weg is gebaan vir verdere studie.

Sleutelwoorde: Sirkulerende vry DNA, kanker, plasma, karakterisering.

Table of contents

ACKNOWLEDGEMENTS ii

ABSTRACT iii

OPSOMMING iv

CHAPTER 1: INTRODUCTION 1

1.1 Problem statement and substantiation 1

1.2 Research aims and objectives 2

1.3 Structure of this dissertation 2

CHAPTER 2: LITERATURE SURVEY 4

2.1 Manuscript prepared for submission
The origin of circulating free DNA 4

CHAPTER 3: MATERIALS AND METHODS 20

3.1 Blood processing 21

3.2 DNA isolation
3.2.1 Phenol/Chloroform method 21
3.2.2 Salting-out method 21
3.2.3 QIAamp blood mini kit 21
3.2.4 MagNA Pure Compact system 22

3.3 Cloning
3.3.1 Blunting 22
3.3.2 Phosphorylation 23
3.3.3 Vector ligation 23
3.3.4 Colony PCR 24
3.3.5 Transformation of competent cells, growing and collecting colonies 24
3.3.6 Sequencing 25
3.3.7 Analysis of sequences 25

3.4 Real-Time PCR
3.4.1 Blood processing and DNA isolation 26
3.4.2 B-*globin* quantification 26

3.5 Agarose gel electrophoresis **28**

3.6 Manuscript prepared for submission
Different methods used for scrutinizing circulating nucleic acids in plasma or serum 29

CHAPTER 4: RESULTS AND DISCUSSION **55**

4.1 Manuscript prepared for submission
A method for characterization of total circulating DNA 55

CHAPTER 5: SUMMARY AND CONCLUSIONS **69**

REFERENCES **71**

ADDENDA **72**

Addendum A
List of figures 72
List of tables 72
Abbreviations 73

Addendum B
Conference proceedings 74

Addendum C
Buffers and media 76

Addendum D
Calculation of Real-time PCR data 78
Real-time PCR Amplification curves 79

Addendum E
Raw sequence data 80

Addendum F
Annotation of Repbase Sequences 85

Chapter 1

Introduction

1.1 Problem statement and substantiation

DNA fragments can be detected in the blood, serum and urine of healthy as well as individuals with a number of conditions. Due to the fact that cancer patients have much higher concentrations of circulating DNA, analysis of these fragments is a very useful tool for noninvasive diagnosis, early detection and a wide variety of other applications (Wu *et al.*, 2002, Ziegler *et al.*, 2002, Chan *et al.*, 2003, Bremnes *et al.*, 2005). Through the years it was demonstrated that tumor-related DNA is not confined to a specific cancer type. It appear to be a common finding across different malignancies, mutant DNA has been found in lung, head and neck, colorectal, gastric, pancreatic, liver, biliary tree, skin, breast, kidney, ovarian, cervical, bladder and prostate cancers as well as hematological malignancies including lymphomas (Bremnes *et al.*, 2005).

The most promising way to improve cancer prognosis is by means of early detection, through identification of the cancer at a stage early enough to be curable by surgery, although the lack of valid screening methods have remained a problem (Bremnes *et al.*, 2005). DNA alterations is present in serum/plasma at an early stage for some cancers and can be used for early detection which can lead to curing (Wu *et al.*, 2002).

To aid the process of early detection the characterization of circulating DNA is vital, this can identify new markers for detection and may be a detection tool in itself. The determining of the nature and appearance of circulating DNA can lead to important developments in this field.

In essence the literature on circulating DNA either describes the origin of circulating DNA for which a conclusion has not been reached (Ziegler *et al.*, 2002, Chan *et al.*, 2003, Chen *et al.*, 2005, Georgiou *et al.*, 2005, McLaren Howard, 2005), or it describes methods for detecting changes or quantifying circulating DNA in different conditions. A gap in the literature exists when it comes to characterization of the total circulating DNA, addressing this issue may shed light on the origin as well as diagnostic and predictive value of these circulating nucleic acid molecules.

Ethics approval for this study was obtained from the Ethics Committee of the North-West University with the title "DNA methylation and breast cancer: a case control study of prognostic outcome measurements", reference number 05M12.

1.2 Research aims and objectives

The aim of this project was to isolate, clone and sequence the total circulating DNA in the plasma of healthy individuals to aid characterization of circulating DNA. Additionally, circulating DNA was quantified and obtained sequences were analyzed as a first step towards finding a biomarker for genetic aberrations in humans.

1.3 Structure of this dissertation

This dissertation was compiled in article form and consists of three unpublished manuscripts which are formatted according to the instructions for authors to Clinical Chemistry, International Journal of Molecular Diagnostics and Laboratory Medicine. The origin of free circulating DNA and the factors which may play a role in the release and clearance of circulating DNA is discussed in chapter Two. Chapter Three consists of a short description of the methods used to isolate, clone and quantify circulating DNA in this study, which is followed by an overview of the factors influencing the concentration and quality of circulating DNA before

isolation, as well as the methods used in the literature with the quantification results obtained by different research groups and laboratories. Chapter Four contains the last manuscript which describes the method for characterization of circulating DNA and the results obtained in this study. A summary and conclusions are given in chapter Five followed by the references used, in general, throughout this dissertation. Note that the references for each manuscript is included as part of the manuscript and not in the general references. Additional information consisting of lists of figures, tables and abbreviations (Addendum A), conference proceedings resulting from this study (Addendum B), real-time PCR data with amplification curves and the standard curve (Addendum D) as well as sequence data (Addendum E), information about Repbase sequences (Addendum F) and the composition of the buffers and media mentioned in chapter Three (Addendum C) are given at the end of this document.

Chapter 2

Literature survey

2.1 Manuscript prepared for submission

The origin of circulating free DNA

M. van der Vaart and P.J. Pretorius

This is the first draft of a review article prepared for submission to Clinical Chemistry, International Journal of Molecular Diagnostics and Laboratory Medicine. The instructions for authors to this journal are available at: http://www.clinchem.org/info_ar/info_a_outline.shtml

The origin of circulating free DNA

Maniesh van der Vaart and Piet, J. Pretorius[1]

Abstract

Background: 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 are still not clear.

Methods: Views from many different research groups about the origin of free circulating DNA were compared, and information from related fields was combined with what is known to compile a thorough review of the literature available to us.

Results: Although most research groups claim that apoptosis or necrosis is the source of circulating DNA, we like to present evidence that does not support this notion.

Conclusion: Apoptosis and necrosis are not the major source of circulating DNA. 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. 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.

School of Biochemistry, North-West University, Potchefstroom Campus, South Africa.
[1] Address correspondence to this author at: e-mail piet.pretorius@nwu.ac.za,
Fax +2718-299-2316, Tel +2718-299-2309

Introduction

It is widely known that higher concentrations of free circulating DNA can be found in the blood of patients with malignant diseases compared to healthy subjects (1). Several studies have been performed to establish whether a significant diagnostic and/or prognostic (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 minireview will address some aspects on the origin of circulating DNA in pathological 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. The release of DNA by living cells will also 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 (2-4). Furthermore it was also in these early days that the low molecular weight and ladder pattern of circulating DNA was revealed by agarose gel electrophoresis (3). 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 – >10 000bp (5). Furthermore it is possible that the ends of these molecules are capped or that these DNA fragments are circulating in the form of nucleosomes or apoptotic bodies (6, 7).

The amount and composition of circulating DNA varies between patients (5) and many reports agree (8, 9) that cancer patients have much larger amounts of circulating DNA in their blood than healthy subjects without severe diseases. Early analysis could not detect any DNA in serum of healthy individuals (3, 10).

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 pathological conditions such as systemic lupus erythematosus (SLE), rheumatoid arthritis, glomerulonephritis, pancreatitis, cholelithiasis, inflammatory bowel disease, peptic ulcer disease, hepatitis, oesophagitis, pulmonary embolism, ulcerative colitis and miliary tuberculosis (for review see (11) and (6)) which are associated with inflammatory processes, other cases that involve increased cell death include trauma, stroke, myocardial infarction, angina (12), sepsis and septic shock (13). Furthermore, over trained athletes (14) also show increased amounts of circulating DNA while fetal DNA can be detected during pregnancy (15). Can the fact that circulating DNA occur in increased amounts in so many different conditions, point to some correlation between them that may reveal a similar mechanism of release or origin?

Many articles report that oncogene mutations and amplifications, microsatellite alterations and epigenetic changes like DNA methylation (for review see (1, 6, 16, 17)) 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 (4, 11, 18, 19). A direct relationship could not be demonstrated between the amount of plasma DNA and the type or clinical status of the cancer (5).

Two main points of view exist in the literature for explaining the origin of circulating free DNA, i.e. these DNA fragments enter the bloodstream following cell death or they are released by living cells (1, 3, 5, 6, 13, 20-24).

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 (25). During apoptosis DNA degradation often occur: 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 (26-28). Because this ladder pattern is also visible after electrophoresis of circulating DNA many believe that apoptosis is the source of the observed DNA fragments in the plasma (5, 23, 29). 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 (30) and the DNA is consequently completely digested into nucleotides by DNase II in lysosomes (26, 28). Thus the possibility exists that DNA fragments released by apoptosis are removed, without a trace, before it can appear in the circulation (5, 23). If this engulfment of apoptotic bodies is impaired or cell death is amplified, tissue injury or autoimmunity will most probably result (23, 30, 31).

Following 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 (23). This implicates macrophages in the generation of circulating DNA. 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 after electrophoresis for both necrotic and apoptotic cells respectively. This may indicate that necrotic human cells are engulfed by mice macrophages or that DNA from necrotic cells are cleaved by the same enzymes functioning in apoptosis thus causing the same ladder pattern as apoptotic cells (23). PCR analysis of circulating DNA in the blood of these mice showed the presence of both human and murine sequences (23). It is possible that macrophage apoptosis is caused by engulfment of high numbers of dead cells or by impaired phagocytic function (23) 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 was 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 (11) and several ingenious mechanisms have been identified in which cancer cells become resistant to apoptosis (32) in order to escape the immune system (33). Various targets for therapeutic intervention in cancer has been explored and many of them is the preferential induction of apoptosis to eliminate cancer cells without affecting normal cells (for review see (33)).

In contrast to apoptosis, necrosis causes random, nonspecific and incomplete digestion of DNA and thus a smear is observed in an electrophoresis gel (22, 27, 34). 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 much more circulating cells should have been present in the blood since the amount of DNA in the plasma undoubtedly exceeds the amount of circulating cells (1, 11, 21), 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. They 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 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 (1, 11, 13, 21, 24) but convincing evidence does not exist to prove this hypothesis. It is quite astounding that even though Anker *et al.* (2) realized the possibility that DNA can be actively released by cells more than 30 years ago, 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.* (24): (a) Instead of increased circulating DNA which is expected if apoptosis is the mechanism of release, Leon *et al.* (10) 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.* (24) agrees with Anker *et al.* (2) who observed in 1975 that human blood lymphocytes actively releases 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. Anker also stated that the quantities of DNA that is released are similar no matter if a quarter of the cells or none at all die, except for cancer cells which can release more DNA than normal cells (11), this shows that cell death are not responsible for the DNA in the plasma. Furthermore, Stroun *et al.* (21) 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 raise significantly because of the amount of proliferating cells that increase and not 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 includes normal and malignant cells in the body.

Other sources of circulating DNA

Based on the observations made by Raptis *et al.* (35) in the early 80's, an exogenous source of free circulating DNA was excluded. However these observations were proven to be wrong by the presence of viral DNA circulating in the plasma of some patients with cancers associated with viral infection, like nasopharyngeal carcinoma where Ebstein-Barr virus DNA can be detected in 96% of cases, cervical cancer where human papillomavirus DNA can be detected in 50% of cases and hepatocellular carcinoma where hepatitis B virus DNA can be detected (36, 37).

Cells that lost their nuclei but remain functional underwent a process termed denucleation or terminal differentiation (38), according to Bischoff *et al.* (7) this may be another source of circulating DNA, but many tumors don't express enough of the molecular or morphological markers of the terminally differentiated state (39) to prove presence 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, because Lo *et al.* (40) observed that the clearance of fetal DNA after delivery occurs in an initial rapid phase followed by a slower second phase, in most of the women all DNA was cleared 2 hours after delivery. Despite this rapid clearance 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 in which fetal DNA is released exceeds its clearance rate. The concentration of fetal DNA thus provides an almost real-time picture of the interaction between DNA release and DNA clearance (40). A mean half life of only 16.3 minutes, because of the rapid metabolism of fetal DNA, was estimated by Bianchi *et al.* (41).

Although the mechanism of clearance has not been elucidated, a few possibilities may be explored. Since free DNA can be detected in urine (19, 42) the kidneys can be expected to play a role in clearance, animal studies also suggested that the liver, spleen and kidneys may be responsible for removal of circulating DNA (40).

Lo *et al.* (40) explored the possibility that plasma nucleases may have a function but proved that it only have a partial role in the removal of circulating fetal DNA. Free circulating DNA found in healthy people, cancer patients and organ transplant recipients most likely have the same clearance mechanism(s) as in maternal plasma, thus it will also be rapidly removed and also display a real-time picture of release which may be useful in monitoring disease and transplant efficiency (40).

Chelobanov *et al.* (43) summarize 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 to possibly be degraded to mononucleotides or be transported 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 inhibit binding of DNA, this may be because of the 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 (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.

Pisetsky *et al.* (23) showed that circulating DNA appeared in the blood within hours of administering dead cells intraperitoneally to mice and that the levels return to the base line after 24 hours.

The mechanism(s) whereby clearance of circulating DNA is achieved is at this point in time still poorly understood (44).

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 (29).

Patients with untreated active systemic lupus erythematosus have much higher concentrations of circulating DNA in their plasma than healthy individuals, but this decline to normal levels after treatment (35). 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 declined (42). 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 factions, i.e. either originating from the bone marrow or from the rest of the body, it was found that a significantly higher concentration originated from the bone marrow (45).

Chang *et al.* (46) observed a 10-fold increase relative to controls in the amount of circulating DNA in patients suffering from a myocardial infarction. They contribute 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 that suffered a myocardial infarction showed a more diffused ladder pattern for the latter (46). It is, however, possible that this may be an artifact 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.* (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 (15, 41). With 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 placental breakdown (7). An interesting observation is that circulating DNA in pregnant woman has a much wider size distribution than circulating DNA in nonpregnant woman (47). According to Bianchi *et al.* (41) the majority of fetal DNA during pregnancy originates from the placenta, but other sources are also possible.

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 (14). An increase in oxidative stress was also observed after exercise overtraining (14), it is known that reactive oxygen species (ROS) cause DNA strand breaks, but can it be implicated in the production of circulating DNA?

A highly significant difference between the concentration of plasma circulating DNA was found between healthy individuals, 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 those who did not develop complications which prove that it may be a valuable prognostic marker in trauma patients (44).

Circulating plasma DNA concentrations was 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 five 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

(48). Although the origin of the circulating DNA in stroke is unknown, ROS is also produced (48) and may be involved in the generation of free circulating DNA, as mentioned above.

Patients with severe sepsis or septic shock had significantly higher circulating plasma DNA concentrations than in normal control persons and even higher concentrations were found in those who did not survive intensive care and those who needed renal or inotropic 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 (13).

Conclusions

Almost every paper on circulating DNA state that apoptosis and/or necrosis is the source of free circulating DNA in serum and plasma because of the ladder pattern of DNA revealed by electrophoresis. However, this ladder pattern can also be found in the culture medium in which lymphocytes grew. 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 free DNA. Radiotherapy or irradiation, chemotherapy and other cancer treatments cause cell death by apoptosis (49) and the amount of circulating DNA is less in cancer patients under treatment than 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. We thus 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 70's (10) and evidence that DNA is released *in vitro* by human blood lymphocytes has been given in the mid 70's (2). To our knowledge it has not been proven that DNA can be released into the circulation system *in vivo* by living cells, but we don't anticipate a reason why this isn't possible. 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 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 degradation 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 found 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. Even though 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 the significance of and effect that these circulating DNA fragments have in the body.

References

1. Goebel G, Zitt M, Zitt M, Muller HM. Circulating nucleic acids in plasma or serum (CNAPS) as prognostic and predictive markers in patients with solid neoplasias. Dis Markers 2005;21:105-20.
2. Anker P, Stroun M, Maurice PA. Spontaneous release of DNA by human blood lymphoctyes as shown in an in vitro system. Cancer Res 1975;35:2375-82.
3. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA. Isolation and characterization of DNA from the plasma of cancer patients. Eur J Cancer Clin Oncol 1987;23:707-12.
4. Stroun M, Anker P, Maurice P, Lyautey J, Lederrey C, Beljanski M. Neoplastic characteristics of the DNA found in the plasma of cancer patients. Oncology 1989;46:318-22.
5. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, Knippers R. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659-65.
6. Ziegler A, Zangemeister-Wittke U, Stahel RA. Circulating DNA: a new diagnostic gold mine? Cancer Treat Rev 2002;28:255-71.
7. Bischoff FZ, Lewis DE, Simpson JL. Cell-free fetal DNA in maternal blood: kinetics, source and structure. Hum Reprod Update 2005;11:59-67.
8. Huang ZH, Li LH, Hua D. Quantitative analysis of plasma circulating DNA at diagnosis and during follow-up of breast cancer patients. Cancer Letters 2006;243:64-70.
9. Sang F, Ren J. Comparisons between capillary zone electrophoresis and real-time PCR for quantification of circulating DNA levels in human sera. Journal of Chromatography B 2006;838:122-8.
10. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 1977;37:646-50.
11. Anker P, Mulcahy H, Chen XQ, Stroun M. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. Cancer Metastasis Rev 1999;18:65-73.
12. Rainer TH, Lam NYL, Man CY, Chiu RWK, Woo KS, Lo YMD. Plasma [beta]-globin DNA as a prognostic marker in chest pain patients. Clinica Chimica Acta 2006;368:110-3.
13. Rhodes A, Wort SJ, Thomas H, Collinson P, Bennett ED. Plasma DNA concentration as a predictor of mortality and sepsis in critically ill patients. Crit Care 2006;10:R60.
14. Fatouros IG, Destouni A, Margonis K, Jamurtas AZ, Vrettou C, Kouretas D, et al. Cell-free plasma DNA as a novel marker of aseptic inflammation severity related to exercise overtraining. Clin Chem 2006;52:1820-4.
15. Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768-75.

16. Bremnes RM, Sirera R, Camps C. Circulating tumour-derived DNA and RNA markers in blood: a tool for early detection, diagnostics, and follow-up? Lung Cancer 2005;49:1-12.
17. Tong Y-K, Lo YMD. Diagnostic developments involving cell-free (circulating) nucleic acids. Clinica Chimica Acta 2006;363:187-96.
18. Shaw JA, Smith BM, Walsh T, Johnson S, Primrose L, Slade MJ, et al. Microsatellite alterations in plasma DNA of primary breast cancer patients. Clin Cancer Res 2000;6:1119-24.
19. Su YH, Wang M, Brenner DE, Ng A, Melkonyan H, Umansky S, et al. Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. J Mol Diagn 2004;6:101-7.
20. Fournie GJ, Courtin JP, Laval F, Chale JJ, Pourrat JP, Pujazon MC, et al. Plasma DNA as a marker of cancerous cell death. Investigations in patients suffering from lung cancer and in nude mice bearing human tumours. Cancer Lett 1995;91:221-7.
21. Stroun M, Lyautey J, Lederrey C, Olson-Sand A, Anker P. About the possible origin and mechanism of circulating DNA: Apoptosis and active DNA release. Clinica Chimica Acta 2001;313:139-42.
22. Wang BG, Huang HY, Chen YC, Bristow RE, Kassauei K, Cheng CC, et al. Increased plasma DNA integrity in cancer patients. Cancer Res 2003;63:3966-8.
23. Pisetsky DS. The immune response to cell death in SLE. Autoimmun Rev 2004;3:500-4.
24. Chen Z, Fadiel A, Naftolin F, Eichenbaum KD, Xia Y. Circulation DNA: Biological implications for cancer metastasis and immunology. Medical Hypotheses 2005;65:956-61.
25. Kanduc D, Mittelman A, Serpico R, Sinigaglia E, Sinha AA, Natale C, et al. Cell death: apoptosis versus necrosis (review). Int J Oncol 2002;21:165-70.
26. Nagata S, Nagase H, Kawane K, Mukae N, Fukuyama H. Degradation of chromosomal DNA during apoptosis. Cell Death Differ 2003;10:108-16.
27. Holdenrieder S, Stieber P. Apoptotic markers in cancer. Clinical Biochemistry 2004;37:605-17.
28. Nagata S. DNA degradation in development and programmed cell death. Annu Rev Immunol 2005;23:853-75.
29. Atamaniuk J, Ruzicka K, Stuhlmeier KM, Karimi A, Eigner M, Mueller MM. Cell-free plasma DNA: a marker for apoptosis during hemodialysis. Clin Chem 2006;52:523-6.
30. Viorritto IC, Nikolov NP, Siegel RM. Autoimmunity versus tolerance: Can dying cells tip the balance? Clin Immunol 2006.
31. Ren Y, Savill J. Apoptosis: the importance of being eaten. Cell Death Differ 1998;5:563-8.
32. Igney FH, Krammer PH. Immune escape of tumors: apoptosis resistance and tumor counterattack. J Leukoc Biol 2002;71:907-20.
33. Bremer E, van Dam G, Kroesen BJ, de Leij L, Helfrich W. Targeted induction of apoptosis for cancer therapy: current progress and prospects. Trends Mol Med 2006;12:382-93.

34. Georgiou CD, Patsoukis N, Papapostolou I. Assay for the quantification of small-sized fragmented genomic DNA. Anal Biochem 2005;339:223-30.
35. Raptis L, Menard HA. Quantitation and characterization of plasma DNA in normals and patients with systemic lupus erythematosus. J Clin Invest 1980;66:1391-9.
36. Lo YM, Chan LY, Lo KW, Leung SF, Zhang J, Chan AT, et al. Quantitative analysis of cell-free Epstein-Barr virus DNA in plasma of patients with nasopharyngeal carcinoma. Cancer Res 1999;59:1188-91.
37. Yang HJ, Liu VW, Tsang PC, Yip AM, Tam KF, Wong LC, et al. Quantification of human papillomavirus DNA in the plasma of patients with cervical cancer. Int J Gynecol Cancer 2004;14:903-10.
38. Gagna CE, Kuo HR, Florea E, Shami W, Taormina R, Vaswani N, et al. Comparison of apoptosis and terminal differentiation: the mammalian aging process. J Histochem Cytochem 2001;49:929-30.
39. Mueller E, Sarraf P, Tontonoz P, Evans RM, Martin KJ, Zhang M, et al. Terminal differentiation of human breast cancer through PPAR gamma. Mol Cell 1998;1:465-70.
40. Lo YM, Zhang J, Leung TN, Lau TK, Chang AM, Hjelm NM. Rapid clearance of fetal DNA from maternal plasma. Am J Hum Genet 1999;64:218-24.
41. Bianchi DW. Circulating fetal DNA: its origin and diagnostic potential-a review. Placenta 2004;25 Suppl A:S93-S101.
42. Zhang J, Tong KL, Li PK, Chan AY, Yeung CK, Pang CC, et al. Presence of donor- and recipient-derived DNA in cell-free urine samples of renal transplantation recipients: urinary DNA chimerism. Clin Chem 1999;45:1741-6.
43. Chelobanov BP, Laktionov PP, Vlasov VV. Proteins involved in binding and cellular uptake of nucleic acids. Biochemistry (Mosc) 2006;71:583-96.
44. Lo YM, Rainer TH, Chan LY, Hjelm NM, Cocks RA. Plasma DNA as a prognostic marker in trauma patients. Clin Chem 2000;46:319-23.
45. Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YM. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem 2002;48:421-7.
46. Chang CP, Chia RH, Wu TL, Tsao KC, Sun CF, Wu JT. Elevated cell-free serum DNA detected in patients with myocardial infarction. Clin Chim Acta 2003;327:95-101.
47. Chan KC, Zhang J, Hui AB, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem 2004;50:88-92.
48. Rainer TH, Wong LK, Lam W, Yuen E, Lam NY, Metreweli C, Lo YM. Prognostic use of circulating plasma nucleic acid concentrations in patients with acute stroke. Clin Chem 2003;49:562-9.
49. Debatin KM, Krammer PH. Death receptors in chemotherapy and cancer. Oncogene 2004;23:2950-66.

Chapter 3

Materials and methods

This diagram is a structured representation of the flow of practical work that was done and the various methods that was used in this study

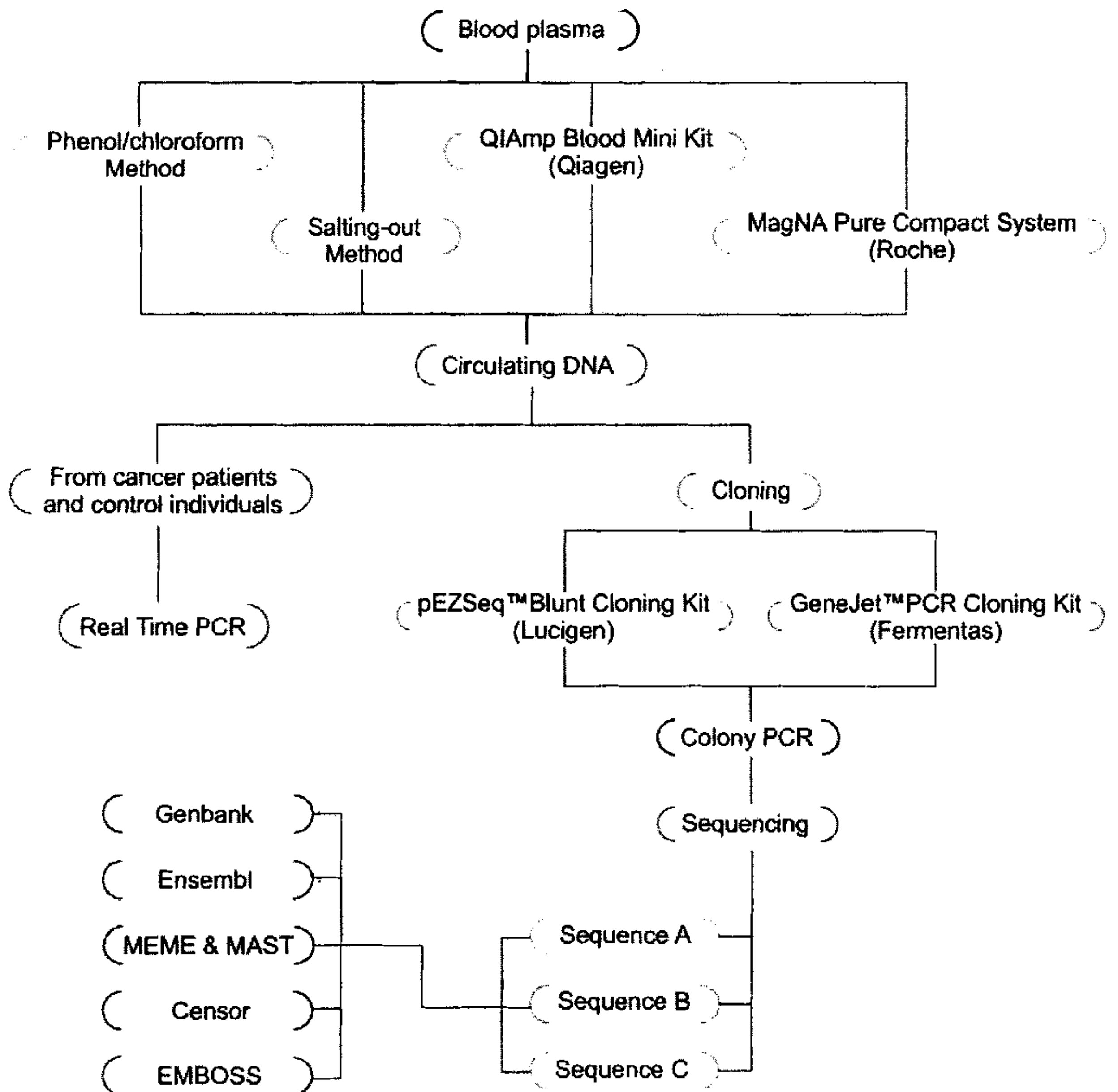

Fig. 3.1 Flow diagram of experimental work

3.1 Blood processing

Blood was collected by venipuncture by AMPATH from various willing healthy individuals into 4 ml BD Vacutainer EDTA tubes from Transgen (REF 368861), it was processed immediately. Centrifugation was carried out in a swinging bucket rotor at 1250g for 20 minutes after which the top plasma part were collected and frozen at -20°C until further use.

3.2 DNA isolation

3.2.1 Phenol/Chloroform method

A starting volume of 1.5 ml plasma was used and most proteins were digested during SDS and proteinase K incubation, after which multiple phenol/chloroform (1:1) extractions was done followed by ethanol precipitation aided by glycogen and Ammonium acetate, precipitated DNA was washed and resuspended (Anker and Stroun, 2001).

3.2.2 Salting-out method

A starting volume of 1 ml plasma was incubated with EDTA, NaCl, SDS and proteinase K. Proteins was precipitated by adding water saturated with sodium chloride after which DNA was ethanol precipitated overnight followed first by phenol/chloroform extraction and then chloroform extraction, subsequently DNA was ethanol precipitated, washed and resuspended (Schmidt *et al.,* 2005).

3.2.3 QIAamp blood mini kit

The QIAamp DNA blood mini kit (Qiagen, # 51104) was obtained from Southern Cross Biotechnology. Isolation was done according to the blood and body fluid protocol supplied with the kit.

3.2.4 MagNA Pure Compact system

The MagNA Pure Compact system (Roche Applied Science) with the MagNA Pure Compact Nucleic Acid Isolation Kit I (Roche Applied Science, # 03730964001) was used to isolate circulating DNA from plasma. DNA was eluted in either 50 µl or 100 µl.

3.3 Cloning

In order to be able to sequence the circulating DNA, a method had to be established to attach small parts of known sequence to the ends of the circulating DNA for sequencing primers to bind to and different single DNA fragments had to be separated for it to be sequenced successfully. Furthermore, the configuration of the ends of the isolated circulating DNA fragments was unknown and it could not be attached to the pEZSeq™ vector (Lucigen® Corporation, # 40464-2), which supplied the known M13 binding areas for the sequencing primers, without blunting and phosphorylation. The normal blunting protocol suggested by Fermentas was not able to blunt the ends of the circulating DNA sufficiently and this protocol had to be altered as described below.

3.3.1 Blunting

T4 DNA Polymerase (Fermentas, # EP0061) was used to blunt the isolated circulating DNA. Alterations made to the supplier's protocol is as follows: 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 Eppendorf tube and incubated at 11°C for 5 minutes. This was followed by adding 0.2 µl of Nucleotide Solution Mix (10 mM ea. dNTP, New England BioLabs®$_{Inc}$ # N0447S). This mixture was incubated for a further 15 minutes at 21°C (room temperature) after which the enzyme was heat inactivated at 70°C for 10 minutes and kept on ice until the phosphorylation step was performed.

3.3.2 Phosphorylation

The 5'-OH group of the double stranded circulating DNA had to be phosphorylated in order for it to be able to be cloned into the vector, this reaction was carried out wit T4 Polynucleotide Kinase (Fermentas, # EK0031). To the blunting reaction mix described above, 0.55 µl ATP (0.04 µmol/ml) and 1 µl T4 Polynucleotide Kinase were added, this mixture was vortexed briefly and then centrifuged for a few seconds and subsequently incubated at 37°C for 30 minutes. This was followed by a single chloroform extraction.

3.3.3 Vector ligation

The pEZSeq™ Blue/White Cloning Kit from Lucigen® Corporation (# 40464-2) was used because of its high cloning efficiency. The vector is supplied with blunt and dephosphorylated ends and it contains a minimal amount of vector DNA between the sequencing primers (M13) and the cloning site. Ligation was done by mixing 6.5 µl of the chloroform extracted, blunted and phosphorylated DNA with 2.5 µl 4x pEZSeq™ vector premix and 1 µl CloneSmart DNA ligase supplied in the kit. This mixture was incubated at room temperature for 2 hours followed by heat inactivation of the enzyme at 70°C for 15 minutes. The mix was then cooled to room temperature for 15 seconds followed by cooling on ice for a further 15 seconds after which it was centrifuged at 13523 x g for 1 minute in a table top Eppendorf centrifuge and was then stored at 4°C until further use.

Alternatively, a sample of the GeneJET™ PCR Cloning Kit (Fermentas # K1221) was obtained from Inqaba Biotec. The pJET1/blunt vector in this kit is linearized and phosphorylated at the 5' end, thus no phosphorylation of the circulating DNA is necessary. Furthermore, pJET1/blunt is a positive selection vector and only recombinant plasmids containing insert DNA will result in viable colonies, which simplify colony screening. The same blunting protocol, as described above, was followed when cloning was performed using this vector and ligation was done according to the blunt-end protocol of the kit. When the pJET1/blunt vector was used for cloning, pJET1 forward and reverse primers, supplied with the kit, was used for colony PCR.

3.3.4 Colony PCR

In order to make sure that the circulating DNA was ligated to the vector, colony PCR was performed on the ligation products. Even though theoretically this is not colony PCR when performed on the ligated DNA, we call it colony PCR because the same protocol is followed and there is no need to describe the same protocol twice as it is later used on bacterial colonies. This protocol was kindly supplied to us by Inqaba Biotec. A master mix of 4.75 µl water, 2.5 µl 10x PCR buffer, 0.5 µl dNTP's, 1 µl of both forward and reverse M13 primers (10 µM, manufactured by Inqaba Biotec) and 0.25 µl Taq polymerase (Promega, # M186E) was put together. When this PCR was performed on the ligated DNA the 15 µl of water normally in the PCR tube was decreased by 2.5 µl and included in the master mix and 2.5 µl of ligated DNA was added to the master mix. This mix was vortexed and centrifuged for a few seconds and the PCR protocol was started, it consisted of the following:

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

The PCR products were subsequently analyzed with agarose gel electrophoresis.

3.3.5 Transformation of competent cells, growing and collecting colonies

Either *E. cloni*® Chemically Competent Cells (Lucigen® Corporation, # 60107-1) or SURE cells (gift from Japie Fourie) where thawed completely on wet ice, 50 -100 µl cells was transferred to a chilled 15 ml tube (Labcon) and 5 µl ligated DNA was added and stirred with pipette tip. This cell mixture was incubated for 30 minutes on ice. This was followed by heat shocking cells for 45 seconds in a 42°C waterbath and returning cells to ice for 2 minutes. Cells was incubated at 37°C for 1 hour in a shaking incubator at 250 rpm after addition of 200 µl of room

temperature Recovery Medium or SOC was added. Transformed cells (100 µl) were plated out on YT +Amp plates without IPTG or X-gal because it was found to interfere with efficient cloning. Plates were incubated overnight at 37°C. Colonies were transferred to master plates and colony PCR as described above were performed after touching each colony with a sterile pipette tip and rinsing it in the 15 µl water in the PCR tube.

Frozen stocks of colonies were made for later use after growing overnight cultures by inoculating 3 ml of LB +Amp culture media with a single colony, 700 µl of the overnight stock were mixed with 300 µl of 50% glycerol for each colony and frozen at -80°C.

3.3.6 Sequencing

Colony PCR products were sent to Inqaba Biotech where it was cleaned with shrimp Alkaline Phosphatase (# EF0511) and Exonuclease I from E. coli (# EN0581) and sequenced with a Genetic analysis system SCE2410 (SpectruMedix) using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and M13 sequencing primers. BaseSpectrum V2.1.1 (SpectruMedix) software was used for analysis. The sequence of the M13 primers is the following:

M13 Forward: 5'-AGCGGATAACAATTTCACACAGGA-3'
M13 Reverse: 5'-CGCCAGGGTTTTCCCAGTCACGAC-3'

3.3.7 Analysis of sequences

Sequences were analyzed after a general course in Bioinformatics was attended at the Bioinformatics and Computational Biology unit of the University of Pretoria, under the guidance of Prof. Fourie Joubert.

3.4 Real-Time PCR

3.4.1 Blood processing and DNA isolation

In order to compare DNA concentrations obtained by real-time PCR with other publications and laboratories, either one or two tubes of blood was collected respectively from 10 breast cancer patients treated at the Oncology Unit of WIlmed Park Hospital by Dr. Samuel J. Fourie and from 10 healthy individuals. Blood from the breast cancer patients was collected by the hospital staff and was processed there like described above and one tube of blood from each healthy individual was either processed like described above or it was centrifuged twice, first at 1600g for 10 minutes and then at 16000g for the same time in a microcentifuge to evaluate the difference in concentration between these two scenarios. Circulating DNA was isolated with the QIAamp DNA blood mini kit (Qiagen) with minor alterations to the blood and body fluid protocol, DNA was isolated from 200 µl plasma which was diluted with 400 µl water and 400 µl buffer AL (instead of 200 µl), according to Herrera *et al.* (2005) these alterations increase the yield of circulating DNA. DNA was eluted in 50 µl of elution buffer after incubation for 5 minutes at room temperature, isolated DNA was stored at 4°C until quantification.

3.4.2 B-*globin* quantification

The β-*globin* gene can be used for quantification of total circulating DNA because it is present in all nucleated cells in the human body (Lo *et al.*, 1998; Lui *et al.*, 2002). The sequence of the β-*globin* primers and dual-labeled fluorescent Taqman probe used for real-time PCR analysis to quantify circulating DNA was obtained from Lo *et al.* (1998), and is the following:

β-*globin* 354F: 5'-GTG CAC CTG ACT CCT GAG GAG A-3';
β-*globin* 455R: 5'-CCT TGA TAC CAA CCT GCC CAG-3';
β-*globin* 402T: 5'-(FAM)AAG GTG AAC GTG GAT GAA GTT GGT GG(TAMRA)-3'

The primers and probe were manufactured by Metabion through Roche and give a PCR product of 102 bp. Sequence data for the β-globin gene can be obtained from GenBank Sequence database (accession number: U01317).

A concentration range of diluted genomic DNA (Promega # G304A) was made with the following concentrations: 100 ng/µl, 50 ng/µl, 10 ng/µl, 50 pg/µl, 10 pg/µl, 5 pg/µl. Real-time PCR was performed on a LightCycler® 2.0 (6 Channel) with the LightCycler® FastStart DNA Master HybProbe kit (Roche, # 303248001). The 530/560 channel was monitored for fluorescence at the end of each extension step while the following protocol was used:

10 min 95°C

50 cycles of:

10 sec at 95°C (template denaturation)

15 sec at 57°C (primer annealing)

10 sec at 72°C (primer extension)

20 min at 40°C

A concentration gradient of $MgCl_2$ was done on the first run like described by the supplier, the optimal concentration was 5 mM $MgCl_2$ thus this concentration was used in subsequent runs. The genomic DNA concentration gradient were run in triplicate followed by duplicate runs with all the samples and the 10 ng/µl genomic DNA standard, which was used for normalization to the standard curve. Blanks were included in each run. The unknown concentrations were calculated by the LightCycler® software (version: LCS4 4.0.0.23) using the second derivative method. The mean and standard deviation between the two sample runs was calculated for all the samples and the quantification results determined by the LightCycler software was converted to ng/ml with the following equation:

$$C = Q\left(\frac{V_{DNA}}{V_{PCR}}\right)\left(\frac{1}{V_{ext}}\right)$$

Where C = target concentration in plasma (ng/ml);

Q = target quantity (ng) determined by the LightCycler;

V_{DNA} = total volume of DNA obtained after extraction (50 µl);

V_{PCR} = volume of DNA solution used in PCR (2 µl); and

V_{ext} = volume of plasma extracted (0.2 ml) (Lo *et al.* 1998).

3.5 Agarose gel electrophoresis

Agarose gel was prepared by weighing the appropriate amount of molecular grade agarose (Separations) and adding 1x TAE buffer. The mixture was microwaved in a closed container until melted, ethidium bromide (Boehringer Mannheim) was added to a final concentration of 0.5 µg/ml. The gel was poured into an electrophoresis tray and allowed to solidify. TEA buffer was added to the electrophoresis chamber until the gel was covered. Samples mixed with loading dye (consisting of sucrose, water and bromophenol blue) were loaded and electrophoresis was conducted at 60 V for 40-60 minutes. All the gels were photographed and viewed on a CHEMI GENIUS Q bio imaging system (Syngene).

3.6 *Manuscript prepared for submission*

Different methods used for scrutinizing circulating nucleic acids in plasma or serum

M. van der Vaart and P.J. Pretorius

This is the first draft of a review article prepared for submission to Clinical Chemistry, International Journal of Molecular Diagnostics and Laboratory Medicine. The instructions for authors to this journal are available at: http://www.clinchem.org/info_ar/info_a_outline.shtml

Different methods used for scrutinizing circulating nucleic acids in plasma or serum

Maniesh van der Vaart and Piet, J. Pretorius[1]

Abstract

Background: Many experimental factors influence the amount of circulating DNA that can be detected in individual mammalians. These factors include time before plasma separation, centrifugation protocols, storage conditions and isolation and quantification methods.

Methods: A literature review was done to identify and compare different preanalytical factors affecting isolation of free circulating DNA. Furthermore, results for a few isolation and quantification methods were compared.

Results: Circulating DNA in plasma of control subjects range between 0.2 and 20.8 ng/ml and in cancer patients between 3.7 and 819 ng/ml.

Conclusions: Before studies can be compared and conclusions can be reached about the levels of circulating DNA in different conditions, the other factors that may influence these levels need to be standardized.

Introduction

Free circulating DNA was isolated from blood of cancer patients for the first time in 1987 by Stroun *et al.* (1), but the technique used could not isolate any DNA from the plasma of healthy individuals. The levels of DNA that could be isolated from only 10 of the 37 cancer patients correlate with values obtained for normal individuals today. Earlier methods may not have been adequately sensitive or specific for measurement of circulating DNA, but the clinical importance that

---

School of Biochemistry, North-West University, Potchefstroom Campus, South Africa.
[1] Address correspondence to this author at: e-mail piet.pretorius@nwu.ac.za,
Fax +2718-299-2316, Tel +2718-299-2309

these circulating DNA fragments may have in diagnosis and monitoring of different conditions make the need for improved and standardized methods fundamental in the optimal use and characterization of circulating DNA. With the wide variety of protocols available today and the always increasing expertise the possibilities are endless.

In this minireview the different factors influencing the levels and quality of circulating DNA in the plasma, even before isolation of DNA, is compared. Furthermore, isolation and quantification methods and results for different research groups and laboratories are put side by side to form a bigger picture of work being done regarding circulating DNA.

Preanalytical preparation of samples

Both quantitative and qualitative analysis of circulating DNA requires uncontaminated starting material, whether serum or plasma (urine and other body fluids not excluded), which should thus be free of cellular DNA, DNA that is shed into the serum or plasma during handling or storage and may interfere with analysis or accurate quantification (2).

Changes in concentration during storage

Whether plasma is stored at -20°C for one month (2), -80°C for two weeks or if it is frozen and thawed up to three times, no significant difference in the yield of free DNA was observed (3). The DNA in the plasma does, however, turn out to be more fragmented after the freeze thaw cycles, but this is not the case with isolated circulating DNA stored in elution buffer (3). When storage of whole-blood is considered, reports differ: Jung et al. (4) did not find a significant difference in the plasma DNA concentration when blood samples were stored for 8 hours at room temperature or for 24 hours at 4°C, while Chan et al. (3) did find a significant increase in concentration. However, delaying separation of blood cells for up to 6 hours, did not have a significant effect on the DNA yield (3, 4 and 5). The same is not true for serum, since the initial amount of DNA in serum was ~2 fold higher than in plasma and increased 3-5 fold after blood was stored for 2-8 hours at room temperature (4).

Influence of the tubes or anticoagulants

Swinkels (2) found that the size of the blood collection tube and thus the amount of blood collected made a difference in the amount of cellular DNA present in a plasma sample. Whether the plasma sample was frozen or not, cellular DNA could still be removed by centrifugation at 16000g before DNA isolation (2). The influence of the brand of blood collection tube used, Vacutainer or S-Monovette, was shown not to have a significant effect on the yield of DNA isolated from either serum or plasma samples (6).

The use of different anticoagulants like EDTA, heparin or citrate, did not have a significant impact on the amount of circulating DNA isolated from plasma if the blood was processed within 6 hours of collection (7). If, however, a delay (>24 hours) in processing of the blood is anticipated, EDTA is a significantly better anticoagulant compared to heparin or citrate. In this case, quantification results could not really be compared to results from undelayed blood processing since a ~2 fold increase in concentration was observed in the former (7). It was also found that heparin, which some investigators suggested inhibits PCR, had no effect on real-time PCR (7).

The influence of drawing blood with a syringe and needle was also tested and no significant difference was observed in the free DNA yield between plasma samples aspirated a number of times, with a syringe and needle, or just once during blood collection (5).

Plasma/serum

The DNA concentration in serum can be 2-24 fold higher than that in plasma (4, 6 and 8) which may suggest the in vitro release of DNA from leukocytes into the serum (3) or hematopoietic cells during the clotting process (5). Furthermore, the fragment length of the circulating DNA was found to be significantly bigger in serum as compared to plasma. This may be attributed to the release of high-molecular-weight DNA during the clotting process (9). With these two important factors in mind, it is advisable to use plasma instead of serum for circulating DNA studies since serum and plasma are not equivalent sources of circulating DNA in

terms of their prognostic potential (6), thus, because of these crucial discrepancies, results using these two sources can not really be compared.

Centrifugation

The possibility that DNA yields could be artificially increased by the processing of blood samples was suggested by Herrera et al (10), when they showed that a two-step centrifugation is necessary in order to rid plasma of cellular DNA. The huge difference in DNA yield, 819 µg/L vs 13 µg/L respectively, between banked samples and fresh samples of esophageal cancer patients in their study was attributed to the fact that the banked samples were centrifuged only once (5 min at 1600g) while the fresh samples were centrifuged 10 min at 1600g and the resulting supernatant was centrifuged again for 10 min at the same speed after which the supernatant was stored for further use (10). Chiu et al (11) observed a significant difference in total DNA concentration when different blood processing protocols were followed. They also found that centrifugation followed by microcentrifugation or filtration is needed to remove all cellular DNA from the plasma. While Lui et al (5) did not get the same result when they compared circulating DNA concentrations in plasma samples centrifuged once or twice and found no significant difference. They also tested the effect of multiple centrifugations and still found no significant effect on the free DNA yield. Interestingly they used the same centrifugation speeds as Chiu et al (11). When plasma was processed by Percol separation a significant increase in the plasma circulating DNA concentration was found and it was concluded that Percol separation enriches plasma with cellular DNA (11). However, this fluctuation in DNA concentration because of centrifugation or filtration was not observed in the quantification of fetal DNA in maternal plasma (11).

Pipetting accuracy may also play a role in the prevention of contamination of free DNA preparations with cellular DNA (2), in fact this is a factor that can influence any quantitative analysis. Although this can be partially overcome by automatic pipetting stations, or experienced handling, it is always necessary to take great care in this regard.

Another aspect that should be kept in mind when quantifying circulating DNA is that natural day to day fluctuations in the total DNA concentration occurs in all individuals (11). The reason for this phenomenon is not known but medication and smoking is known to have an effect (12) and it is even possible that lifestyle may play a role in this fluctuation.

To be sure that plasma is free of cellular DNA, many preanalytical factors should be considered. In depth studies to compare the preanalytical conditions that have an influence on the concentration of circulating DNA in serum or plasma is scarce and need more attention.

Extraction methods

Through the years many methods for the isolation of circulating DNA has been used by different research groups, these include salting-out, phenol chloroform and other forms of organic extraction methods, Guanidine/Promega Wizard resin method and even automated extraction with the MagNA Pure (Roche) and Kingfisher (ThermoLifeSciences) robotic systems. However the extraction method that is made use of most often is the QIAamp Blood Mini Kit from Qiagen and according to Wu et al. (13) this kit can recover 82 to 92% of circulating DNA from serum with a sensitivity of 2 ng/ml. Sang et al. (14) agrees with this and found a relative standard deviation of 14.37% when measuring DNA yield using capillary zone electrophoresis with laser-induced fluorescence detection. Su et al. (15) disagree with the use of this kit because it theoretically purifies primarily high molecular-weight DNA. They examined this visually on polyacrylamide gel electrophoresis and chose to use the Guanidine/Promega Wizard resin method instead. Schmidt et al. (16) agree that the QIAamp columns are effective in binding larger molecules (100-150 bp) and they compared the columns to a modified salting-out method. They found that the yield of the QIAamp columns ranged from 40 – 60% and when circulating DNA is repurified with the QIAamp columns approximately 50% of the DNA stayed bound to the columns. According to the quantitative data they obtained, they concluded that the time-consuming salting-out method compensated with a higher yield of DNA, the low circulating DNA values they obtained was possibly a result of the real-time PCR target they used (see table).

De Kok et al. (17) compared three DNA isolation methods and they also calculated how PCR efficiency was influenced by different isolation methods, efficiencies of 70%, 34% and 51% was observed respectively for the PureGene kit , PureGene + QIAquick column and phenol-chloroform method. Furthermore it was observed that PCR was enhanced after isolation with the PureGene (10%) and phenol-chloroform (52%) methods (17).

The growing number of DNA isolation kits available on the market makes the choice between them increasingly difficult (17). With exception to pre-analytical factors that plays a significant role in circulating DNA concentration, a well-considered choice of an isolation method is very important in qualitative as well as quantitative analysis.

A table was constructed to list and compare isolation and quantification methods used and results obtained by different research groups and laboratories, alterations to the specified methods and also some other factors which may influence the results is listed (see table 1). Subsequently graphs to graphically compare the levels of circulating DNA in plasma (Figure 1) and serum (Figure 2) respectively, was constructed which clearly show the variability between methods, conditions and laboratories.

Table 1: Comparison of different isolation and quantification methods with subsequent results

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
61 breast cancer 33 non-neoplastic breast diseases (before treatment)	27 woman	5 ml EDTA tubes, centrifuged 10 min at 2000g, supernatant centrifuged 10min at 12000g	200 µl plasma	Within 2 hours, -80°C	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, Elution volume 50 µl		Real-time PCR B-*globin* (Taqman)	Healthy: 12.9 (3-73) Benign: 22.3 (5-106) Malignant: 64.6 (9-566)	(18)
19 lung cancer 4 colon cancer (patients was under treatment)	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 silicate magnetic beads and a KingFisher ML robotic magnetic particle processor (ThermoLifeSciences) And QIAamp DNA Midi kit (Qiagen) according to blood and body fluid protocol	Modifications to KingFisher kit: 300 µl plasma was mixed with 950 µl lysis buffer and 80 µl magnetic beads, DNA was eluted in Tris-EDTA for 20 min.	Flourometry (Fluoroskan; ThermoLifeSciences) whith Picogreen reagent (Molecular Probes)	Healthy: (3-22) Cancer: Qiagen: 15.7* (0-76) Concentrated plasma: KingFisher: 45.7* (13-127) Qiagen: 34.3* (9-128) No significant difference in yield was observed if DNA was isolated from concentrated plasma with either the KingFisher method or the Qiagen kit	(19)
84 non-small cell lung cancer (before surgery)	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	DNA eluted in bidistilled water	DipStick TM Kit (Invitrogen)	Healthy: 18.0 Patients: 318.0 Measurable amounts of DNA could not be found in 3 of the patents and 11 of the controls	(20)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
	10	EDTA, heparin or citrate	200 µl plasma	0,2,6 and 24 hours after collection (kept at room temp during this time), -80°C	QIAamp blood kit (Qiagen) according to blood and body fluid protocol, Elution volume 50 µl		Real-time PCR B-*globin* Taqman	EDTA: 0h: 3.9† 24h: 6.2† Heparin: 0h: 3.6† 24h: 27.1† Citrate: 0h: 3.6† 24h: 28.4†	(7)
173 non-small-cell-lung cancer	46	10 ml EDTA, two consecutive centrifugations of 1500g for 10 min 10 ml plain tubes for serum	1000 µl Plasma and 1000 µl serum	-70°C	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, Elution volume 70 µl	Washes with buffer AW2 was done 3 times to remove PCR inhibitors, DNA was eluted in Tris-HCL	Real-time PCR Glyceraldehyde-3-phosphate dehydrogenase gene Taqman	Healthy: Plasma: 1.8 Serum: 12.6 Patients: Plasma: 3.7 (0.1-93.5) Serum: 39.6 (0.3-641.3)	(6)
Banked samples: 20 esophageal cancer (EC) 23 gastroesophageal reflux disease (GERD) Fresh samples: 38 esophageal cancer 25 lung cancer (LC) 28 GERD	11	10 ml EDTA, Banked samples: centrifuged once for 5 min at 1600g, Fresh samples: Centrifuged twice for 10 min at 1600g	200 µl plasma	1 hour, -80°C	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol	Added 400 µl water and 400 µl buffer AL1 to sample, to increase DNA yield	Real-time PCR β-*actin* Taqman	Banked samples: EC: 819 (46.2-4738) GERD: 432 (6.0-2888) Fresh samples: Healthy: 10.6 (7-14) EC: 13 (4.5-46.5) LC: 14.6 (3-30) GERD: 10.5 (4-23.5)	(10)
31 with heart, liver or renal transplants	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 β-*globin* Taqman	Healthy: 6.0† Patients: 8.8†	(21)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
58 with chest pain suggestive of cardiac ischemia	21	4 ml EDTA, divided in 2 portions and centrifuged once for 10 min at 1500g, plasma from 1 portion was filtered (0.22 µm pore) and used for quantification	400-800 µl plasma	-80°C	QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol		Real-time PCR β-*actin* Taqman	Healthy: 2.3† Minor cardiac injury: 3.1† ST elevation angina (STEA): 3.5† ST elevation myocardial infarction (STEMI): 4.4† Died: 7.7†	(22)
84 with acute blunt traumatic injury	27	4 ml heparin, centrifuged at 3000g, time not available	400-800 µl plasma	-80 or -20°C	QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol		Real-time PCR β-*actin* Taqman	Healthy: 20.8† Minor/moderate trauma: 91.2† Major trauma: 1196.6†	(23)
	12 38	Plain tubes, centrifuged 10 min at 3000g	200 µl Serum	12 samples : 2 hours 38 samples: 8 hours, -20°C	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 100µl sterile water	DNA was either extracted (CZE^{1}) or digested wit proteinase K (CZE^{2}) for capillary zone electrophoresis	Capillary zone electrophores is (CZE) with laser-induced fluorescence detection And Real-time PCR B-*globin* Taqman	2h PCR: 67.7 (29-122) 2h CZE^{1}: 70.8 (27-149) 2h CZE^{2}: 71.1 (33-140) 8h PCR: 188.6 (24-362) 8h CZE^{1}: 192.6 (48-317) 8h CZE^{2}: 195.7 (50-335) The tree methods compared very well although CZE^{2} give slightly higher results	(14)
10 Hemodialysis (HD) patients	30	EDTA, centrifugation time and speed not available	800 µl plasma	-20°C	DNA isolation reagent set (Roche) according to protocol, elution volume 50 µl		Real-time PCR Vista Green	Healthy: 12.5±4.7 Before HD: 13.9±5.2 After 20 min of HD: 20.5±10.3 End of HD: 129.4±83.4	(24)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
	25 half-marathon runners	EDTA, centrifugation time and speed not available	800 µl plasma	-20°C	DNA isolation reagent set (Roche) according to protocol, elution volume 50 µl		Real-time PCR Vista Green	Before race: 18.0±2.8 (max 27) After race: 334.4±139.4 (max 702) 2h after race: 30.4±18.99 (max 112)	(25)
30 with cancer	20	EDTA, centrifugation for 20 min at 3000g	Plasma		QIAamp DNA blood kit (Qiagen) according to blood and body fluid protocol		Competitive PCR using lamin B2 locus as single copy gene and densitometric scanning of ethidium bromide-stained gel after PAGE And Real-time PCR β-*globin* LC-Red 640	Healthy: 3.7 (10-15) Cancer: 219.0 (10-1200) A difference of ~13% between the two methods was observed.	(26)
	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)	Type of tubes not available, centrifugation for10 min at 800g followed by 10 min at 16000g	400 µl plasma		QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 50 µl sterile water		Real-time PCR Β-*globin* SYBR Green I	Baseline: 0.2 (0.1-0.4)[†] T1: 0.9 (0.5-1.2)[†] T2: 1.9 (1.3-2.5)[†] T3: 4.0 (1.7-6.3)[†] T4: 0.5 (0.2-0.9)[†]	(27)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
55 myocardial infarction (MI) patients	274		200µl serum		QIAamp 96 Spin blood DNA extraction kit (Qiagen) according to the blood and body fluid protocol, eluted in 200 µl sterile water	First digested with proteinase K, three consecutive washes was performed during extraction	PicoGreen DNA kit (Molecular probe) and fluorescence measured by spectro-fluorometer	Healthy: 36.3 ± 23.8 MI: 510.6 ± 398	(28)
88 with a stroke-like syndrome	24	10 ml EDTA, centrifuged once for 5 min at 1500g	200 µl plasma	-80°C	QIAamp blood kit (Qiagen) according to the blood and body fluid protocol, eluted in 200 µl sterile water		Real-time PCR B-*globin* Taqman	Healthy: no value available Patients: 9.5* (1.6 –54.6)[†]	(29)
186 with breast cancer 161 with ovarian cancer 74 with prostate cancer	100		1000 µl serum	-70°C for ~1 year	QIAamp 96 Spin blood DNA extraction kit (Qiagen) according to the blood and body fluid protocol, eluted in 200 µl sterile water or Tris-EDTA	Proteins was digested with proteinase K before DNA extraction and three consecutive washes was performed during isolation	PicoGreen Double strand DNA detection kit (Molecular probe) and flouresence measured by spectro-fluorometer	Healthy: 57.0±30 Breast cancer:** CA15-3<100: 74.0±49 CA15-3 100-500: 111.0±104 CA15-3: 307.0±601 Ovarian cancer:** CA125<100U/ml: 67.0±36 CA125>500U/ml: 49.0±27 CA125<500: 118.0±125 Prostate cancer:** PSA <4ng/ml: 458.0±790 PSA >20 ng/ml: 490.0±471	(13)
52 patients in a general intensive care unit (ICU)	10	EDTA, initially centrifuged 6 min at 3000 rpm followed by 10 min at 14000 rpm	200 µl plasma	Within 3 hours, -20°C	High Pure PCR template preparation kit (Roche)		Real-time PCR B-*globin* Taqman	Healthy: 17.0*** (14-19) Patients: 80.0*** (48-260)	(30)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
9 with lung cancer 7 with benign lung disease		EDTA, initially centrifuged 10 min at 350g followed by 15 min at 2000g	1000 µl plasma	-80°C	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol And Salting-out method		Real-time PCR ERV-3 Taqman	Lung cancer: Qiagen: 4.0 Salting-out: 7.8 Benign: Qiagen: 3.0 Salting-out: 15.8	(31)
	5	10 ml EDTA,	Plasma	-80°C	a: PureGene DNA isolation kit (Genta Systems), b: PureGene DNA isolation kit + QIAguick column method c: phenol-chloroform method	DNA was dissolved in or eluted with 50 ml Tris-HCL in all methods, Plasma was pooled and spiked with 50 copies of marker DNA before (samples) or after (controls) DNA isolation. For further alterations see (kok 1998)	Real-time PCR *Bcr-abl* marker Taqman	Control***: a: 51 copies b: 26 copies c: 70 copies Sample***: a: 33 copies b: 16 copies c: 24 copies In a pilot study only a few copies of the marker could be measured in the isolate of the Qiagen kit.	(17)
32 with colorectal cancer (CRC) 19 with periampullary cancer (PACs)	51	1 ml CORVAC serum separator tubes, centrifugation for 15 min at 1000g and filtered (3 mm serum filter, Fisher Scientific)	Serum	Within 6 hours, -80°C	No isolation, only proteinase K digestion and further centrifugation for 5 min at 10000g		Real-time PCR ALU SYBR Gold (Molecular Probe)	Healthy: 0.34 ±0.25 CRC: Stage I/II: 1.63±0.43 Stage III/IIII: 1.73±0.45 PACs: Stage I/II: 0.84±0.53 Stage III/IIII: 0.66±0.62	(32)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
22 bone marrow transplantation recipients	14	EDTA or plain, Centrifuged first at 3000g for 10 min followed by 10 min at 16000g	400-800 µl Plasma and serum (from only 7 patients)	Within 6 hours	QIAamp blood kit (Qiagen) according to blood and body fluid protocol		Real-time PCR B-*globin* Taqman	Healthy: 8.3†*** Patient: Plasma: 7.7†*** Serum: 107.9†***	(5)
	34 pregnant woman (getational age 13-38 weeks)	10 ml EDTA, divided in 2 parts: 1. initially centrifuged 10 min at 1600g followed by either: 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 1200g on a discontinuous percol (Amersham Pharmacia) gradient followed by either: a, b or c (like described above)	800 µl plasma	Within 2 hours, -20°C	QIAamp blood kit (Qiagen) according to blood and body fluid protocol		Real-time PCR B-*globin* Taqman	1a: 15.9***(4-71)†‡ 1b: 8.7*** (2-34)†‡ 1c: 8.7*** (2-59)†‡ 1d: 9.5*** (2-34)†‡ 2a: 117.8*** (25-353)†‡ 2b: 8.7*** (2-38)†‡ 2c: 17.4*** (4-163)†‡	(11)
	25 woman in early pregnancy (11-7 wk gestation) 25 woman in late pregnancy (37-43 wk gestation)	10 ml EDTA or plain, centrifuged at 3000g and supernatant resentrifuged at 3000g, times not available	400-800 µl plasma or serum	-20°C	QIAamp blood kit (Qiagen) according to blood and body fluid protocol, elution volume 50 µl		Real-time PCR B-*globin* Taqman	Plasma: Early: 6.5 (2.4-12.3) Late: 39.2 (7.4-210.4) Combined: 22.9 (2.4-210.4) Serum: Combined: 334.3 (38.4-1608.8)	(8)

Number of Patients	Number of healthy individuals	Tubes and centrifugation	Plasma/serum and starting volume for isolation	Time before plasma collection and temp. stored at before isolation	Isolation method	Alterations before and during kit protocol	Quantification method	Quantity of circulating DNA Average (range) ng/ml	Reference
100 with NSCLC	100 matched by sex, age and smoking habits	7.5 ml EDTA, two times centrifugation for 10 min at 2500 rpm	1 ml plasma	Blood stored at -140°C before plasma collection	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 50 ml water		Real time PCR *hTERT* Taqman	Healthy: 4.6 (0.1-27) Patients: 75.0 (0.5-3010)	(12)
10 breast cancer patients under treatment	10	4 ml EDTA, (a) Centrifuged once for 20 min at 1600g or (b) Centrifuged first at 3000g for 10 min followed by 10 min at 16000g (for half of plasma from normal patients only)	200 µl plasma	Immediately, -20°C	QIAamp DNA mini blood kit (Qiagen) according to blood and body fluid protocol, eluted in 50 ml water	Added 400 µl water and 400 µl buffer AL1 to sample, to increase DNA yield	Real-time PCR B-*globin* Taqman	Healthy: (a) 1.5 (0.3-7.1) (b) 0.5 (0.0-1.6) Patients: 10.8 (0.3-43.7)	****

* Average calculated from supplementary data

** CA15-3, CA125 and PSA are tumor markers for breast, ovarian and prostate cancers respectively, normal PSA: 4-20 ng/ml

*** Median levels (average not available)

**** Results from our study

† ng/ml calculated from genome equivalents per milliliter or liter (rounded to 1 decimal)

‡ Values estimated from box plots in article

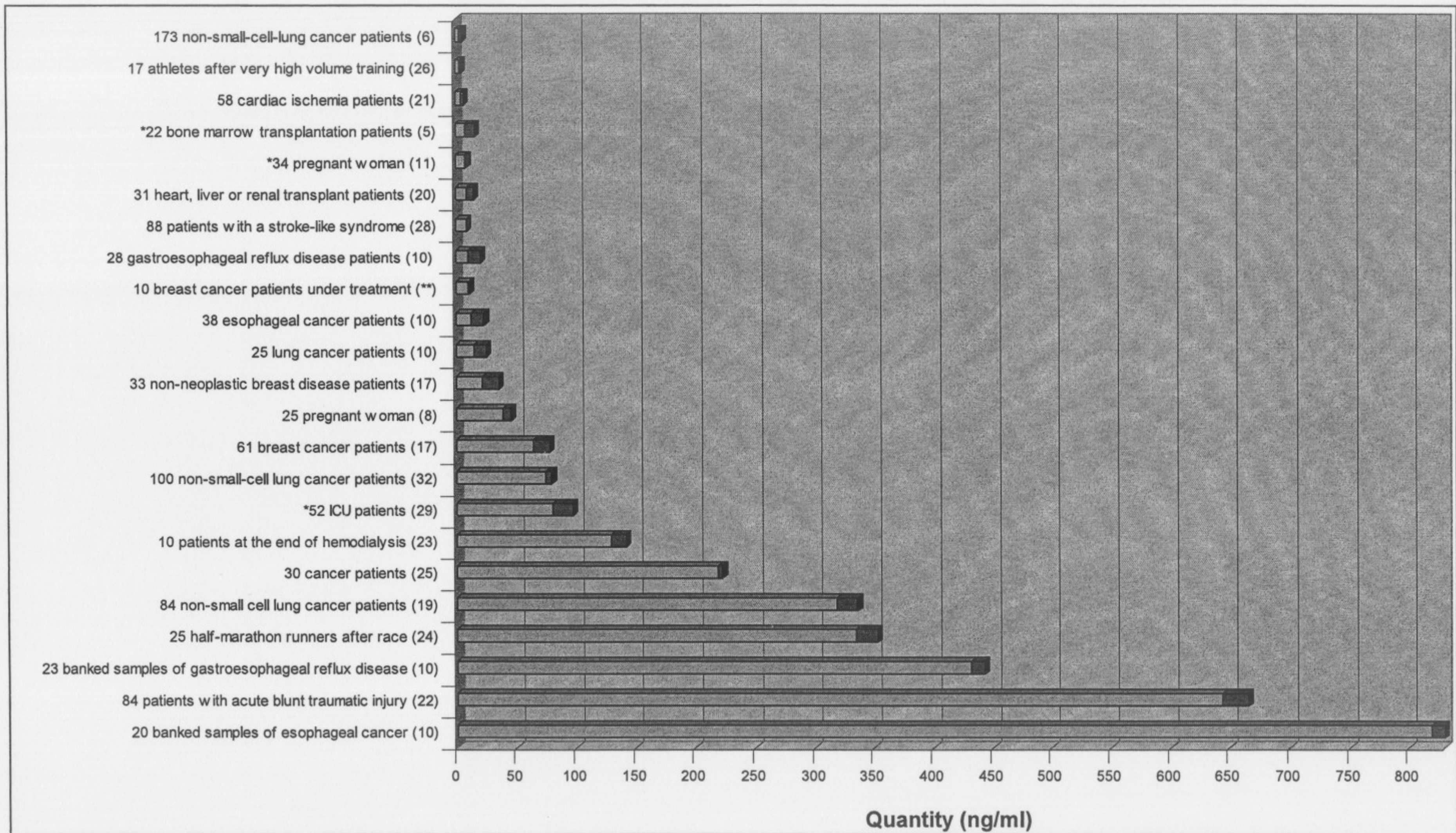

Fig 1: Levels of circulating DNA in plasma in different conditions

Levels in controls are shown in ■ on top of levels of conditions shown on the y-axis. Exclusion criteria: serum samples, samples with no mean or medial levels available, cancer samples without control values. Data is arranged in ascending order for ease in comparison.

* Median levels (average not available)

** Results from our study

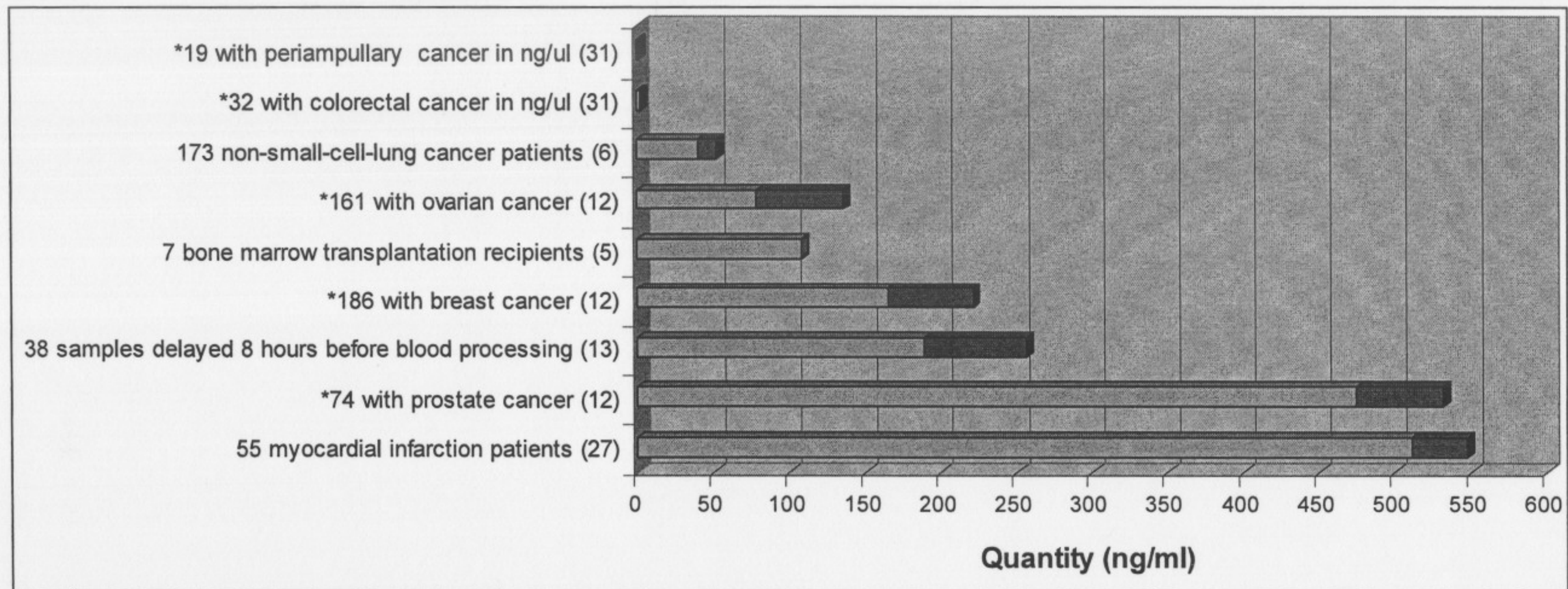

Fig 2: Levels of circulating DNA in serum in different conditions

Levels in controls are shown in ■ on top of levels of conditions shown on the y-axis. Data is arranged in ascending order for ease in comparison.
* Average calculated from different cancer stages.

Results presented in Table 1 and the graphs (Figure 1 and 2) show that all the values for control subjects in both plasma and serum analysis is less than 21 ng/ml (range 0.0-20.8), but for cancer patients the values ranged between 3.7 and 819 ng/ml. This shows a clear rise in circulating DNA with malignancy. It also seems as though levels obtained from plasma circulating DNA is higher than those in serum, but this in not the case as has been shown by studies using the same method for both serum and plasma DNA isolation and quantification (6). The reason for this shift in values is the impact that preanalytical factors and the choice of isolation and quantification methods have on the results.

Quantification methods

Real-time-PCR

The enormous power of real-time PCR that can detect the DNA from a single cell which corresponds to about 6.6 pg, count in favor of quantification of circulating free DNA (18, 29). Furthermore, no post-PCR processing of samples is necessary which minimizes the risk of carryover contamination and the fast turnaround time is another advantage (18). An important aspect of real-time PCR, like PCR, is that the quality, purity and design of the primers and probe are determining factors in the success of its application. Lo et al. (8) gave an overview of the Taqman analysis, this system use a probe which is specific for the analyzed sequence and fluorescence from the selected dye is only observed if the Taq DNA polymerase cleaves the reporter part from the probe in the extension step of the PCR reaction. The principle on which the Taqman system is based, make it the method of choice above application in which SYBR Green is used to quantify circulating DNA, since SYBR Green bind to any double stranded DNA and is not specific to the sequence that is analyzed.

PicoGreen

The wide detection range, 25 pg – 1000 ng/ml, of the PicoGreen Double strand DNA quantification kit was chosen by Wu et al. (13), and they also showed that only half of the PicoGreen reagent can be used with half the DNA to give the same result, thus saving reagent and requiring less sample.

DNA Dipstick®

The DNA Dipstick® assay (Invitrogen) can also be used for DNA quantification but, although it can detect low concentrations of DNA (0.1 ng/µl), it detects both double and single stranded DNA, as well as RNA. As in all flourometric based methods the DNA Dipstick® assay depends on a colour change and not direct quantification. Frattini et al (33) reported that this method gave reliable and rapid quantification of circulating DNA in plasma.

Radioimmunoassay

Leon *et al.* (34) developed a direct competitive binding radioimmunoassay method to detect DNA in biological material in the mid seventies. Leon *et al.* and Shapiro *et al.* (35) used this method and, respectively, detected quantities of up to 5000 ng/ml and 10000 ng/ml circulating DNA directly in serum with the use of antibody from a systemic lupus erythematosus (SLE) patient. Compared to the numbers generally published, these quantities seem to be abnormally high.

Capillary zone electrophoresis

Capillary zone electrophoresis with laser-induced fluorescence detection is characterized by simplicity, automation, short analysis time, small sample and reagent requirement and extremely high sensitivity and good reproducibility can be obtained with a good covalent-coated capillary (14). A big advantage of this method (CZE[2] with serum digestion by proteinase K) is that DNA extraction from serum, and possibly plasma, is not needed which saves time and money. Data obtained by Sang (14) proved that capillary zone electrophoresis with

laser-induced fluorescence detection is a practical and reliable method for quantification of circulating DNA in serum with a detection limit of 0.5 pg/µl.

Competitive PCR

Absolute quantification of nucleic acids can be achieved by a PCR based method described by Diviacco (36). With the addition of a known amount of competitor DNA molecules that share the same primer recognition sites as the sample, but are of different length, the concentration of sample DNA can be determined by the ratio of amplified products after gel electrophoresis. The advantage of this method is that the amplification of both the sample and the competitor is affected by variables in the PCR. Jahr (26) compared quantification of this competitive PCR method using the lamin B2 locus as template, with that of real-time PCR for β-globin using LC-Red 640 as flourophore, and reproducibility of ~12.5% and ~10% respectively and a difference of ~ 13% between the two methods was observed.

The different methods used and variables in methods make it difficult to compare the free DNA yield reported by different research groups (18), but despite considerable variation, all of these methods demonstrated the elevated levels of circulating DNA in cancer patients compared to healthy individuals (14). Due to the increased sensitivity of PCR-based quantification compared to other flourometric methods (14), real-time PCR is more promising for quantitative analysis of circulating DNA.

Methods to evaluate qualitative changes in DNA

Detection of genetic and epigenetic changes like microsatellite alterations, gene hypermethylation and mutations in Ras and P53 in circulating DNA, as well as other circulating entities like RNA transcripts and mitochondrial and viral DNA, are reviewed elsewhere (37, 38, 39 and 40). Other techniques which include mutant allele-specific PCR (MASA), restriction fragment length polymorphism (RFLP) PCR and methylation specific PCR (MSP) are needed to detect these qualitative changes in the free DNA.

MS analysis

A method based on a single-allele base extension reaction (SABER) and mass spectrometry was developed by Ding et al. (41). According to them, this method can detect fetal specific alleles which include point mutations and single nucleotide polymorphisms (SNPs), if present in maternal plasma. They showed the analytical sensitivity and specificity of the MS analysis by detecting the paternal mutations coding for β-thalassemia in maternal blood and finding the same fetal genotype as determined by amniotic fluid, chorionic villus, or fetal blood analysis. The method was performed on maternal blood obtained from 62 women, 50 of which carried normal fetuses and 12 were at risk for β-thalassemia major. In this blind study they detected 6 out of 6 fetal genotypes with β-thalassemia. MS analysis can thus play an important role in the future research concerning circulating DNA, not just in pregnancies but also in cancer and other areas (41).

Characterization of total circulating DNA

To our knowledge only one research group published a paper in which they described amplification of total circulating DNA. Li et al. (42) used a blunt-end ligation-mediated whole genome amplification method to amplify small as well as intermediate and large DNA fragments obtained from plasma. In short, the isolated plasma circulating DNA was blunted with T4 DNA polymerase after which it was self or cross-ligated with T4 DNA ligase, followed by heat inactivation of the enzyme. This was followed by amplification using the GenomiPhi whole genome amplification kit (Amersham Biosciences). The entire protocol takes more than 16 hours to be completed and can be carried out in a single tube. With this method screening for allelic imbalance and other genetic alterations including SNPs are made possible, furthermore, with the use of microarrays, total circulating plasma DNA can be examined for cancer fingerprints or new biomarkers (42).

Some research groups have sequenced the PCR products they generated (43) but these sequences represented only a specific small fraction of the circulating DNA. Although mutations can be detected in the amplified regions of the circulating DNA, characterization or analysis of the circulating DNA as a whole have not been performed and as a result conclusions about the origin or function can not be reached.

Conclusions

Preanalytical sample preparation is an important factor in obtaining uncontaminated starting material, circulating DNA, for subsequent quantitative or qualitative analysis. The use of plasma rather than serum as source of circulating DNA is encouraged for the purpose of comparing results. Furthermore it is concluded that plasma should be separated from blood sooner rather than later, but always within 6 hours of blood collection. No significant effect is observed in the circulating DNA if plasma is freeze-thawed for less than three cycles, but aliquating of plasma before freezing at -20°C or -80°C is advisable. Different results from the effects of diverse centrifugation procedures has been observed, our results show a 66% decrease in circulating DNA between plasma centrifuged once or twice, thus it may be concluded that centrifugation of whole blood for 10 minutes at 1600g followed by microcentrifugation (10 minutes at 16000g) of the resulting supernatant may be optimal for ridding plasma of most of the cellular particles including cellular DNA. This second centrifugation step may even be carried out after plasma was frozen. The daily fluctuation of the levels of circulating DNA in all individuals may also have an influence on levels of circulating DNA and must be compensated for by choosing variably matched control subjects for each study. Further work regarding this fluctuation and its cause is considered necessary.

A well considered choice of isolation and quantification method is important as it influence comparability of results. This is clearly demonstrated by the widely varying levels of circulating DNA between different studies shown in this review.

All values for plasma of control subjects was less than 21 ng/ml (range 0.0-20.8), but for cancer patients the values ranged between 3.7 and 819 ng/ml (see Table 1). Although it is generally accepted that cancer patients have higher levels of circulating DNA than healthy subjects, whether plasma or serum is used for analysis, the huge concentration difference observed between the different studies is most likely a result of the preanalytical factors and the methods used for isolation and quantification. A few things in this regard is worth mentioning: (a) the studies which reported the highest levels of circulating DNA made use of a single centrifugation step for obtaining plasma, (b) those reporting the lowest levels used alternative targets or detection systems for real-time PCR and/or different isolation methods, this include plasma and serum analysis and (c) even though it may look as if plasma concentrations is higher than those in serum (see fig 1 and 2), this is not the case. This just prove that the methods used to isolate and quantify circulating DNA and especially the preanalytical factors affecting the amount of cellular DNA in plasma or serum can cause quantification results to be misleading and worthless.

References

1. Stroun M, Anker P, Lyautey J, Lederrey C, Maurice PA. Isolation and characterization of DNA from the plasma of cancer patients. Eur J Cancer Clin Oncol 1987;23:707-12.
2. Swinkels DW, Wiegerinck E, Steegers EA, de Kok JB. Effects of blood-processing protocols on cell-free DNA quantification in plasma. Clin Chem 2003;49:525-6.
3. Chan KC, Yeung SW, Lui WB, Rainer TH, Lo YM. Effects of preanalytical factors on the molecular size of cell-free DNA in blood. Clin Chem 2005;51:781-4.
4. Jung M, Klotzek S, Lewandowski M, Fleischhacker M, Jung K. Changes in concentration of DNA in serum and plasma during storage of blood samples. Clin Chem 2003;49:1028-9.
5. Lui YY, Chik KW, Chiu RW, Ho CY, Lam CW, Lo YM. Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. Clin Chem 2002;48:421-7.
6. Gautschi O, Bigosch C, Huegli B, Jermann M, Marx A, Chasse E, et al. Circulating deoxyribonucleic Acid as prognostic marker in non-small-cell lung cancer patients undergoing chemotherapy. J Clin Oncol 2004;22:4157-64.
7. Lam NY, Rainer TH, Chiu RW, Lo YM. EDTA is a better anticoagulant than heparin or citrate for delayed blood processing for plasma DNA analysis. Clin Chem 2004;50:256-7.
8. Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768-75.
9. Chen Z, Fadiel A, Naftolin F, Eichenbaum KD, Xia Y. Circulation DNA: Biological implications for cancer metastasis and immunology. Medical Hypotheses 2005;65:956-61.
10. Herrera LJ, Raja S, Gooding WE, El-Hefnawy T, Kelly L, Luketich JD, Godfrey TE. Quantitative analysis of circulating plasma DNA as a tumor marker in thoracic malignancies. Clin Chem 2005;51:113-8.
11. Chiu RW, Poon LL, Lau TK, Leung TN, Wong EM, Lo YM. Effects of blood-processing protocols on fetal and total DNA quantification in maternal plasma. Clin Chem 2001;47:1607-13.
12. Sozzi G, Conte D, Leon M, Ciricione R, Roz L, Ratcliffe C, et al. Quantification of free circulating DNA as a diagnostic marker in lung cancer. J Clin Oncol 2003;21:3902-8.
13. Wu TL, Zhang D, Chia JH, Tsao KH, Sun CF, Wu JT. Cell-free DNA: measurement in various carcinomas and establishment of normal reference range. Clin Chim Acta 2002;321:77-87.
14. Sang F, Ren J. Comparisons between capillary zone electrophoresis and real-time PCR for quantification of circulating DNA levels in human sera. Journal of Chromatography B 2006;838:122-8.

15. Su YH, Wang M, Block TM, Landt O, Botezatu I, Serdyuk O, et al. Transrenal DNA as a diagnostic tool: important technical notes. Ann N Y Acad Sci 2004;1022:81-9.
16. Schmidt B, Weickmann S, Witt C, Fleischhacker M. Improved method for isolating cell-free DNA. Clin Chem 2005;51:1561-3.
17. de Kok JB, Hendriks JC, van Solinge WW, Willems HL, Mensink EJ, Swinkels DW. Use of real-time quantitative PCR to compare DNA isolation methods. Clin Chem 1998;44:2201-4.
18. Huang ZH, Li LH, Hua D. Quantitative analysis of plasma circulating DNA at diagnosis and during follow-up of breast cancer patients. Cancer Letters 2006;243:64-70.
19. Stemmer C, Beau-Faller M, Pencreac'h E, Guerin E, Schneider A, Jaqmin D, et al. Use of magnetic beads for plasma cell-free DNA extraction: toward automation of plasma DNA analysis for molecular diagnostics. Clin Chem 2003;49:1953-5.
20. Sozzi G, Conte D, Mariani L, Lo Vullo S, Roz L, Lombardo C, et al. Analysis of circulating tumor DNA in plasma at diagnosis and during follow-up of lung cancer patients. Cancer Res 2001;61:4675-8.
21. Lui YY, Woo KS, Wang AY, Yeung CK, Li PK, Chau E, et al. Origin of plasma cell-free DNA after solid organ transplantation. Clin Chem 2003;49:495-6.
22. Rainer TH, Lam NYL, Man CY, Chiu RWK, Woo KS, Lo YMD. Plasma [beta]-globin DNA as a prognostic marker in chest pain patients. Clinica Chimica Acta 2006;368:110-3.
23. Lo YM, Rainer TH, Chan LY, Hjelm NM, Cocks RA. Plasma DNA as a prognostic marker in trauma patients. Clin Chem 2000;46:319-23.
24. Atamaniuk J, Ruzicka K, Stuhlmeier KM, Karimi A, Eigner M, Mueller MM. Cell-free plasma DNA: a marker for apoptosis during hemodialysis. Clin Chem 2006;52:523-6.
25. Atamaniuk J, Vidotto C, Tschan H, Bachl N, Stuhlmeier KM, Muller MM. Increased concentrations of cell-free plasma DNA after exhaustive exercise. Clin Chem 2004;50:1668-70.
26. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, Knippers R. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659-65.
27. Fatouros IG, Destouni A, Margonis K, Jamurtas AZ, Vrettou C, Kouretas D, et al. Cell-free plasma DNA as a novel marker of aseptic inflammation severity related to exercise overtraining. Clin Chem 2006;52:1820-4.
28. Chang CP, Chia RH, Wu TL, Tsao KC, Sun CF, Wu JT. Elevated cell-free serum DNA detected in patients with myocardial infarction. Clin Chim Acta 2003;327:95-101.
29. Rainer TH, Wong LK, Lam W, Yuen E, Lam NY, Metreweli C, Lo YM. Prognostic use of circulating plasma nucleic acid concentrations in patients with acute stroke. Clin Chem 2003;49:562-9.

30. Rhodes A, Wort SJ, Thomas H, Collinson P, Bennett ED. Plasma DNA concentration as a predictor of mortality and sepsis in critically ill patients. Crit Care 2006;10:R60.
31. Schmidt B, Engel E, Carstensen T, Weickmann S, John M, Witt C, Fleischhacker M. Quantification of free RNA in serum and bronchial lavage: a new diagnostic tool in lung cancer detection? Lung Cancer 2005;48:145-7.
32. Umetani N, Kim J, Hiramatsu S, Reber HA, Hines OJ, Bilchik AJ, Hoon DS. Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. Clin Chem 2006;52:1062-9.
33. Frattini M, Balestra D, Verderio P, Gallino G, Leo E, Sozzi G, et al. Reproducibility of a semiquantitative measurement of circulating DNA in plasma from neoplastic patients. J Clin Oncol 2005;23:3163-4; author reply 4-5.
34. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 1977;37:646-50.
35. Shapiro B, Chakrabarty M, Cohn EM, Leon SA. Determination of circulating DNA levels in patients with benign or malignant gastrointestinal disease. Cancer 1983;51:2116-20.
36. Diviacco S, Norio P, Zentilin L, Menzo S, Clementi M, Biamonti G, et al. A novel procedure for quantitative polymerase chain reaction by coamplification of competitive templates. Gene 1992;122:313-20.
37. Ziegler A, Zangemeister-Wittke U, Stahel RA. Circulating DNA: a new diagnostic gold mine? Cancer Treat Rev 2002;28:255-71.
38. Bremnes RM, Sirera R, Camps C. Circulating tumour-derived DNA and RNA markers in blood: a tool for early detection, diagnostics, and follow-up? Lung Cancer 2005;49:1-12.
39. Goebel G, Zitt M, Zitt M, Muller HM. Circulating nucleic acids in plasma or serum (CNAPS) as prognostic and predictive markers in patients with solid neoplasias. Dis Markers 2005;21:105-20.
40. Tong Y-K, Lo YMD. Diagnostic developments involving cell-free (circulating) nucleic acids. Clinica Chimica Acta 2006;363:187-96.
41. Ding C, Chiu RW, Lau TK, Leung TN, Chan LC, Chan AY, et al. MS analysis of single-nucleotide differences in circulating nucleic acids: Application to noninvasive prenatal diagnosis. Proc Natl Acad Sci U S A 2004;101:10762-7.
42. Li J, Harris L, Mamon H, Kulke MH, Liu WH, Zhu P, Mike Makrigiorgos G. Whole genome amplification of plasma-circulating DNA enables expanded screening for allelic imbalance in plasma. J Mol Diagn 2006;8:22-30.
43. Diehl F, Li M, Dressman D, He Y, Shen D, Szabo S, et al. Detection and quantification of mutations in the plasma of patients with colorectal tumors. Proc Natl Acad Sci U S A 2005;102:16368-73.

Chapter 4

Results and discussion

4.1 *Manuscript prepared for submission*

A method for characterization of total circulating DNA

M. van der Vaart and P.J. Pretorius

This is the first draft of an article prepared for submission to Clinical Chemistry, International Journal of Molecular Diagnostics and Laboratory Medicine or another suitable journal. The instructions for authors to Clinical Chemistry are available at: http://www.clinchem.org/info_ar/info_a_outline.shtml

A method for characterization of total circulating DNA

Maniesh van der Vaart and Piet, J. Pretorius[1]

Abstract

Background: Although much work has been done in the field of circulating DNA, no publications on the sequencing 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 with various methods and cloned into a blunt vector. Resulting colonies was sequenced and analyzed.

Results: It was found that the majority of the DNA that ligated to the vector was about 200 bp in length. A total of thirty-five sequences were analyzed and it was found that circulating DNA consists of parts of the human genome that has not been characterized yet. A number of large repeats comprise large parts of the sequences and Alu repeats are in abundance.

Conclusion: Cloning and sequencing of free circulating DNA was successful and this first attempt on characterizing free circulating DNA by analysis of sequence data confirmed results of previous studies. Further characterization has to be done in order to determine the composition of total free circulating DNA.

Introduction

Circulating DNA, which is fragments of nucleic acids that circulate in the blood of healthy individuals, cancer patients and patients with a number of diseases as well as during pregnancy. It is known that the former group has lower levels of

---

School of Biochemistry, North-West University, Potchefstroom Campus, South Africa.
[1] Address correspondence to this author at: e-mail piet.pretorius@nwu.ac.za,
Fax +2718-299-2316, Tel +2718-299-2309

these circulating nucleic acids compared to the other groups (1-3). 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 little has been published on the analysis of the sequences that total circulating DNA comprises. PCR and Real-time PCR amplification showed the presence of a number of genes like *P53*, the *Ras* family, *Ig* Heavy chain, *TP53*, *SRY* and *DYS1* (for fetal DNA), *lamin* B2, *leptin*, β-*Globin*, β-*actin* and Alu repeats, and a number of other polymorphic markers were used to detect alterations which are similar in the tumor and plasma DNA (1-9). 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 unknown. Characterization of total circulating DNA may shed some light on this aspect, which may be beneficial in diagnosis and prognosis and may also contribute to determining the source and function of circulating DNA.

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 healthy individuals.

Methods

Blood collection and isolation of DNA

Blood was drawn in two 4ml EDTA tubes from each individual. Plasma was immediately separated from the cellular fraction by centrifugation at 1600g for 20 min after which the resulting supernatant (plasma) was stored at -20°C. Circulating DNA was isolated with a variety of methods like the QIAamp DNA Blood Mini kit (Qiagen), a salting-out method (Schmidt 2005), phenol/chloroform extraction and the MagNA Pure Compact system (Roche).

Cloning and sequencing

Blunting of DNA was performed using T4 DNA Polymerase (Fermentas) and 10 μl of DNA with Tango buffer (Fermentas). The reaction mixture was

incubated without dNTP's for 5 min at 11°C, dNTP's 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 then 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 heat inactivated and a chloroform extraction was performed. The blunt, phosphorylated DNA is then ligated into the pEZSeq vector (Lucigen) with subsequent heat inactivation of the enzyme after which colony PCR was performed to estimate the success of the ligation (Fig 1). SURE cells are then transformed with the vector by heat shock and cells are plated on YT +Amp agar plates after recovery. Colonies that was shown to have inserts (Fig 2), with colony PCR, was sequenced by Inqaba Biotechnical Industries (South Africa).

Sequence analysis

Sequences was analyzed with a number of web based sequence analysis tools like Repbase CENSOR (10), EMBOSS suite (11), MEME/MAST (Multiple Em for Motif Elicitation/Motif Alignment & Search Tool) (12, 13), Sequence Manipulation Suite (14) and NCBI BLAST against various databases (15).

Results and conclusions

Cloning of total circulating DNA was not as easy as first expected. A possible reason is that the ends of the DNA fragments are capped or that something is preventing the blunting enzyme from functioning. Blunting of the free circulating DNA fragments was made possible by altering the DNA blunting protocol provided with the enzyme.

As shown in Figure 1, a large amount of DNA ligated to the vector. The majority of the DNA fragments which bound to the vector was about 200 bp (length of vector DNA attached to the circulating DNA is 110 bp) and larger and smaller fragments can also be observed, this correlate with results from other groups (3, 4). This ligated DNA was transformed into competent cells which was

plated out and produced a number of colonies. Colony PCR of thirty-five of the colonies that grew (shown in Figure 2) also show different sizes of inserts which indicate that different fragments were cloned.

It was observed that different methods of DNA isolation had a significant impact on the amount of colonies that grew, circulating DNA isolated with the QIAamp blood mini kit produced the least amount of colonies while the phenol/chloroform and salting-out extraction method produced many colonies. A total of thirty-five colonies were sent for sequencing.

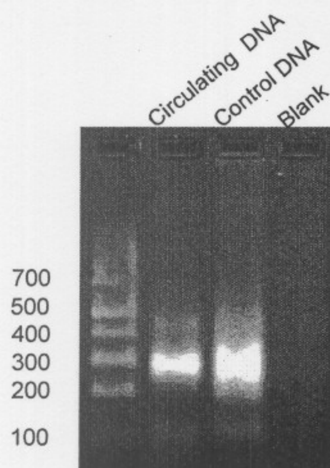

Fig 1: Products of PCR on circulating DNA ligated to the pEZSeq vector, before transformation into cells. The control DNA is genomic DNA which was fragmented.

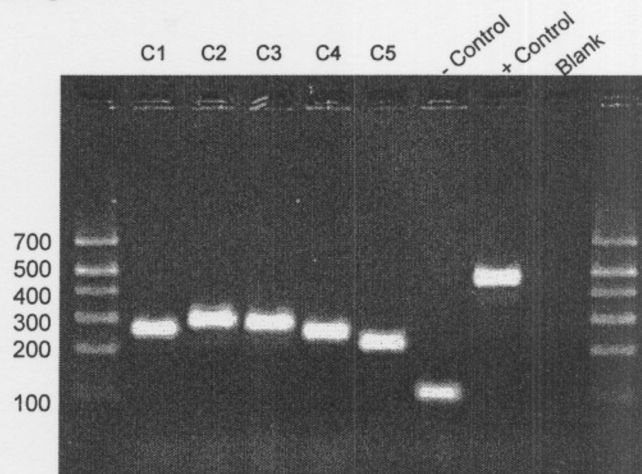

Fig 2: Products of colony PCR of pEZSeq transformants, thirty-five similar colonies were sequenced of which C1 to C5 is representative.

When sequence analysis was done, it was found that many of the sequences coordinated to parts of the human genome which has not been characterized yet, but a number of large repeats were found to comprise a large amount of the sequences. Table 1 was constructed to compare the sequences to each other in an attempt to make sense of the large amount of data which were obtained. When sequences were submitted for a nucleotide-nucleotide BLAST search against the Human genome build 36 database, it was found that the sequences represented almost every chromosome, but a match in this database could not

be found for all sequences. When the nucleotide-nucleotide BLAST was done against the Alu database, it was interesting to find many matches with high E-values in almost all classes matched with Alu repeats. Where Alu-J represents the oldest of the Alu families in the proposed Alu evolution, the Alu-S family represents Alu's of intermediate age while the Alu-Y family is the youngest (16). Furthermore, if the fact that Alu sequences, which is short interspersed elements (SINEs), only comprise ~13% of the genome while genes and related sequences comprise ~25% of the human genome is taken into account, one should expect to see more sequences representing genes (17, 18). The appearance of so many Alu repeats must be significant, but at this stage, it is not known what the significance and the meaning of this is. The reason why genes or other sequences were not cloned is also unknown.

When sequences was analyzed with CENSOR, which searches for known human repeats in the given sequences, other repeats with high scores were found, again the meaning of circulating DNA containing these repeats will have to be elucidated. Only abbreviations of the names of the repeats are given in the table because of space constraint, more information about these repeats are available in addendum E of this document.

MEME/MAST analysis revealed more repeats and also some similarity between the sequences, these results were not included because of inconsistency. Results obtained from the EMBOSS suite of programs did not produce any significant results. It is obvious that more work need to be done for a conclusion to be reached and the results obtained look promising and is encouraging for further analysis because this is, as far as we know, the first attempt on characterizing free circulating DNA by analysis of sequence data.

The table which was constructed was inserted in the article for the sake of completeness and exam purposes, and will be revised before submitting this manuscript for publication.

Table 1 of sequence analysis data

Sequence name (Length bp) Analysis name (Number of hits)	Length of match	Hit sequence name	Similarity	Score (Bits)	E-value*	Number of instances of same family or with alternate assemblies
>m6m13R(124)						
CENSOR	122	SVA	77%	423		
BLASTN Alu (133)		Alu-Sx,	90%	151	3.00E-39	44
		Alu-Sb1,	89%	135	2.00E-34	1
		Alu-Sb,	89%	135	2.00E-34	5
		Alu-Sc,	89%	127	4.00E-32	4
		Alu-Sq,	88%	121	2.00E-30	7
		Alu-J,	87%	119	1.00E-29	28
		Alu-Sp,	86%	105	1.00E-25	4
		Alu-Sb2,	82%	81.8	2.00E-18	1
		Alu-FLA?	83%	73.8	5.00E-16	4
>m5m13R(188)						
CENSOR	188	L1PBB_5	78%	1053		
BLASTN Alu (4)	13	Alu-J	100%	26.3	0.11	2
	11	Alu-S	100%	22.3	1.7	1
	11	Alu-Sx	100%	22.3	1.7	1
>m2m13R(161)						
CENSOR	126	SVA	75%	434		
BLASTN Alu (168)	153	Alu-Sx	86%	113	6.00E-28	44
	146	Alu-Sp	86%	107	4.00E-26	5
	146	Alu-Sq	85%	99.6	9.00E-24	8
	146	Alu-Sc	84%	91.7	2.00E-21	4
	153	Alu-Sb	83%	73.8	5.00E-16	7
	69	Alu-J	88%	65.9	1.00E-13	25
	44	Alu-Spqx	95%	63.9	5.00E-13	1
	81	Alu-Sb1	85%	58	3.00E-11	1
	75	Alu-Fla?	86%	54	5.00E-10	4
	123	Alu-Sb2	82%	44.1	5.00E-07	1
>m10m13R(121)						
CENSOR	56	SVA	69%	186		
	61	Pb1	80%	326		
BLASTN Alu (143)	74	Alu-Sq	91%	99.6	9.00E-24	7
	74	Alu-Sx	91%	99.6	9.00E-24	43
	74	Alu-Sc	90%	91.7	2.00E-21	5
	74	Alu-Sp	90%	91.7	2.00E-21	10
	45	Alu-J	95%	69.9	8.00E-15	23
	74	Alu-Sb	86%	67.9	3.00E-14	5
	74	Alu-Sb1	85%	60	8.00E-12	1
	74	Alu-Sb2	83%	52	2.00E-09	1
	45	Alu-S	88%	50.1	7.00E-09	1
		Alu-Spqx		44.1	5.00E-07	1
		Alu-FLA?		36.2	1.00E-04	3

Sequence name (Length bp) Analysis name (Number of hits)	Length of match	Hit sequence name	Similarity	Score (Bits)	E-value*	Number of instances of same family or with alternate assemblies
>m9m13R(157)						
CENSOR	141	L1PA7_5	87%	775		
>m1m13R(152)						
BLASTN Alu (1)	12	Alu-J	100%	24.3	0.43	1
BLASTN Human genome build 36 (2)	143	Chromosome 6	99%	259	5.00E-67	2
>m7m13R(187)						
BLASTN Alu (46)	21	Alu-Sq	95%	34.2	4.00E-04	4
	20	Alu-J	95%	32.2	0.002	22
	17	Alu-Sx	94%	26.3	0.11	7
	12	Alu-Sp	100%	24.3	0.43	5
	11	Alu-FLA?	100%	22.3	1.7	1
	11	Alu-Sc	100%	22.3	1.7	2
	11	Alu-Sb	100%	22.3	1.7	3
	11	Alu Sb1	100%	22.3	1.7	1
	11	Alu-Sb2	100%	22.3	1.7	1
BLASTN Human genome build 36 (5)	178	Chromosome 7	97%	303	2.00E-80	3
>m14m13R(181)						
CENSOR	44	MEN	77%	221		
BLASTN Alu (104)	35	Alu-Sq	97%	61.9	2.00E-12	8
	32	Alu-J	96%	56	1.00E-10	52
	39	Alu-Sb1	92%	54	5.00E-10	1
	39	Alu-Sb2	92%	54	5.00E-10	1
	39	Alu-Sc	92%	54	5.00E-10	4
	39	Alu-Sp	92%	54	5.00E-10	6
	39	Alu-Sx	92%	54	5.00E-10	22
	39	Alu-Sb	92%	54	5.00E-10	4
	32	Alu-Spqx	90%	40.1	7.00E-06	2
BLASTN Human genome build 36 (5)	172	Chromosome 7	100%	318	8.00E-85	3
>E1001M13R(201)						
BLASTN Alu (6)	12	Alu-J	100%	24.3	0.43	3
	12	Alu-Sp	100%	24.3	0.43	1
	11	Alu-Sx	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	192	Chromosome 9	99%	348	1.00E-93	2

Sequence name (Length bp) Analysis name (Number of hits)	Length of match	Hit sequence name	Similarity	Score (Bits)	E-value*	Number of instances of same family or with alternate assemblies
>FC1M13R(148)						
CENSOR	146	L1Pt_5end	96%	1059		
BLASTN Alu (22)	13	Alu-Sx	100%	26.3	0.11	4
	12	Alu-J	100%	24.3	0.43	13
	11	Alu-Spqx	100%	22.3	1.7	1
	11	Alu-FLA?	100%	22.3	1.7	2
	11	Alu-Sq	100%	22.3	1.7	1
>FC2M13R(198)						
CENSOR	179	TIGGER1	84%	1064		
BLASTN Alu (26)	12	Alu-Sx	100%	24.3	0.43	8
	12	Alu-Sq	100%	24.3	0.43	1
	12	Alu-J	100%	24.3	0.43	15
	11	Alu-Spqx	100%	22.3	1.7	1
	11	Alu-FLA?	100%	22.3	1.7	1
>FC3M13R(164)						
BLASTN Alu (11)	13	Alu-Sx	100%	26.3	0.11	3
	12	Alu-J	100%	24.3	0.43	5
	12	Alu-Sp	100%	24.3	0.43	1
	11	Alu-FLA?	100%	22.3	1.7	1
	11	Alu-Sq	100%	22.3	1.7	1
BLASTN Human genome build 36 (3)	164	Chromosome 7	100%	303	2.00E-80	3
>FC4M13R(177)						
CENSOR	100	L1PREC2	89%	724		
	75	L1PA15	89%	584		
BLASTN Alu (5)	12	Alu-Sc	100%	24.3	0.43	1
	12	Alu-J	100%	24.3	0.43	3
	11	Alu-Sx	100%	22.3	1.7	1
>FC5M13R(156)						
BLASTN Human genome build 36 (2)	140	Chromosome 4	98%	259	5.00E-67	2
>FC6M13R(82)						
BLASTN Alu (4)	11	Alu-J	100%	22.3	1.7	2
	11	Alu-Sx	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	69	Chromosome 13	98%	121	2.00E-25	2
>FC7M13R(176)						
BLASTN Alu (5)	13	Alu-J	100%	26.3	0.11	4
	13	Alu-FLA?	100%	26.3	0.11	1
BLASTN Human genome build 36 (2)	172	Chromosome 4	100%	318	8.00E-85	2

Sequence name (Length bp) Analysis name (Number of hits)	Length of match	Hit sequence name	Similarity	Score (Bits)	E-value*	Number of instances of same family or with alternate assemblies
>FC9M13R(191)						
CENSOR	191	L1	98%	1743		
Alu Blast (3)	12	Alu-J	100%	24.3	0.43	3
>FC10M13R(189)						
CENSOR	172	AluY	97%	1285		
BLASTN Alu (191)	172	Alu-Sb1	95%	248	2.00E-68	1
	172	Alu-Sb	93%	232	9.00E-64	7
	88	Alu-Sc	98%	167	5.00E-44	5
	93	Alu-Sb2	96%	161	3.00E-39	1
	144	Alu-J	90%	153	7.00E-40	39
	87	Alu-Sx	96%	143	7.00E-37	31
	87	Alu-Sp	95%	135	2.00E-34	5
	87	Alu-Sq	95%	135	2.00E-34	9
	42	Alu-Spqx	95%	67.9	3.00E-14	2
>FC11M13R(596)						
BLASTN Alu (17)	13	Alu-Sx	100%	26.3	0.11	6
	11	Alu-J	100%	22.3	1.7	8
	11	Alu-Sc	100%	22.3	1.7	1
	11	Alu-FLA?	100%	22.3	1.7	1
	11	Alu-Sp	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	592	Chromosome 16	99%	1081	0	2
>FC12M13R(164)						
BLASTN Human genome build 36 (4)	164	Chromosome 10	100%	303	2.00E-80	4
>FC13M13R(37)						
>FC14M13R(162)						
BLASTN Alu (20)	13	Alu-Sx	100%	26.3	0.11	1
	13	Alu-J	100%	26.3	0.11	15
	12	Alu-FLA?	100%	24.3	0.43	2
	11	Alu-Spqx	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	163	Chromosome 5	98%	289	6.00E-76	2
>FC15M13R(175)						
CENSOR	51	L1ME2	91%	346		
BLASTN Alu (9)	11	Alu-J	100%	22.3	1.7	8
	11	Alu-Sx	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	174	Chromosome 17	98%	302	8.00E-80	2

Sequence name (Length bp) Analysis name (Number of hits)	Length of match	Hit sequence name	Similarity	Score (Bits)	E-value*	Number of instances of same family or with alternate assemblies
>FC16M13R(575)						
BLASTN Alu (31)	14	Alu-Sq	100%	28.2	0.027	3
	13	Alu-J	100%	26.3	0.11	21
	13	Alu-Sx	100%	26.3	0.11	4
	11	Alu-FLA?	100%	22.3	1.7	1
	11	Alu-Sc	100%	22.3	1.7	1
BLASTN Human genome build 36 (3)	581	Chromosome 7	96%	935	0	3
>FC17M13R(206)						
BLASTN Alu (11)	12	Alu-J	100%	24.3	0.43	5
	12	Alu-Sc	100%	24.3	0.43	2
	12	Alu-Sx	100%	24.3	0.43	2
	11	Alu-Spqx	100%	22.3	1.7	1
	11	AluSq	100%	22.3	1.7	1
>FC18M13R(242)						
CENSOR	117	AluJB	92%	842		
BLASTN Alu (145)	117	Alu-J	91%	153	7.00E-40	61
	117	Alu-Sq	88%	129	1.00E-32	8
	117	Alu-Sx	88%	129	1.00E-32	17
	117	Alu-Sp	87%	113	6.00E-28	4
	117	Alu-Sc	85%	91.7	2.00E-21	2
	117	Alu-Sb2	84%	83.8	5.00E-19	1
	117	Alu-Sb	84%	83.8	5.00E-19	4
	47	Alu-Sb1	95%	77.8	3.00E-17	1
	45	Alu-Fla?	93%	65.9	1.00E-13	2
BLASTN Human genome build 36 (12)	194	Chromosome 1	100%	359	5.00E-97	2
	42	Chromosome 14	95%	65.8	1.00E-08	2
>FC19M13R(183)						
BLASTN Alu (3)	11	Alu-J	100%	22.3	1.7	2
	11	Alu-Sx	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	181	Chromosome X	99%	327	1.00E-87	2
>FC20M13R(156)						
BLASTN Alu (7)	12	Alu-J	100%	24.3	0.43	4
	12	Alu-Sx	100%	24.3	0.43	1
	11	Alu-Sb	100%	22.3	1.7	1
	11	Alu-Sp	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	152	Chromosome 13	98%	268	8.00E-70	2

Sequence name (Length bp) Analysis name (Number of hits)	Length of match	Hit sequence name	Similarity	Score (Bits)	E-value*	Number of instances of same family or with alternate assemblies
>FC21M13R(152)						
CENSOR	62	B1F1	83%	321		
BLASTN Alu (144)	47	Alu-Sp	97%	85.7	1.00E-19	5
	51	Alu-Sx	94%	77.8	3.00E-17	22
	44	Alu-Sb1	95%	71.9	2.00E-15	1
	44	Alu-J	95%	71.9	2.00E-15	52
	44	Alu-Sb2	95%	71.9	2.00E-15	1
	44	Alu-Sc	95%	71.9	2.00E-15	4
	44	Alu-Sq	95%	71.9	2.00E-15	8
	44	Alu-Sb	95%	71.9	2.00E-15	5
	41	Alu-Spqx	90%	50.1	7.00E-09	2
BLASTN Human genome build 36 (2)	153	Chromosome 5	98%	270	2.00E-70	2
>FC22M13R(169)						
BLASTN Alu (2)	12	Alu-Sx	100%	24.3	0.43	1
	11	Alu-J	100%	22.3	1.7	1
BLASTN Human genome build 36 (2)	163	Chromosome 3	99%	296	4.00E-78	2
>FC23M13R(198)						
CENSOR	176	L1PA16_5	75%	695		
BLASTN Alu (6)	11	Alu-J	100%	22.3	1.7	5
	11	Alu-Sx	100%	22.3	1.7	1
>FC24M13R(174)						
CENSOR	152	L1MC1	80%	661		
BLASTN Alu (2)	11	Alu-Sx	100%	22.3	1.7	1
	11	Alu-FLA	100%	22.3	1.7	1
>FC25M13R(174)						
CENSOR	124	BSRa	72%	444		
	37	BSR	86%	250		
BLASTN Alu (2)	11	Alu-Sp	100%	22.3	1.7	1
	11	Alu-Sx	100%	22.3	1.7	1
>FC26M13R(41)						
>FC27M13R(173)						
BLASTN Alu (11)	12	Alu-J	100%	24.3	0.43	9
	12	Alu-Sb	100%	24.3	0.43	1
	12	Alu-Sq	100%	24.3	0.43	1
BLASTN Human genome build 36 (2)	174	Chromosome 20	98%	302	8.00E-80	2

* Hits with lower E-values are more significant than those with higher values.

Only the first instance of matches in the same family or alternate assemblies are shown in the 'Hit sequence name' column, the number of occurrences of following instances are given in the last column. BLAST analysis was done against different databases as shown in the first column and default parameters were used.

References

1. Lo YM, Tein MS, Lau TK, Haines CJ, Leung TN, Poon PM, et al. Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. Am J Hum Genet 1998;62:768-75.
2. Anker P, Mulcahy H, Chen XQ, Stroun M. Detection of circulating tumour DNA in the blood (plasma/serum) of cancer patients. Cancer Metastasis Rev 1999;18:65-73.
3. Ziegler A, Zangemeister-Wittke U, Stahel RA. Circulating DNA: a new diagnostic gold mine? Cancer Treat Rev 2002;28:255-71.
4. Jahr S, Hentze H, Englisch S, Hardt D, Fackelmayer FO, Hesch RD, Knippers R. DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. Cancer Res 2001;61:1659-65.
5. Silva JM, Silva J, Sanchez A, Garcia JM, Dominguez G, Provencio M, et al. Tumor DNA in plasma at diagnosis of breast cancer patients is a valuable predictor of disease-free survival. Clin Cancer Res 2002;8:3761-6.
6. Bianchi DW. Circulating fetal DNA: its origin and diagnostic potential-a review. Placenta 2004;25 Suppl A:S93-S101.
7. Chan KC, Zhang J, Hui AB, Wong N, Lau TK, Leung TN, et al. Size distributions of maternal and fetal DNA in maternal plasma. Clin Chem 2004;50:88-92.
8. Bremnes RM, Sirera R, Camps C. Circulating tumour-derived DNA and RNA markers in blood: a tool for early detection, diagnostics, and follow-up? Lung Cancer 2005;49:1-12.
9. Umetani N, Kim J, Hiramatsu S, Reber HA, Hines OJ, Bilchik AJ, Hoon DS. Increased integrity of free circulating DNA in sera of patients with colorectal or periampullary cancer: direct quantitative PCR for ALU repeats. Clin Chem 2006;52:1062-9.
10. Jurka, J, Klonowski, P, Dagman, V, Pelton, P. CENSOR - a program for identification and elimination of repetitive elements from DNA sequences. Computers and Chemistry 1996;20 :1: 119-22.
11. Rice, P, Longden, I, Bleasby, A. EMBOSS*:* The European Molecular Biology Open Software Suite Trends in Genetics 2000;16:6:*276-7.*
12. Bailey, TL, Elkan, C. "Fitting a mixture model by expectation maximization to discover motifs in biopolymers", Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology. 1994;28-36
13. Bailey, TL, Gribskov, M. Combining evidence using p-values: application to sequence homology searches. Bioinformatics. 1998;14:48-54.
14. Stothard P. The Sequence Manipulation Suite: JavaScript programs for analyzing and formatting protein and DNA sequences. Biotechniques 2000;28:1102-4.
15. NCBI BLAST [Web:] http://www.ncbi.nlm.nih.gov/blast
16. Mighell AJ, Markham AF, Robinson PA. Alu sequences. FEBS Lett 1997;417:1-5.

17. Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, et al. Standardized nomenclature for Alu repeats. J Mol Evol 1996;42:3-6.
18. Jasinska A, Krzyzosiak WJ. Repetitive sequences that shape the human transcriptome. FEBS Lett 2004;567:136-41.

Chapter 5
Summary and conclusions

Circulating DNA is present in the blood of healthy as well as diseased individuals. It has been found by many research groups that the levels of circulating DNA is higher in cancer patients than in healthy individuals and results obtained in this study agree with this.

Although the origin of circulating DNA is unknown, a few hypotheses about its source are published. The first and most popular is that DNA is released after apoptosis or necrosis of cells and the second is that DNA may be released by living cells. It was concluded in this study that apoptosis or necrosis may not be the major source of circulating DNA in the blood because of a number of reasons. (a) Malignant cells become resistant to apoptosis. (b) Major apoptosis causes inflammation and possibly an autoimmune reaction which is not seen in cancer. (c) During cancer treatment, which causes apoptosis, lower levels of circulating DNA can be observed in the blood of patients. Apoptosis thus may play a minor role in the origin, but is not the main source of circulating DNA in the blood, while release of DNA by living cells may be the major source. However, the disturbance of the equilibrium between the release of DNA, by living cells, and the mechanisms used for clearing this DNA may play the key role in the appearance of increased amounts of circulating DNA in the blood.

Before circulating DNA can be studied, it has to be isolated from the blood. It has been shown in this study that a variety of preanalitycal factors affect the quantity as well as the quality of the circulating DNA. Furthermore, the choice of isolation and quantification method also has a great impact on the resultant study. In this study it was decided to use the methods most frequently used in other studies to isolate DNA for quantification in order to be able to compare our results with those published. The resultant quantification of the levels of free

DNA is in concordance with the published results. It was therefore concluded that this analysis was successful. A number of other methods were also used for isolation of circulating DNA, but the obtained DNA was used for cloning as the QIAamp blood DNA mini kit did not isolate sufficient amounts of DNA for cloning.

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 unknown. Therefore, this study was an attempt to characterize total circulating DNA. A method was developed to clone and sequence circulating DNA and a total of thirty-five sequences were obtained and partially analyzed. It was found that the majority of the cloned sequences contain Alu repeats, therefore, further analysis is needed to determine the meaning of these repeats. Furthermore characterization of total circulating DNA may be beneficial in diagnosis and prognosis and may also contribute to determining the source and function of circulating DNA.

The goal of this study was to isolate, clone and sequence total circulating DNA from plasma. This goal was reached and the study was thus successful. However, it was found that more work need to be done in order to reach the end goal of identifying biomarkers for genetic aberrations in humans. The method is in place, thus more sequences can be generated from healthy as well as diseased individuals in the future. In addition, more or better bioinformatics tools need to be employed more effectively to reduce the massive amount of data generated by one sequence and to make sense of the seemingly endless amount of information.

References

ANKER, P & STROUN, M (2001) DNA Extraction from Serum and Plasma Samples. [*web:*] *http://www.unige.ch/LABPV/symposium/cnaps/cnaps.html* [Date of access: 28 Jun. 2005].

BREMNES, R. M., SIRERA, R. & CAMPS, C. (2005) Circulating tumour-derived DNA and RNA markers in blood: a tool for early detection, diagnostics, and follow-up? *Lung Cancer,* 49, 1-12.

CHAN, A. K., CHIU, R. W. & LO, Y. M. (2003) Cell-free nucleic acids in plasma, serum and urine: a new tool in molecular diagnosis. *Ann Clin Biochem,* 40, 122-30.

CHEN, Z., FADIEL, A., NAFTOLIN, F., EICHENBAUM, K. D. & XIA, Y. (2005) Circulation DNA: Biological implications for cancer metastasis and immunology. *Medical Hypotheses,* 65, 956-961.

GEORGIOU, C. D., PATSOUKIS, N. & PAPAPOSTOLOU, I. (2005) Assay for the quantification of small-sized fragmented genomic DNA. *Anal Biochem,* 339, 223-30.

HERRERA, L. J., RAJA, S., GOODING, W. E., EL-HEFNAWY, T., KELLY, L., LUKETICH, J. D. & GODFREY, T. E. (2005) Quantitative analysis of circulating plasma DNA as a tumor marker in thoracic malignancies. *Clin Chem,* 51, 113-8.

LO, Y.M., TEIN, M.S., LAU, T.K., HAINES, C.J., LEUNG, T.N., POON, P.M. WEINSCOAT, J.S., JOHNSON, P.J., CHANG, A.M. & HJELM, N.M. (1998) Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am J Hum Genet*, 62, 768-75.

LUI, Y.Y., CHIK, K.W., CHIU, R.W., HO, C.Y., LAM, C.W. & LO, Y.M. (2002) Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin Chem*, 48, 421-7.

MCLAREN HOWARD, J. (2005) Cell Free DNA [web:] *http://www.biolab.co.uk/docs/dnacellf.pdf* [Date of access: 27 Oct. 2005].

SAMBROOK, J. & RUSSELL, D.W. (2001) Molecular cloning: A Laboratory Manual. 3rd Ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

SCHMIDT, B., WEICKMANN, S., WITT, C. & FLEISCHHACKER, M. (2005) Improved method for isolating cell-free DNA. *Clin Chem*, 51, 1561-3.

WU, T. L., ZHANG, D., CHIA, J. H., TSAO, K. H., SUN, C. F. & WU, J. T. (2002) Cell-free DNA: measurement in various carcinomas and establishment of normal reference range. *Clin Chim Acta,* 321, 77-87.

ZIEGLER, A., ZANGEMEISTER-WITTKE, U. & STAHEL, R. A. (2002) Circulating DNA: a new diagnostic gold mine? *Cancer Treat Rev,* 28, 255-71.

Addenda

Addendum A

List of figures

Flow diagram of experimental work 20
Levels of circulating DNA in plasma in different conditions 44
Levels of circulating DNA in serum in different conditions 45
Products of colony PCR of pEZSeq ligated DNA 58
Products of colony PCR of pEZSeq transformants 58
Amplification curve of run 1 of all the samples 78
Amplification curve of run 2 of all the samples 78
Standard curve constructed after a concentration range of genomic DNA was run in triplicate. 78

List of tables

Comparison of different isolation and quantification methods with subsequent results 36
Table of sequence analysis data 60
Table of Real-time PCR data 77

Abbreviations

+Amp	With Ampicillin
ATP	Adenosine triphosphate
bp	Base pairs
DNA	Deoxyribonucleic acid
EDTA	Ethylenediamine tetra-acetate
EMBOSS	European Molecular Biology Open Software Suite
FAM	6 carboxyflourescein
g	Times gravity
kb	Kilo base pairs
MAST	Motif Alignment & Search Tool
MEME	Multiple Em for Motif Elicitation
NaCl	Sodium Chloride
PCR	Polymerase chain reaction
SDS	Sodium dodecyl sulfate
TAMRA	6 carboxy-tetramethylrhodamine
V	Volt

Addendum B

Conference proceedings

Abstract prepared for "a Molecular Meander in the Midlands" SASBMB XXth conference held from the 2nd to the 5th of July 2006, a poster was presented by the first author.

Characterization of circulating DNA as a biomarker for genetic aberrations in humans

[1]**Maniesh van der Vaart,** [2]Samuel J Fourie and Piet J Pretorius

[1]Division of Biochemistry, North-West University, Private Bag X6001, Potchefstroom 2520, SOUTH AFRICA, Maniesh.vanderVaart@nwu.ac.za. [2] Wilmed Park Oncology, Wilmed Park Hospital, Klerksdorp, SOUTH AFRICA, sfourie@oncology-sa.co.za

Free DNA or circulating DNA in humans is fragmented DNA that are fairly short, it contains multiples of 180 to 200 bp and can be up to 1000 bp long. Free DNA is circulating in the blood of healthy as well as sick individuals and can be detected in the blood, serum and urine of healthy as well as sick individuals. Through the years it was demonstrated that tumor-related DNA is not confined to a specific cancer type, it appears to be a common finding across different malignancies. The fact that cancer patients have much higher concentrations of circulating DNA makes it a very useful tool for non-invasive diagnosis. The origin of circulating DNA is still enigmatic. Apoptosis or necrosis is seen by many as the source, but release by living cancer cells is an alarming possibility which may prove that cancer is an infectious disease.

A method for isolating total circulating DNA from human plasma was standardized by adjusting a commercially available DNA isolation kit. The DNA isolated was on average 300 – 400 bp long and significantly more circulating DNA was isolated form a diagnosed breast cancer patient than from control persons. Since the aim of this study is to determine whether free DNA can be utilized as a biomarker(s) for specific malignancies. We cloned and sequenced the total isolated circulating DNA to identify and characterize any sequences specific for a specific malignancy. The presence of specific markers, e.g. P16 and RASSF1A, in the total isolated free DNA was investigated with PCR. These results will be presented for the first time at the Molecular Meander in the Midlands congress 2006.

Addendum C

Buffers and media

LB+Amp culture media

For 1 liter, mix:

10 g	Tryptone
5 g	Yeast extract
10 g	NaCl

In 950 ml of ddH_2O and adjust pH to 7 with NaOH, fill to 1 L and autoclave 20 minutes. When cooled to ~55°C add Ampicillin to a final concentration of 100 µg/ml (Sambrook and Russel, 2001).

SOC recovery media

For 1 liter, mix:

20 g	Tryptone
5 g	Yeast extract
0.5 g	NaCl

In 900 ml of ddH_2O, add 10 ml 250 mM KCL and adjust pH to 7 with NaOH, autoclave 20 minutes and add 5 ml sterile 2 M $MgCl_2$ when cooled, mix and add glucose to a final concentration of 20 mMm, fill to 1 L and store at 4°C (Sambrook and Russel, 2001).

TAE buffer for agarose gel electrophoresis

For 1 liter 50x stock solution, mix:

242 g	Tris base
57.1 ml	glacial acetic acid
100 ml	0.5 M EDTA (pH 8.0)

In 800 ml of ddH_2O and fill to 1 L, for agarose gel electrophoresis, dilute the stock solution to a 1x working solution (Sambrook and Russel, 2001).

YT+Amp growth media

For 1 liter, mix:

5 g	Yeast extract
8 g	Tryptone
5 g	NaCl
15 g	Agar

In 950 ml of ddH_2O and adjust pH to 7 with NaOH, fill to 1 L and autoclave 20 minutes. When cooled to ~55°C add Ampicillin to a final concentration of 100 µg/ml. Pour media into Petri dishes and allow to solidify, store at 4°C for up to 1 week (pEZSeq™ Blunt Cloning kit manual).

Addendum D

Calculation of Real-time PCR data

Table of Real-time PCR data

	Run 1 (ng)	Run 2 (ng)	Average (ng)	Standard deviation (ng)	Circulating DNA (ng/ml)	Average (ng/ml)	Range (ng/ml)
Samples of healthy controls centrifuged once							
C1 E	0.0048	0.0075	0.0061	0.0019	0.76		
C1 Li	0.0019	0.0043	0.0031	0.0017	0.39		
C1 C	0.0094	0.0102	0.0098	0.0006	1.22		
C1 A	0.0010	0.0041	0.0025	0.0022	0.32		
C1 Z	0.0239	0.0309	0.0274	0.0049	3.43		
C1 P	0.0131	0.0022	0.0076	0.0077	0.95		
C1 M	0.0004	Capillary broke					
C1 L	0.0417	0.0724	0.0571	0.0217	7.13		
C1 1	0.0043	0.0080	0.0062	0.0026	0.77		
C1 3	0.0003	0.0042	0.0022	0.0028	0.28	1.53	(0.28-7.13)
Samples of healthy controls centrifuged twice							
C2 E		0.0003	0.0003		0.04		
C2 Li	0.0062	0.0003	0.0032	0.0042	0.40		
C2 C		0.0052	0.0052		0.65		
C2 A			0.0000		0.00		
C2 Z	0.0007	0.0052	0.0029	0.0031	0.37		
C2 P	0.0034		0.0034		0.43		
C2 M	0.0044	0.0212	0.0128	0.0119	1.60		
C2 L	0.0063		0.0063		0.79	0.53	(0.00-1.60)
Samples of treated breast cancer patients							
Patient 1	0.0619	0.0278	0.0449	0.0241	5.61		
Patient 2	0.0073	0.0032	0.0053	0.0029	0.66		
Patient 3	0.0025	0.0020	0.0023	0.0004	0.28		
Patient 4	0.1580	0.1510	0.1545	0.0049	19.31		
Patient 5	0.0020	0.0030	0.0025	0.0007	0.32		
Patient 6	0.0905	0.0681	0.0793	0.0158	9.91		
Patient 7	0.0214	0.0706	0.0460	0.0348	5.75		
Patient 8	0.3330	0.3660	0.3495	0.0233	43.69		
Patient 9	0.0472	0.0163	0.0318	0.0218	3.97		
Patient 10	0.1800	0.1200	0.1500	0.0424	18.75	10.82	(0.28-43.69)
STD 10ng	10.0000	10.0000	10.0000	0.0000			

Real-time PCR Amplification curves

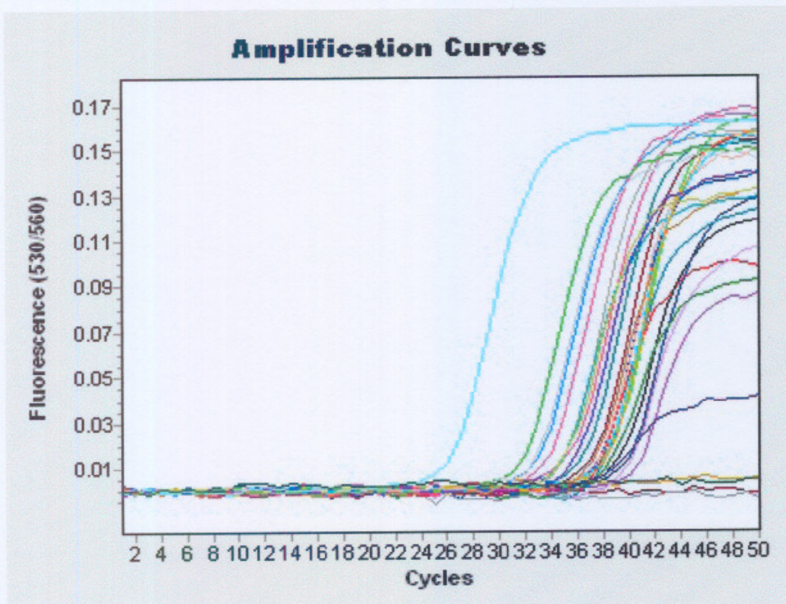

Fig 1: Amplification curve of run 1 of all the samples.

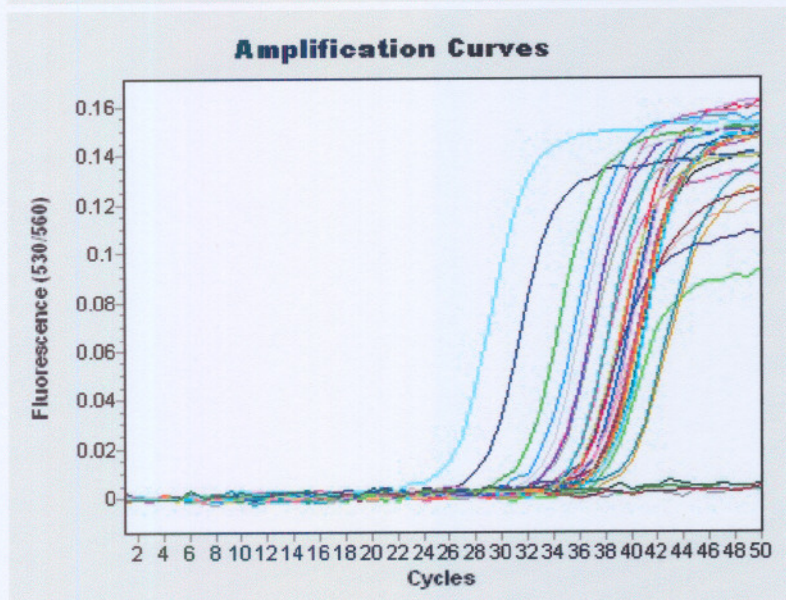

Fig 2: Amplification curve of run 2 of all the samples.

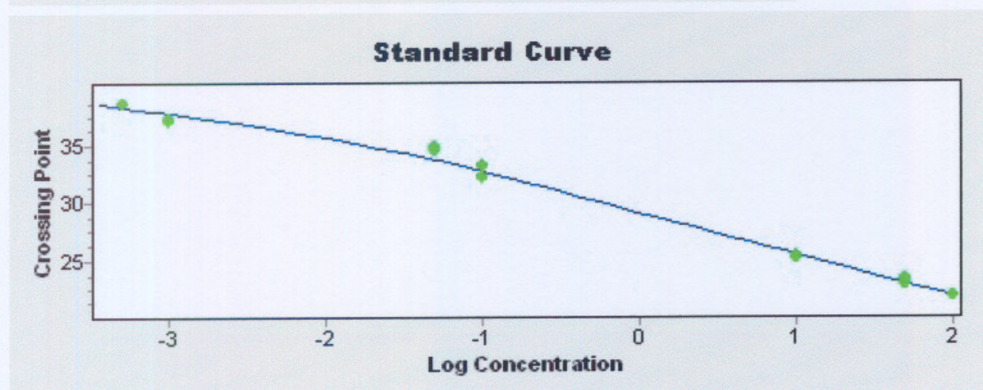

Fig 3: Standard curve constructed after a concentration range of genomic DNA was run in triplicate.

Addendum E

Raw sequence data

Sequence "m6m13R" 124 residues

```
  1 GTAATCCTAG CTACTTGGGA GGCTGAGGCA GGAGAAGCAC TTGTACCCAG
 51 GCGGTGGAGC TTGCAGTGAG CCGAGATTGT GCCACTGCAC TCCAGCCTGG
101 GCGACAGAGC GAGACTCTGT CTCA
```

Sequence "m5m13R" 188 residues

```
  1 TGCAGGAATT CTGACGTTCC AAGTAGAGAG AGATTGTGAT TCTATGCCTT
 51 ATGCAAATCT GAATCTGGAG GGGACTCCTC CTCTGGGGAT GCAGTTAACT
101 TGAGGTGTTC CAGAAAGGCT ATCTACAGGT CCACTCATGC TAAATTCCCA
151 TAAGAGAAGT ACCAGCTGTG TCTGCAGCAG TGAATGAG
```

Sequence "m2m13R" 161 residues

```
  1 AGCTTGCTGC CTCTACTAAA AATACAAAAA ATAGCCGGGT GTGGTGGTGC
 51 ATGCTGGTAA TCCCACCCTA CTCAGGAGCT AAGGCAGGAG AATCACTTGA
101 ACCCAGGAGT GGAGGTTGCA GTGAGCTGAG ATCATGCCAC GGCACTCCAG
151 CCTGAGCGAC A
```

Sequence "m10m13R" 121 residues

```
  1 AGTTTGCTCG CTATGTCACC TGGGCTGGAA TGCAATAGCA CGAATTCCGG
 51 CTCACTGCAA CCTCTGCCTC CTGGGTTCAA GCGATTCTCC TGCTACAGGC
101 TCCTGAGTAG CTGGGATTAC A
```

Sequence "m9m13R" 157 residues

```
  1 AGCATTGCTC CCCACAGCCG CCCCTTTCCC CCAGATGTTC TGTCCCAGGA
 51 AAATGGTGGG TTTTATCTTT TAAGTCCCTG CTGGGGCTGC TCGCCTTTTT
101 TTTCAGAGAC GCCCTGCCCA GAGAGAAGAA ATCTGGCAGT CTGGCCACAG
151 CATTCTT
```

Sequence "m1m13R" 152 residues

```
  1 AGCGTTGTCA GCATCAGGAC AAAAGTCGTC ATCCAAAAAT TCCAATTCTA
 51 GTTTTCTCTT GGTCTGGGCT TAAAGCCAGA GGCGATTCTC CATAGGTGGC
101 ATGCCAGTGC CATCTGACTG CTGACATATT CTGTGTGGTG GTTGAAAAAA
151 CT
```

Sequence "m7m13R" 187 residues

```
  1 CAGCATCGTC TCCCACAAGT GCGTGGGCAT TATGGGCATG AGCCACCGCC
 51 GCCCAACCCC AATACCTATT TCTGATGTGA TAGGGTAGAA AATTGAAAAG
101 GGAGGACCAC CCAGAGATTG GCCAATGGTA CGGAAAGCTG TGGACACGTG
151 AGTTTCTCCT GAAACGCATT TTGTTTAGCA TCTCCTT
```

Sequence "m14m13R" 181 residues

```
  1 AGCGTCGTCC CCAAAGTGCT GGGATTATGG GCATGAGCCA CCGCGCCCAA
 51 CCCCAATACC TATTTCTGAT GTGATAGGGT AGAAAATTGA AAAGGGAGGA
101 CCACCCAGAG ATTGGCCAAT GGTACGGAAA GCTGTGGACA CGTGAGTTTC
151 TCCTGAAACG CATTTTGTTT AGCATCTCCT T
```

Sequence "E1001M13R" 201 residues

```
  1 AGCGTTGTCT CCCCATCTGC TGGTTGAATG CAGGGGAATC CAGTAGAAGA
 51 TCCTGAAATC CAAGAGGTTG GTGGGGCCAT TAGCTAGAAA GAGTGTTCCT
101 GAGTTATTGG TGGAGCAAGC TGCCCCTCCC CACAACAACC TTCATTGTAT
151 TGTCACATGT CACTGAGATT TTAAGGCTGT TTTTTTATTT TGGTTATCTT
201 A
```

Sequence "FC1M13R" 148 residues

```
  1 CTCCGCCATT GCCCAGGGCT TGCTTAGGTA AACAAAGCAC GCCAGGAAGC
 51 TCGAACTGAG GTGGAGCCCA CTACAGCTCA AGGAGGCCCT GCCTGCCTCT
101 GTAGGCTCCA CCTCTGGGGG CAGGGCACAG ACAAACAAAA AGACAGCA
```

Sequence "FC2M13R" 198 residues

```
  1 ACGCCAAGCC TTGTCCATTC AAGAACTTTT CCTTTGCATT CACAATTTGG
 51 CTAACTCTTT GGCACAGTGG CCTAGCTCTC AGCCTATCAT TGTTTTTGAC
101 CTTCCTCTCT CAGCAAATTT AATTATTTCT AGCTTTTAAT TTGAAAAGAG
151 AGGCGTGCAA CTCTTCCTTT CACTTGAACA CTTAGAGGCC ATTGTACG
```

Sequence "FC3M13R" 164 residues

```
  1 TGACATCAGT GTTTCACATC AACCTGACTG TCTTAATTTC TTAGATTTAG
 51 TAAATGTCTG TATTCCCTCA GGAGAGCTAC ATCTAAGTGT GTGTGTTTAT
101 ATAAATACAC ACATAATTAT ATTTATATAT TGTTCACAAT GGAAAACATA
151 AATGGAATCC TTGC
```

Sequence "FC4M13R" 177 residues

```
  1 ATTCTGTTCC AGTTGGTCTA TGTGTCTGTT TTTGTACTAG TACCATGCTA
 51 TTTTGGTTAT TGTAACCTTG CAGTATAGTT TGAATTCAGG TAATGTGATG
101 GCACACATAT GTTCATCGTA GCACTATTCA CAATAGTAAA GACATAGAAT
151 CAACCTAAAT GCCCAGCAGT GGTAGGT
```

Sequence "FC5M13R" 156 residues

```
  1 GCCAAGCATT GTCTTGATGT ATAATTTATA ATTTATAAAA TTTATAAATT
 51 ATACATAGAT AATTTATAAT TTATAATTTA TAAAAGATAC TATTGATCAG
101 TCCCACTAGC ATTCAAAATA TTTTTCTCAT CTTCAATTTC GTTAAAGAGT
151 TTGTGT
```

Sequence "FC6M13R" 82 residues

```
  1 CGCCAAGATT GTCGTATTCA CACAAGAACC CATTTAGTAA AGAACTTCAC
 51 ACTTACAGTT CTTTGGCTAG TCTGCACCTC AC
```

Sequence "FC7M13R" 176 residues

```
  1 CGTCTTGTAA ATTATAGCCA CAATGATAAT GAGAGAAAAA AATGTGGACT
 51 TAACTATATT TGGAAAACAA TTTTTTGGGA AACAAGTAAA TTTATAAATT
101 ACTTACTAAA GCATCGAGTT CATGAATTCA GAAAAACATT AACATCGCTA
151 AAAAATTAAA ATCTAAAATA ATATAA
```

Sequence "FC9M13R" 191 residues

```
  1 CTAAAATTCT CTTTTTTTGT TGTGTCTCTG CCAGGCTTTG GTATCAGGAT
 51 GATGCTGGCC TCATAAAATG AGTTAGGGAG GATTCCCTCT TTTTCTATTG
101 ATTGGAATAG TTTCGGAAGG AATGGTACCA GCTCCTCCTT GTACCTCTGG
151 TAGAATTCGG CTGTGAATCC AACTGGTCCT GGACTTTTTT T
```

Sequence "FC10M13R" 189 residues

```
  1 GGGACTACAG GCGCCCGCCA CTACGCCCAG CTAATTTTTT GTATTTTTAG
 51 TAGAGACGGG GTTTCACCGT TTTTTTTTTA GCCGGGATGG TCTCGATCTC
101 CTGACCTCGT GATCCGCCCG CCTCGGCCTC CCAAAGTGCT GGGATTACAG
151 GCGTGAGCCA CCGCGCCCGG CCGAACATGA GTTTTTTGA
```

Sequence "FC11M13R" 596 residues

```
  1 TGTCGCAGGC AAGTCGTTTT ATGTTTCCTG AGCAGAAAGA AGGCAATCTC
 51 TTCTGGCAGT GAGGTGACAA GGGTTCCGAT GGATTCCCCT CTCCCAGATC
101 AGAATTAAAC TAGAGCCTGT CAGCCAACTC GCTGAACACT CAATCTTCCA
151 TCTTTATGGG TTTAGCTGTG ATGTTATTGC AGGAGTCTTT CAGCATGTGA
201 AAGGGAGGTC ATCCTGGAAG AATATATGCA CAGAGCCTTC TGTGTTCTCT
251 TTCTATGGGC TCCGGACATT TACCTTGACT TGAAAGGTAA CATCTTGGAA
301 GGACAGATGA ACAGACAGAC AGATGGACAG TACTCGCCAT GAATCCCAGC
351 CACAGCCAAA GCCAACTCCA TCCTTAGATA CCCCATTCTC AAAAGCCAAT
401 ACATTTTCTT TTGTGCCTCT GCTAGGTGGA GTTGGTTTCT GTCTTGTGTT
451 TCAAAATAGT CCCATCTATT ATGCTATCTC TGGCTTACTC TTTCTGAGCC
501 TTGGTTATTT TCACTCACAA AATGGAAACA AATAGCATCA GTCCCGCCTT
551 GCTCATATGG CTGGCATGAA ATTGTTCTTG CAAAATGTCA AGTGTT
```

Sequence "FC12M13R" 164 residues

```
  1 CATTTGGCTT CACAGTCCAG CCAAGTCATT TGTAACAAAG TTATTTGAGG
 51 AATGACTAGG CAGGTTGGGA GAGAAGGTTC AGGAATCTAG CCTTGTTATT
101 AGCATCATTT TAAAAGGCAA ATGAACGATC ATTTAGAGGG ACAAAACTGC
151 ACGTTTTCCG ATGG
```

Sequence "FC13M13R" 37 residues

```
  1 AGTTTATTAG CTCCCGTTTG CCTCATAGCC TCATTTT
```

Sequence "FC14M13R" 162 residues

```
  1 GCTTGACTAG TCCGGGTCAC AGAGGACAAT CCTCCTGAAT AGAGGTGGTG
 51 TGGATGGCAT CAGTGATGTA TTGCTTTAGG CCGATCCAGC AAGCACTTCC
101 AATTAATTGT ACCTCGGCAG CAGTCCTCTG CTTCATACAG GGCAATGGTC
151 CTCAACAGGA GA
```

Sequence "FC15M13R" 175 residues

```
  1 GCCTATTTGA TTTCCCTGAT GACTAATGAT GTCACCATCT CTTCATGTGC
 51 TTATTTGTTG ATGCATTTTA AAAATGTGAT TTCTACTTTC TCACTTGATC
101 GGCAAAGACA TGTTGGCTGT AACTATTAAC TTGGCATCCA CAGCTTTCCC
151 CACTGATTTT TTTTTTTTTG AGACG
```

Sequence "FC16M13R" 575 residues

```
  1 GCTTGTTTAT CTAGTGTCTG TCCATGCCTG TCTCCCTATA CACTGTAAGC
 51 TTCCCAAGAG CATACACAGG ACCATCCCAT CTTTGGGATG CCTAGCAAGA
101 TGTTCATCGG AGGACATAAA ATGTTTGATC GGTGAGTGAA TCGAATGAAT
151 CTGGCACCAT TTCCTTCCTC TTAACCAAAC ACGTGACGTC TTCCTCAGAT
201 TCGCACATTG TCTCACGCTA TGCTCGGGAG CTGGAGTGGA AGACACTTAC
251 AGCTGCGTGT CTTAATCTGA TGATACAATT GCAAGAGACC ACCCATCTGC
301 ATGGAACACA CACAAAAATA AAGCCTGAGT CCAGTGAAGC AAAACCTCTA
351 AAGCAGGAAA CACATGAAAT CACTTGAGAT TCAAGAAATA GGAAGAATGA
401 GAGTTGAAAG GAAGCACGGT AGTTCAGAGA CAAATGGATT TGCCGTGGAT
451 TTTCAAAAAG CACATAACAA GAAGTCTGCC CAGTTATGGG CTTTTGGTTT
501 TTTTTCATAT GGACTGGACC TCTCCTAGTC CTCCTCAGGC CCTTCTCTCC
551 AGCCAGTGGC CTTACATGAC TATGT
```

Sequence "FC17M13R" 206 residues

```
  1 GCTTAAAAAA AAAAAAAGTT AGGTGGGGTG GGGGGGGCCA CATGGCCCGG
 51 CAGGTCTCCA GGAAGGCCTC CAGGCTTACC TCCCGGGAAG CCAAAAAAGA
101 CTGGGAAGGA ATTGGCCTTG GACGCCTTGG GGAAAGGCCA GGGGCCCTCG
151 CAGGTGAGAC GCCCGTGGTG GTGGCGCGCC ACCTTGGGGA CAGATATGTC
201 ACCTTG
```

Sequence "FC18M13R" 242 residues

```
  1 GCTTGAGATT GGGAGCATGG GGATTTGATG GGGATGAGGA TTTGATGGGT
 51 TCGGAAAACT TTGAAACATT TTACTTTCAG CCGGGCGCGG TGGCTCATGC
101 CTGTAATCCC AGCACTTTGG AAGGCCAAAG CTGGAGGATC ACTTGAGCCC
151 AGGAGTTGGA GACCAGCCTG GACAACATGG CCAAACCCCG TCTCTATTTA
201 TTATTTTTTA AATTACGTCT ATTAATTCAT TGAATCCTCA CA
```

Sequence "FC19M13R" 183 residues

```
  1 GTCCTGTAAG AATTAAACAA TCAGCTTTGC ATTACATCTT TCCTGAGCTT
 51 TACCTCTCCG GAACTTCCTT ATAGTTAACT GACACTATAT TTTGTGACTT
101 GACTTCTGTG GGAGCTGAGT GCTGATACAC CCTATGCATC AAAGAAGTTC
151 TTGATGTTTT TCTGACTTTG AATCACAGTT AAC
```

Sequence "FC20M13R" 156 residues

```
  1 CGTTGAGCAG AGCAGACAGG CGCTGGACAG AGGGGCAGGG GAGTGCCAAG
 51 TTGTCCTGGA GGCAGACAGC CCAGCTGAGC CTCCTTACCT CCCTTCAGCC
101 AAGCCCACCT GCACGTGATC TGCTGGCCTC AGGCTGCTGC TCTGCCTTCA
151 TTGCTG
```

Sequence "FC21M13R" 152 residues

```
  1 GCCTGTGACC GTCCACCTCA GCCTCCCAGA GTGCTGGGAT TACAGGCGTG
 51 AGCCACCGCA TCCAGCAGCT AACCTTGATA TTCTAAACAC CACTACTGCT
101 TGCTTCCCTA CACTCATTGG AAGTGGTACC CTGACCTTTT CAATTAATAA
151 GT
```

Sequence "FC22M13R" 169 residues

```
  1 GCTCGGGCCA GAAAGACTTG TGGTGTACTT GGTTTGAGCA GTACCTCTAG
 51 TATGCACCAT TCCAGGGAAC ATGGTTGTCA ACATAACAAG TTTAGTGTGT
101 GTATTTGCAG ACAACATAGC CCTGAAAGCA TGAAATGTAG AATAGGTAAG
151 GTCTGTCTTA GAAAAGGAG
```

Sequence "FC23M13R" 198 residues

```
  1 ATATAGTCTT CTGATATTCC GGTACTGGCA GAAGTTCTCT GTTGCCTCAG
 51 GCAATGGGCT GATTCATAGA ATACACAGTC ATCTGAGCTC CCATGCTTAG
101 CCACCAGAGT GCAGGGCCCA CTAAGTGTAA GGACCAAGAC CAGGCAGGGG
151 TTGCCTACAG ATCCCTCAAT GGCAGGGCAC AAGCACCAGG CACTGAGG
```

Sequence "FC24M13R" 174 residues

```
  1 ATTGCTCTTG TACAGTAGAG GAAAGTGTAC TGTTAGGTAA ACTGTGGACT
 51 TTAGATGATG ATGTGTCACT GTAGGTTCAT CCATTGGAAT AAATGCACCA
101 CTCTTGTGTG GGATATTGAT AGTGGGAAGA CTGCCCAATT AAGAAATCTG
151 TACTTTCTAC TCAATTTTGC TGTA
```

Sequence "FC25M13R" 174 residues

```
  1 GCTTATCAGG CAAGGTATGT CACAAAGCCC CCTGTAGCAA AGCCTTGACG
 51 ATAGTTACAT CACTTCGGTG ATCAGTGGCG AGATCTCTCA TAATTCCCCT
101 GTAGGCAGGG CTTATACAAC AGTTACATCA CCTGGGTGAT CAGTGCAGAG
151 ATATGTCACA AGAATCCTGT ATGC
```

Sequence "FC26M13R" 41 residues

```
  1 TATGTCTGCA TAATTGAACA GCCTTGCTCC CAGGATGAAG C
```

Sequence "FC27M13R" 173 residues

```
  1 GCTCCTGGAT ACAGTTAGAT GCAGTTTCTG TAATTCCAGT ATCCAATTCT
 51 GAAATGCTCA CCAGTTTGGA ATGATTATTT TGCTGATATT ATAAAGATGT
101 ATTGCTTCTT CCACTGAGTC ACCCTGGTGC CTAAAGCTTA CCTCTCCACC
151 CCGATCCTTG CCCTGAACAA AGA
```

Addendum F

Annotation of Repbase Sequences

```
ID   AluJb repbase; DNA; PRI; 283 BP.
DT   20-AUG-1998 (Rel. 1, Created)
DT   20-AUG-1998 (Rel. 1, Last updated,
DT   Version 1)
DE   Alu-Jb subfamily - a consensus.
KW   SINE1; SINE; Non-LTR Retrotransposon;
KW   Interspersed Repeat; Alu-J; Alu-Jb;
KW   AluJ; AluJb; Repetitive sequence.
OS   Primates
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria.
RN   [1]
RA   Jurka J.;
RT   "Origin and evolution of Alu repetitive
RT   elements.";
RL   Molecular Biology Intelligence Unit:The
RL   impact of short interspersed elements
RL   (SINEs) on the host genome (ed. Richard
RL   J.Maraia), R.G. Landes Company, Austin,
RL   pp.25-41 (1995).
CC   [1] (Consensus)
SQ   Sequence 283 BP; 59A; 82C; 98G; 44T;
SQ   0 other;
```

```
ID   AluY repbase; DNA; PRI; 282 BP.
DT   20-AUG-1998 (Rel. 1, Created)
DT   20-AUG-1998 (Rel. 1, Last updated,
DT   Version 1)
DE   Alu-Y subfamily - a consensus.
KW   SINE1; SINE; Non-LTR Retrotransposon;
KW   Interspersed Repeat; Alu-Sb; Alu-Y;
KW   AluY; Repetitive sequence.
OS   Primates
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria.
RN   [1]
RP   180-186
RA   Willard C., Nguyen H.T., Schmid C.W.;
RT   "Existence of at least three distinct
RT   Alu subfamilies.";
RL   J. Mol. Evol 26, (1987).
RN   [2]
RP   1-282
RA   Jurka J., Smith T.;
RT   "A fundamental division in the Alu
RT   family of repeated sequences.";
RL   Proc. Natl. Acad. Sci. USA 85, (1988).
RN   [3]
RP   1-282
RA   Batzer M.A., Kilroy G.E., Richard P.E.,
RA   Shaikh T.H.,Desselle T.D., Hoppens
RA   C.L., Deininger P.L.;
RT   "Structure and variability of recently
RT   inserted Alu family members.";
RL   Nucleic Acids Res 18, (1990).
CC   [2] (Consensus)
SQ   Sequence 282 BP; 62A; 80C; 98G; 42T;
SQ   0 other;
```

```
ID   B1F1 repbase; DNA; ROD; 149 BP.
DT   16-MAR-2006 (Rel. 11.03, Created)
DE   SINE1 from Muridae.
KW   SINE1; SINE; Non-LTR Retrotransposon;
KW   Interspersed Repeat; B1F; B1F1.
OS   Muridae
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia;
OC   Sciurognathi.
RN   [1]
RP   1-149
RA   Smit A.F.;
RT   "B1F1 - a subfamily of SINEs from
RT   Muridae.";
RL   Direct Submission to Repbase Update
RL   (11-NOV-2005).
CC   [1] (Consensus)
SQ   Sequence 149 BP; 38A; 40C; 48G; 23T;
SQ   0 other;
```

```
ID   BSR repbase; DNA; PRI; 68 BP.
DT   01-OCT-1995 (Rel. 1.09, Created)
DT   18-APR-1997 (Rel. 2.03, Last updated,
DT   Version 2)
DE   Human beta satellite DNA - a
DE   consensus.
KW   SAT; Satellite; Simple Repeat; BSR;
KW   Satellite repetitive element.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   1-68
RA   Waye S.J., Huntington F.W.;
RT   "Human beta satellite DNA: Genomic
RT   organization and sequence
RT   definition of a class of highly
RT   repetitive tandem DNA.";
RL   Proc. Natl. Acad. Sci. USA 86, 6250-
RL   6254 (1989).
CC   [1] (Consensus)
CC   Position 25 can be either a or t
CC   Position 31 can be either c or t.
SQ   Sequence 68 BP; 20A; 16C; 18G; 14T;
SQ   0 other;
```

```
ID   BSRa repbase; DNA; PRI; 142 BP.
DT   14-NOV-2005 (Rel. 10.11, Created)
DT   09-DEC-2005 (Rel. 10.11, Last updated,
DT   Version 1)
DE   Satellite from primates.
KW   Satellite; Simple Repeat; BSRa; BSRb.
OS   Primates
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria.
RN   [1]
RP   1-142
RA   Smit A.F.;
```

```
RT   "BSRa - Satellite from primates.";
RL   Direct Submission to Repbase Update
RL   (11-NOV-2005).
CC   [1] (Consensus)
SQ   Sequence 142 BP; 30 A; 44 C; 46 G; 22T;
SQ   0 other;


ID   L1 repbase; DNA; PRI; 5403 BP.
DT   01-OCT-1995 (Rel. 1.09, Created)
DT   18-APR-1997 (Rel. 2.03, Last updated,
DT   Version 2)
DE   Primate L1 consensus.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat; L1 (subfamily
KW   L1PA2); LINE1.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini;
OC   Hominidae; Homo.
RN   [1]
RP   1-5403
RA   Smit A.F.;
RT   "L1.";
RL   Direct Submission to Repbase Update
RL   (1996).
CC   [1] (Consensus)
CC   This is not the complete sequence of
CC   the L1PA2 element. 3' ends
CC   of a variety of L1 subfamilies starting
CC   at position 5254(overlapping 150 bp)
CC   are given separately in the database.
CC   ORF1 is from bp 1030 to 2046, ORF2 from
CC   2110 to > 5403.
SQ   Sequence 5403 BP; 2077A; 1233C; 1102G
SQ   991T; 0 other;


ID   L1MC1 repbase; DNA; PRI; 1080 BP.
DT   23-JUN-2000 (Rel. 5.05, Created)
DT   23-JUN-2000 (Rel. 5.05, Last updated,
DT   Version 1)
DE   3'-end of L1 repeat (subfamily L1MC1) -
DE   a consensus.
KW   L1; Non-LTR Retrotransposon;
KW   interspersed Repeat; L1 (LINE) family;
KW   L1M4; L1MC1; L1MC1 subfamily; MER16;
KW   Repetitive sequence.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini;
OC   Hominidae; Homo.
RN   [1]
RP   868-1017
RA   Kaplan J.D., Jurka J., Solus F.J.,
RA   Duncan H.C.;
RT   "Medium reiteration frequency
RT   repetitive sequences in the human
RT   genome.";
RL   Nucleic Acids Res 17, 4731-4738 (1991).
RN   [2]
RP   1-1080
RA   Smit A.F, Toth G, Riggs D.A, Jurka J;
RT   "Ancestral, mammalian-wide subfamilies
RT   of LINE-1 repetitive sequences.";
RL   J. Mol. Biol 246, 401-417 (1995).
RN   [3]
RP   1-1080
RA   Smit A.F.;
RT   "Direct submission.";
RL   Direct Submission to Repbase Update
RL   (03-MAY-2000).
CC   [3] (Consensus)
CC   ORF2 ends at bp 675; average
CC   divergence
CC   of copies from consensus:15%.
SQ   Sequence 1080 BP; 395A; 196C; 229G;
SQ   259T; 1 other;


ID   L1ME2 repbase; DNA; PRI; 911 BP.
DT   01-OCT-1995 (Rel. 1.09, Created)
DT   23-JUN-2000 (Rel. 5.05, Last updated,
DT   Version 4)
DE   3'-end of L1 repeat (subfamily L1ME2)
DE   - a consensus.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat;
KW   L1 (LINE) family; L1M4; L1ME2; L1ME2
KW   subfamily; Repetitive sequence.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   1-911
RA   Smit A.F, Toth G, Riggs D.A, Jurka J;
RT   "Ancestral, mammalian-wide subfamilies
RT   of LINE-1 repetitive sequences.";
RL   J. Mol. Biol 246, 401-417 (1995).
RN   [2]
RP   1-911
RA   Smit A.F.;
RT   "Direct submission.";
RL   Direct Submission to Repbase Update
RL   (03-MAY-2000).
CC   [2] (Consensus)
CC   ORF2 ends at bp 675; average
CC   divergence of copies from consensus:
CC   22%.
SQ   Sequence 911 BP; 354A; 165C; 171G;
SQ   216T; 5 other;


ID   L1PA15 repbase; DNA; PRI; 912 BP.
DT   01-OCT-1995 (Rel. 1.09, Created)
DT   18-APR-1997 (Rel. 2.03, Last updated,
DT   Version 3)
DE   3'-end of L1 repeat (subfamily L1PA15)
DE   - a consensus.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat;
KW   L1 (LINE) family; L1P4; L1PA15; L1PA15
KW   subfamily; MER13; Repetitive sequence.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   696-885
RA   Kaplan J.D., Jurka J., Solus F.J.,
RA   Duncan H.C.;
RT   "Medium reiteration frequency
RT   repetitive sequences in the human
RT   genome.";
RL   Nucleic Acids Res 17, 4731-4738
RL   (1991).
RN   [2]
```

```
RP   1-912
RA   Smit A.F, Toth G, Riggs D.A, Jurka J;
RT   "Ancestral, mammalian-wide subfamilies
RT   of LINE-1 repetitive sequences.";
RL   J. Mol. Biol 246, 401-417 (1995).
RN   [3]
RP   1-912
RA   Smit A.F.;
RT   "Direct submission.";
RL   Direct Submission to Repbase Update
RL   (03-MAY-2000).
CC   [3] (Consensus)
CC   ORF2 ends at bp 684; average divergence
CC   of copies from consensus: 12%.
SQ   Sequence 912 BP; 364A; 191C; 175G;
SQ   182T; 0 other;


ID   L1PA16_5 repbase; DNA; PRI; 4083 BP.
DT   23-JAN-1998 (Rel. 3, Created)
DT   07-FEB-2000 (Rel. 5.01, Last updated,
DT   Version 2)
DE   Primate L1PA16_5 LINE1 repetitive
DE   element - a consensus.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat; L1 repeat; L180;
KW   L1PA16_5.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini;
OC   Hominidae; Homo.
RN   [1]
RP   1-1696
RA   Smit A.F.;
RT   "L1PA16_5.";
RL   Direct Submission to Repbase Update
RL   (1997).
RN   [2]
RP   1697-4083
RA   Jurka J.;
RT   "L1PA16_5.";
RL   Direct Submission to Repbase Update
RL   (JAN-2000).
CC   5' end of LINE elements with L1PA15-16
CC   subfamily 3' ends, comprising the 5'
CC   UTR and extending to position 3296
CC   relative to L1 sequence in this
CC   database. The UTR region from 714 to
CC   1238 contains 5 (a variable number of)
CC   113 bp tandem repeat units. -77%
CC   identity with L1.
SQ   Sequence 4083 BP; 1461A; 951C; 825G;
SQ   799T; 47 other;


ID   L1PA7_5 repbase; DNA ; PRI ; 1727 BP.
DT   23-JUN-2000 (Rel. 5.05, Created)
DT   23-JUN-2000 (Rel. 5.05, Last updated,
DT   Version 1)
DE   L1PA7_5 - a consensus.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat; L1P5A1; L1PA7_5;
KW   Repetitive sequence.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Primates; Catarrhini;
OC   Hominidae; Homo.
RN   [1]
RP   1-1727
RA   Kapitonov V.V., Jurka J.;
RT   "L1PA7_5.";
RL   Direct Submission to Repbase Update
RL   (1996).
CC   [1] (Consensus)
CC   L1PA7_5 is a consensus sequence for a
CC   subfamily of L1. This 1.7 kb consensus
CC   corresponds to the 1.4 kb 5' region of
CC   L1 from REPBASE.
CC   Average similarity of L1PA7_5
CC   sequences to the consensus is 0.93.
CC   The subfamily consensus contains
CC   multiple diagnostic positions and
CC   multiple long insertions of 180 bp
CC   (between positions 348 and 349 in L1
CC   from REPBASE), 176 bp (positions 586-
CC   587), 27 bp (380-381), 19 bp (469-
CC   470), 23 bp (759-760) and 25 bp (952-
CC   953).
SQ   Sequence 1727 BP; 486A; 462C; 472G;
SQ   302T; 5 other;


ID   L1PBB_5 repbase; DNA; PRI; 945 BP.
DT   23-JAN-1998 (Rel. 3, Created)
DT   23-JAN-1998 (Rel. 3, Last updated,
DT   Version 1)
DE   Primate L1PBB_5 LINE1 repetitive
DE   element - a consensus.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat; IN25; L1 repeat;
KW   L1PBB_5.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   1-847
RA   Jurka J., Kapitonov V.V.;
RT   "L1PBB_5.";
RL   Direct Submission to Repbase Update
RL   (1997).
RN   [2]
RP   1-945
RA   Smit A.F.;
RT   "L1PBB_5.";
RL   Direct Submission to Repbase Update
RL   (1997).
CC   Consensus of an insertion just
CC   upstream of ORF1 in L1PBa (at pos.
CC   1620-1621).
SQ   Sequence 945 BP; 281A; 258C; 204G;
SQ   180T; 22 other;


ID   L1PREC2 repbase; DNA; PRI; 8145 BP.
DT   31-OCT-2000 (Rel. 5.09, Created)
DT   28-JUL-2005 (Rel. 5.09, Last updated,
DT   Version 2)
DE   L1PREC2 is an ancient subfamily of L1
DE   - a consensus sequence.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat; L1 subfamily;
KW   L1P12_5; L1PA17_5; L1PREC1; L1PREC2;
KW   LINE1; ORF1; ORF2; endonuclease;
KW   reverse transcriptase.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
```

```
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   1062-4234
RA   Jurka J.;
RT   "L1PREC2.";
RL   Direct Submission to Repbase Update
RL   (OCT-2000).
RN   [2]
RP   1-8145
RA   Kapitonov V.V.;
RT   "L1PREC2.";
RL   Direct Submission to Repbase Update
RL   (OCT-2000).
CC   [2] (Consensus)
CC   It is a complete consensus sequence of
CC   the L1PREC2 subfamily of L1. Average
CC   divergence of L1PREC2 copies from the
CC   consensus sequence is 11%.
CC   A ~900-bp 3' tail of L1PREC2 is 96%
CC   identical to L1PA12, L1PA13 and
CC   L1PA14. A portion of L1PREC2, which
CC   starts at position 3420 is 87% and 90%
CC   identical with L1 and L1PREC1,
CC   respectively.
CC   The 5'-terminal 3419 bp long portion is
CC   ~90% identical to L1PA12_5 and L1PA17_5
CC   (including several long gaps).
FH   Key    Location/Qualifiers
FT   CDS    3011..4024
FT          /product="L1PREC2_1p"
FT          /note="RNA-binding protein"
SQ   Sequence 8145 BP; 2837A; 2134C; 1714G;
SQ   1460T; 0 other;


ID   L1Pt_5end repbase; DNA; PRI; 2137 BP.
DT   06-SEP-2005 (Rel. 10.08, Created)
DT   06-SEP-2005 (Rel. 10.08, Last updated,
DT   Version 1)
DE   L1 Non-LTR Retrotransposon from Pan
DE   troglodytes.
KW   L1; Non-LTR Retrotransposon;
KW   Interspersed Repeat; L1Pt_5end.
OS   Pan troglodytes
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Pan.
RN   [1]
RP   1-2137
RA   Smit A.F.;
RT   "L1Pt_5end - L1 Non-LTR Retrotransposon
RT   from Pan troglodytes.";
RL   Direct Submission to Repbase Update
RL   (06-SEP-2005).
CC   [1] (Consensus)
SQ   Sequence 2137 BP; 733A; 512C; 532G;
SQ   360T; 0 other;


ID   MEN repbase; DNA; ROD; 269 BP.
DT   06-SEP-2005 (Rel. 10.08, Created)
DT   19-JUN-2006 (Rel. 10.08, Last updated,
DT   Version 2)
DE   SINE2 SINE from Menetes.
KW   SINE2; SINE; Non-LTR Retrotransposon;
KW   Interspersed Repeat; MEN.
OS   Menetes berdmorei
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi; Mammalia;
OC   Eutheria; Rodentia; Sciurognathi;
OC   Sciuridae; Sciurinae; Menetes.
RN   [1]
RP   1-269
RA   Serdobova I.M., Kramerov D.A.;
RT   "Short retroposons of the B2
RT   superfamily: RT  evolution and
RT   application for the study of
RT   rodent phylogeny.";
RL   J Mol Evol 46(2), 202-214 (1998).
RN   [2]
RP   1-269
RA   Smit A.F.;
RT   "MEN - SINE2 SINE from Menetes.";
RL   Direct Submission to Repbase Update
RL   (06-SEP-2005).
CC   [2] (Consensus)
CC   Groundsquirrel.  Acc. Nos: X80312,
CC   X80313, X80314, Y09599,Y09600, Y09601.
SQ   Sequence 269 BP; 82A; 63C; 73G; 46T;
SQ   5 other;


ID   PB1 repbase; DNA; ROD; 127 BP.
DT   14-NOV-2005 (Rel. 10.11, Created)
DT   09-DEC-2005 (Rel. 10.11, Last updated,
DT   Version 1)
DE   SINE1 SINE from rodents.
KW   SINE1; SINE; Non-LTR Retrotransposon;
KW   Interspersed Repeat; ALU; PB1.
OS   Rodentia
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria.
RN   [1]
RP   1-127
RA   Smit A.F.;
RT   "PB1 - SINE1 SINE from rodents.";
RL   Direct Submission to Repbase Update
RL   (11-NOV-2005).
CC   [1] (Consensus)
SQ   Sequence 127 BP; 32A; 36C; 41G; 18T;
SQ   0 other;


ID   SVA repbase; DNA; PRI; 1640 BP.
AC   L09706;
DT   01-OCT-1995 (Rel. 1.09, Created)
DT   11-NOV-2005 (Rel. 2.03, Last updated,
DT   Version 3)
DE   Composite retroposon.
KW   SINE; Non-LTR Retrotransposon;
KW   Interspersed Repeat;
KW   Repetitive sequence; SINE-R; SVA.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata;
OC   Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RA   Ono M., Kawakami M., Takezawa T.;
RT   "A novel human nonretroviral
RT   retroposon derived from an
RT   endogenous retrovirus.";
RL   Nucleic Acids Res 15, 8725-8737
RL   (1987).
RN   [2]
RA   Shen L, Wu C.L, Sanlioglu S, Chen R,
RA   Mendoza R.A, Dangel W.A, Carroll C.M.,
RA   Zipf B.W., Yu Y.C.;
RT   "Structure and genetics of the
RT   partially duplicated gene RP
```

```
RT   located immediately upstream of the
RT   complement C4A and the C4B genes in
RT   the HLA class III region. Molecular
RT   cloning, exon-intron structure,
RT   composite retroposon, and breakpoint
RT   of gene duplication.";
RL   J. Biol. Chem 269, 8466-8476 (1994).
DR   GenBank; L09706; Positions 7919 6280.
CC   This element consists of fragments of
CC   two Alu elements and a fragment of an
CC   HERV-K LTR (LTR5) flanking a variable
CC   number of 40 bp tandem repeats.
SQ   Sequence 1640 BP; 286A; 539C; 497G;
SQ   318T; 0 other;


ID   TIGGER1 repbase; DNA; PRI; 2418 BP.
DT   01-MAY-1996 (Rel. 1.04, Created)
DT   18-APR-1997 (Rel. 2.03, Last updated,
DT   Version 2)
DE   Autonomous DNA transposon.
KW   Mariner; DNA transposon; Interspersed
KW   Repeat; MER37; Repetitive sequence;
KW   Tigger1.
OS   Homo sapiens
OC   Eukaryota; Metazoa; Chordata; Craniata;
OC   Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates;
OC   Catarrhini; Hominidae; Homo.
RN   [1]
RP   108-229
RA   Iris F, Bougueleret L, Prieur S,
RA   Caterina D, Primas G, Perrot V, Jurka
RA   J, Rodriguez-Tome P, Claverie J, Cohen
RA   D, Dausset J;
RT   "Dense Alu clustering and a potential
RT   new member of the NFkappaB family
RT   within a 90 kilobase HLA class III
RT   segment.";
RL   Nature Genet 3, 137-145 (1993).
RN   [2]
RP   1-550
RA   Lutfalla G., McInnis G.M., Uze G.;
RT   "Structure of the human CRFB4 gene:
RT   comparison with its IFNAR neighbour.";
RL   J. Mol. Evol 41, 338-348 (1995).
RN   [3]
RA   Smit A.F, Riggs D.A;
RT   "Tiggers and other DNA transposon
RT   fossils in the human genome.";
RL   Proc. Natl. Acad. Sci. USA 93, 1443-
RL   1448 (1996).
CC   [3] (Consensus)
CC   23 bp terminal inverted repeats, TA
CC   target site ORF1: bases 425 to 1789;
CC   ORF2: bases 1811 to 2206.
SQ   Sequence 2418 BP; 753A; 472C; 517G;
SQ   668T; 8 other;
```