Epigenetics of cell-free plasma DNA for non-invasive prenatal diagnosis of fetal aneuploidies

William Matthew Puszyk BSc (Hons)

Thesis submitted for the degree of Doctor of Philosophy

University of Warwick

Department of Biological Sciences

September 2008

Contents

Table of Contents

List of Figures

List of Tables

Acknowledgements

First of all I would like to thank my supervisor Dr. Robert Old for his support and patience, without both of which I would not have made it this far. I would also like to thank Paulina Hanson-Manful; her support throughout the project and during the writing of this thesis has been invaluable. Thanks also go to Dr. Francesco Crea, who taught me everything I know in the lab and has supported the project from its inception. Thanks also go to Dr. Rhagu Adya for providing much fun in the office and keeping the whole lab entertained for many years.

I would also like to thank everyone at Warwick Student Cinema, without their friendship this thesis would not have been possible.

I would also like to thank my advisory committee for sound advice and guidance given throughout the project.

Special thanks also go to my parents and grandparents for all of their love and support; without which I would not be where I am today.

This work was funded by the SAFE Network of Excellence;
reference number LSHB-CT-2004–503243.

Declaration

All of the results presented in this thesis were obtained by the author unless specified in the text.

All sources of information have been acknowledged by means of a reference.

None of the work contained within this thesis has been used for any previous application for a degree.

Abbreviations used

AIRE: Autoimmune regulatory element

APC: Anaphase promoting complex

AR: Allelic ratio

βhCG: Beta-subunit of human chorionic gonadotropin

BSPs: Bisulfite specific primers

BS-PCR: Bisulfite specific PCR

CEB: Candidate epigenetic biomarker

Cf DNA: Cell-free DNA

Cf RNA: Cell-free RNA

Cff DNA: Cell-free fetal DNA

Cff RNA: Cell-free fetal RNA

Cfm DNA: Cell-free maternal DNA

Cfp DNA: Cell-free plasma DNA

Cfp RNA: Cell-free plasma RNA

CGH: Comparative genomic hybridisation

CGI: CpG island

CHD: Congenital heart defect

COBRA: Combined bisulfite restriction analysis

CpG: Cytosine followed by Guanine on a linear sequence of DNA

CPM: Confined placental mosaicism

Ct: Cycle threshold

CTBs: Cytotrophoblasts

CVS: Chorionic villus sampling

DAPI: 4,6 diamino-2-phenylindole

DDH_2O: Double distilled H_2O

DNA: Deoxyribonucleic acid

DNMT: DNA methyltransferase

Dntp: Deoxyribonucleotide triphosphate

DOP-PCR: Degenerate-oligonucleotide PCR

DS: Downs syndrome

ECM: Extracellular matrix

EDTA: Ethylene diamine tetra acetic acid

ES cells: Embryonic stem cells

FNRBCs: Fetal nucleated red blood cells

FISH: Fluorescent *in situ* hybridisation

FITC: Fluorescein isothiocyanate

GNAS1: Guanine nucleotide binding protein 1
ge: Genomic equivalent

hCG: Human chorionic gonadotropin

HDACs: Histone deacetylase complexes

HMT: Histone methyltransferase

hPL: Human placental lactogen

ICM: Inner cell mass

LB medium: Luria Bertani medium

LRD: Limb reduction defect

M I: Meiosis I

M II: Meiosis II

MCA: Melt curve analysis

MeDIP: Methylated DNA immunoprecipitation

mRNA: Messenger RNA

ml: Millilitre

mM: Millimolar

MSPs: Methylation specific primers

MSRE analysis: Methylation-sensitive restriction endonuclease analysis

ND: Nondisjunction

ng: nanogram

NICHD: National Institute of Child Health and Human Development

NIPD: Non-invasive prenatal diagnosis

NT: Nuchal translucency

PAPP-A: Pregnancy-associated plasma protein A

PCR: Polymerase chain reaction

PLAC4: Placenta specific 4

RCD: Relative chromosome dosage

RNA: Ribonucleic acid

SAM: S-adenosylmethionine

SC: Synaptonemal complex

SLE: Systemic lupus erythematosus

SNP: Single nucleotide polymorphism

SSc: Systemic sclerosis

TBE: Tris borate EDTA

TDF: Testis determining factor

TfR: Transferrin receptor

TIMP: Tissue inhibitor of metalloproteinase

Tm: Melting temperature

TRD: Transcriptional repressor domain

TRITC: Tetra methylrhodamine isothiocyanate

TUNEL: Terminal UdTP nuclear end labelling

USPs: Unmethylation specific primers

µg: Microgram

µl: Microlitre

µM: Micromole

Thesis summary

Since the discovery of cell-free fetal DNA in the circulation of pregnant women fetal-specific DNA biomarkers for non-invasive prenatal diagnosis of fetal aneuploidy have been sought. A model system assessing the DNA methylation of placental DNA and adult peripheral leukocyte DNA has been developed previously to represent fetal and maternal plasma DNA.

To use DNA methylation to detect specific DNA molecules it is desirable that cell-free plasma DNA maintains the methylation profile of its tissue source. Using the imprinted gene *GNAS1*, a test has been developed to assess, for the first time the relative abundance of methylated and unmethylated DNA circulating in plasma. Methylated and unmethylated DNA sequences were found in equal abundance. If this finding is applicable to all plasma DNA sequences, we conclude that the steady-state concentration of DNA in methylated and unmethylated form is a reliable indicator of its input into the circulation.

We have also investigated whether the abundances of different single copy gene sequences in cell-free plasma DNA are equal. Using real-time PCR, the relative abundances of six unique genomic DNA sequences in plasma were assessed. We find that plasma DNA from different sequences is not present in equal abundance in normal healthy individuals. The relative abundance of sequences tested differed by as much as 12.3 fold.

We present a panel of novel candidate epigenetic biomarkers which have been identified using the model system. Of 366 DNA regions tested 3% were found to be differential. Further characterisation of these candidate epigenetic biomarkers has revealed limitations of the model system.

In view of these results, future epigenetic biomarker development should be achieved by a direct assessment of first trimester plasma DNA.

Introduction

Chapter 1

1.1 Overview of meiosis

1.1.1 Fetal aneuploidy risk factors

When assessing risk factors for fetal aneuploidy there are several important questions which must be addressed: 1) What are the underlying mechanisms of fetal aneuploidy? 2) What is the parental origin of the additional copy of the chromosome? 3) What are the health risks associated with aneuploid births to the infants born? (Hassold *et al*, 2007). No fewer than 5% of all pregnancies are aneuploid pregnancies, either trisomic or monosomic for a specific chromosome; most of which terminate *in utero* and are thought to be the leading cause of miscarriages (Hassold and Hunt, 2001). However, some aneuploidies are not sufficiently deleterious as to cause termination of the fetus *in utero* and may be carried to term. Trisomy 21 is one such aneuploidy, making fetal aneuploidy the leading cause of congenital birth defects (Hassold *et al*, 2007), with an incidence of 1-2/1000 births (Hook and Cross, 1983). For this reason the main focus of this research project has been on Chromosome 21 trisomy, however the strategic approaches of the work presented are applicable to other aneuploidies such as trisomy of Chromosome 18 Edwards syndrome, and trisomy of Chromosome 13 Patau Syndrome.

1.1.2 Meiotic errors in fetal aneuploidy

Errors during meiosis can cause aneuploidies, causing some gametes to gain an extra chromosome, and become trisomic. Meiosis is a specialised cell division process, which is essential for the propagation of all sexually reproducing organisms. During meiosis a diploid germ cell divides to generate daughter cells. Meiosis differs from mitosis in that the germ cell divides twice after a single DNA replication event. During meiosis I homologous chromosomes are separated into two daughter cells; the cells then proceed to meiosis II without an additional replication event. During meiosis II sister chromatids are separated into two daughter cells creating a total of four haploid gametes in males; and generating a single haploid gamete in females (in meiosis I half of the chromosomal material is packaged into a separate cellular structure known as a polar body, the same process occurs during meiosis II, giving a total of two polar bodies per oocyte created) (Morelli and Cohen, 2005).

Meiosis I is divided into four stages: prophase, metaphase, anaphase and telophase. Most of the events which define meiosis occur during prophase I. During prophase I a proteinaceous meiosis-specific structure called the synaptonemal complex (SC) forms (figure 1.1). The SC is composed of two lateral elements which form along the entire length of each sister chromatid and one central element which joins both lateral elements and binds them to the two homologous chromosomes (Morelli and Cohen, 2005).

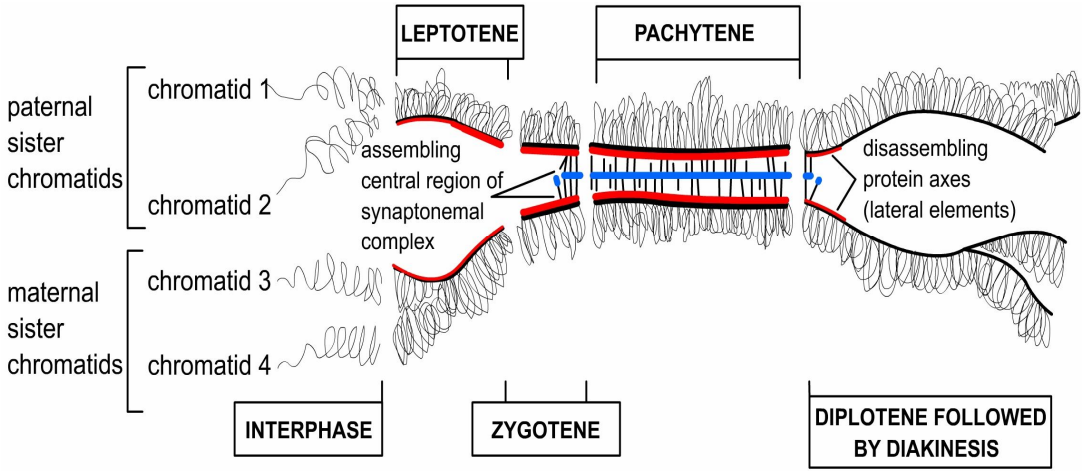


Figure 1.1: Chromosome synapsis during the different stages of meiotic prophase I
A single bivalent is shown. The pachytene stage is defined as the period during which the synaptonemal complex is fully formed. As meiosis progresses, the two homologs become tightly connected by proteins that form the central region of the synaptonemal complex; which is composed of a central element (blue), transverse filaments (thin black lines) and lateral elements (red) that anchor the chromatin loops. (Taken from Alberts, 2002, 4th Edn).

The SC facilitates recombination and enables chiasma to form between homologous chromatids. The formation of chiasma during prophase I is important as it increases genetic diversity by enabling recombination and also creates tension between the homologous chromosomes which is necessary for successful cell division during meiosis I (Carpenter, 1994). This is an important factor in preventing the generation of aneuploid gametes.

1.1.3 Nondisjunction during maternal meiosis I

90% of all cases of Chromosome 21 nondisjunction (ND) are due to meiotic errors, the majority of which occur during meiosis I (Freeman *et al*, 2007) ND is characterised as an error in cell division where the chromosomes fail to disjoin.

Meiosis is initiated at about 11-12 weeks of gestation, and after pairing, synapsis and recombination events, cells arrest until just prior to ovulation. This means that homologous chromosomes are arrested at prophase I for approximately 10-50 years, during a period which is termed dictyate arrest (Oliver *et al*, 2008). The formation of chiasma during prophase I (in addition to a role in recombination) functions to stabilise paired homologous chromosomes and enables their correct orientation on the meiotic spindle (Carpenter, 1994).

One cause of ND at meiosis I is a lack of chiasma formation between homologous chromosomes, this is termed achiasmate ND. Without the formation of chiasma to generate tension and ensure correct alignment on the meiotic spindle, chromosomes may segregate incorrectly during anaphase I. Another cause of ND at meiosis I occurs when chiasmata are formed during prophase I, this is termed chiasmate ND. Incorrect segregation in the case of chiasmate ND is not fully understood, but it is thought that either meiotic spindle error may be the cause, or that the position of the cross-over event may be important (Oliver *et al*, 2008). Research has shown that the number of observed aneuploidies increased in the oocytes of women under the age of 29, when the chiasmata are formed in the most distal 3.8 Mb of the long arm of Chromosome 21 (Oliver *et al*, 2008), confirming telomeric recombination as a risk factor for ND. Although ND during meiosis I is the largest risk factor for generation of aneuploid gametes, ND at meiosis II creates an additional barrier to successful gametogenesis (Oliver *et al*, 2008).

1.1.3.1 Precocious separation at meiosis I

Another form of meiotic error which occurs early during prophase I, is caused by the precocious separation of the centromere allowing the sister chromatids to separated prematurely (Angell *et al*, 1997). One study demonstrated that MI aneuploidies were commonly due to the addition or loss of a single chromatid rather than an entire chromosome (Angell *et al*, 1997). Another study of mouse oocytes demonstrated that at anaphase I sister chromatids (univalents) lying far from the equator are still included with bivalents and pass to one or other poles, before cell division (Hunt *et al*, 1995). MI metaphase chromosomes from the oocytes of young women are long and extended, but with increasing age can become contracted, it may be that the

failure of condensation proteins such as condensin, due to age related structural damage, may reduce the capacity for chromosomes to condense into a stable bivalent prior to division (Angell, 1997). Other potential factors, which may lead to precocious separation of sister chromatids could be age related structural damage of cohesins at the centromere, or even possibly the early action of the enzyme separase, which cleaves cohesins at the centromere (see section 1.1.5 page 6).

1.1.4 Nondisjunction during maternal meiosis II

ND events can also occur in meiosis II, these events can arise from defects in prophase II, where there is a failure to resolve chromosome crossovers before segregation into daughter cells. These events can be distinguished from meiosis I meiotic errors by detection of recombination-specific markers. In meiosis I the presence of a meiotic cross-over event occurring in the telomeric region of Chromosome 21 increases the risk of ND. However, during meiosis II the presence of a single meiotic cross-over event occurring closer to the pericentromeric region increases the risk of a ND event (Oliver *et al*, 2008). This may be due to the fact that during prophase II it is the sister chromatids which separate and the formation of crossovers closer to the centromere may impede this separation (figure 1.2).

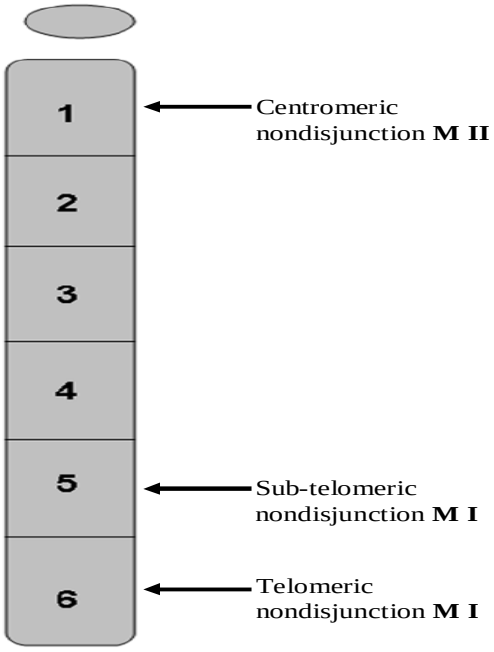


Figure 1.2: Ideogram of Chromosome 21q, with location of meiotic nondisjunction events
During MI most meiotic ND events occur in sub-telomeric and telomeric regions with an average of 41% of MI ND events occurring within interval 5 in all age groups and 41% of MI ND events occurring within interval 6 in women under the age of 29. During MII ND events occur predominantly in the centromeric region of Chromosome 21. In the combined group of women over the age of 29 the average rate of ND events occurring within interval 1 is 37.5%. (Adapted from Oliver *et al*, 2008).

One risk factor of fetal aneuploidy which has been identified is maternal age, which is part of the criteria used to determine which women should undergo prenatal screening during pregnancy.

1.1.5 Maternal age effect

There are multivariate factors which mean that with increasing maternal age there is an increased risk of fetal aneuploidy. Some of the risk factors and cellular events have been characterised, however there is still much unknown about all of the factors which contribute to the maternal age effect.

Degradation of meiotic proteins Bub1 and MAD2 is correlated with maternal age (Baker *et al*, 2004; Steuerwald *et al*, 2001). These proteins are part of the spindle assembly checkpoint pathway, which interacts with the kinetochore in order to delay anaphase by inhibition of the anaphase promoting complex (APC) in order to ensure correct chromosome segregation. With increasing maternal age, the probability of a single crossover occurring within the pericentromeric region in meiosis II also increases, which also increases the risk of ND (Oliver *et al*, 2008).

The chance of a ND event leading to a trisomy 21 birth increases with maternal age, or more accurately by increase in oocyte age. The likelihood that a woman under 30 who becomes pregnant with a Down syndrome (DS) fetus is less than 1 in 1,000. However, the chance increases to 1 in 400 for women who become pregnant at age 35. The likelihood of DS continues to increase as a woman ages, so that by age 42, the chance of a DS birth is 1 in 60 and by age 49, the chance is 1 in 12 (http://www.bgcdownsyndrome.org/about.cfm).

1.1.6 Paternal-specific nondisjunction

Meiosis in males in much more closely regulated than meiosis in females. During spermatogenesis there is no prolonged arrest during meiosis I, which reduces the risk of ND. However, ND in meiosis does occur and it is an additional paternal copy of Chromosome 21, which is responsible for 5-10% of trisomy 21 conceptions (Muller *et al*, 2000; Hook *et al*, 1995; Sherman *et al*, 1991; Yoon *et al*, 1996). Other factors which have also been suggested to play a role in the successful completion of

meiosis in males include: the lower temperature of the testicular environment (in relation to the rest of the body) and the presence of spermatogenic specific versions of heat shock proteins, which act as chaperones during spermatogenesis (Allen *et al*, 1996).

1.1.7 Trisomy 21 and congenital birth defects

Down syndrome (DS) was first described by Langdon Down in 1866 and linked to trisomy 21 in 1959 by Lejeune and Jacobs. It is the most common chromosomal abnormality causing learning difficulties and mental defects. There is no racial predilection, males and females are affected in the almost the same ratios (Byard *et al*, 2007). Mortality ratios provide evidence of excessive risk of leukaemia as well as other haematopoietic malignancies, as well as congenital defects resulting in heart failure. In addition, respiratory tract infections and pneumonia were also shown to be higher in the DS population compared to the general population (Scholl *et al*, 1982).

In a population-based study of 2,814 individuals with DS the overall incidence of cancer was comparable to that of the general population, but showed a different distribution of tumour types and malignancies. The incidence of solid tumours was less in all age groups by approximately half, whereas leukaemia occurred 10–20 times more frequently in individuals with DS (Hasel *et al*, 2000). The largest relative risk is observed in young children (under the age of 5 years), in whom acute myeloid leukaemia (AML) is 153 fold more frequent in children with DS compared with the general population and the incidence of acute megakaryoblastic leukaemia is 500 fold higher in children with DS than in the general population (Hasle *et al*, 2000; Hitzler, 2007; Zipursky *et al*, 1992).

It has also been shown that there is a 40-50% risk for congenital heart defects (CHDs) among children born with DS, whereas the risk to the population in general is about 1% (Frid *et al*, 1999). When offering diagnosis of DS it is important to remember that DS is also a potential indicator of these other conditions, which can be associated with a Downs birth.

1.2 Current prenatal diagnosis

Current prenatal diagnosis relies on a combination of techniques, which fall in to two broad categories: prenatal screening tests, which are non-invasive and are designed to detect those cases which have a higher probability of fetal abnormality, and prenatal diagnostic techniques, which are more accurate but usually more invasive. There are many forms of prenatal diagnosis, which are currently in use (see pages 8-12).

Amniocentesis is an invasive diagnostic technique and is the most widely accepted method for obtaining fetal cells for genetic analysis. After amniocentesis, chorionic villus sampling (CVS) is the most common form of invasive prenatal diagnosis. Both of these techniques are invasive and carry an associated risk of approximately 1% fetal loss (Brambati and Tului, 2005).

Non-invasive techniques are used as screening tools and lack the predictive power to form diagnostic tests individually. Ultrasound examinations are routinely used to diagnose congenital defects such as DS, and specific characteristics such as nuchal translucency can be measured to indicate risk. Second trimester serum screening is another non-invasive technique, which uses the measurement of fetal-specific proteins as an indicator for trisomy 21.

1.2.1 Invasive methods for prenatal diagnosis: amniocentesis and chorionic villus sampling

Women choose to have invasive prenatal diagnosis for a number of reasons including; having a child previously affected with a congenital defect, or aneuploidy, or the presence of markers for aneuploidy as indicated by an ultrasound test. On average approximately 5-10% of pregnant women in the U.K opt for some form of invasive prenatal diagnostic testing (Eddleman *et al*, 2006; Mujezinovic *et al*, 2007; Nicolaides *et al*, 2005 a). The most common invasive prenatal diagnostic techniques are amniocentesis and chorionic villus sampling (CVS). Amniocentesis is the drainage of amniotic fluid from the amniotic sac, in order to obtain fetal cells which are predominantly extravasted from the fetus itself, this is achieved by insertion of a

needle crossing the maternal abdomen and penetrating the amniotic sac (Wapner *et al*, 2005).

CVS is the collection of fetal cells from multiple tissue types which form the chorionic villus. Cell types within the chorionic villus include villus cells, syncytiotrophoblasts and mesodermal cells which make up fetal capillaries (Wapner *et al*, 2005).

Amniocentesis is regarded as the 'gold standard' of second trimester testing, however it can only be safely performed after 14 weeks of gestation, first trimester testing with amniocentesis has been linked with increased risk of fetal loss, amniotic sac fluid leakage and limb reduction defects (LRDs) such as talipe equinovarus (Wapner *et al*, 2005).

CVS can be used in the first trimester (approximately 10-12 weeks), and has been used clinically since the 1980s. However, similarly to amniocentesis, earlier testing has been linked to LRDs and increased risk of fetal loss (Wapner *et al*, 2005). Additional difficulties with CVS include the diagnostic errors caused by maternal cell contamination and confined placental mosaicism (CPM). Chorionic villus samples typically contain a mixture of fetal and maternal cells, although samples are thoroughly cleaned and separated under the microscope, some maternal cells may occasionally remain and grow in the subsequent cell culture. In some cases this may lead to the development of two cell lines, and in rare circumstances the maternal cell line may out-grow the fetal cell line (Boehm *et al*, 1993; Ledbetter *et al*, 1992; Wapner *et al*, 2005; Williams *et al*, 1987).

CPM is one source of potential diagnostic error for CVS. CPM can develop in either of two ways. One mechanism by which CPM occurs is when an aneuploid embryonic cell undergoes 'mitotic rescue', whereby the additional chromosome is expelled from the cell during mitosis. The other mechanism by which CPM occurs is known as post zygotic error. During the 32-64 cell stage of embryogenesis only 3-4 cells make up the inner cell mass (ICM) of the embryo, the rest of the cells become precursors for extra-embryonic tissues. If one of these precursors somehow becomes aneuploid by way of post zygotic error this can lead to CPM (Wolstenholme *et al*,

1996). Another form of CPM can occur when an embryonic cell undergoes 'mitotic rescue' and by chance expels the wrong copy of a trisomic chromosome, the cell can be left with two chromosomes from the same parent; this is known as uniparental disomy (Cassidy *et al*, 1992).

In one trial where the CVS test was performed on pregnancies between 8 to 9 weeks after conception, severe LRDs were reported in 5/289 cases (Firth *et al*, 1991). In another clinical trial, CVS and amniocentesis were performed during the first trimester. The associated risk of fetal loss was 7.6% for CVS and 7.0% for amniocentesis (Mujezinovic *et al*, 2007). A systematic review of all clinical trials conducted in relation to CVS and amniocentesis has established that with optimal conditions for each test; including experienced medical teams and laboratory staff as well as testing at the ideal gestational age during the second trimester; the associated risk of fetal loss was between 0.7-2.0% for CVS and 0.6-1.9% for amniocentesis (Mujezinovic *et al*, 2007).

The slightly higher risk in CVS is due to the slightly different gestational ages, with CVS testing being slightly earlier than amniocentesis. Overall the data suggests that these tests are optimal for late first trimester to second trimester testing, as earlier testing results in a greatly increased risk of fetal loss, which is unacceptable.

1.2.2 Non-Invasive methods for prenatal diagnosis

Current non-invasive methods of prenatal diagnosis involve the use of screening technologies and procedures in order to assess risk of fetal aneuploidy (usually trisomy 21). The judicious application of non-invasive techniques such as those based on ultrasound and the detection of serum biomarkers, greatly reduces the number of pregnancies which are subjected to testing by invasive methods.

1.2.2.1 Nuchal translucency

Nuchal translucency (NT) is an ultrasound technique, which is used to measure the fluid behind the neck of the fetus; the translucency is measured from the skin of the

fetus to the soft tissue overlying the cervical spine. NT changes throughout gestation and the optimal period for NT measurement is between 11 and 13 weeks (Nicolaides, 2005).

NT can only be measured if the fetus is within the normal size range; this is measured as fetal crown to rump length, with the acceptable range being between 45mm and 84mm. The lower limit is due to the technical aspects of sonography and deals with issues of image resolution, and the upper limit is in place to offer women earlier and safer forms of termination (Nicolaides, 2005). NT can be measured either by transabdominal or transvaginal sonography and the results achieved are similar (Nicolaides, 2005). One potential barrier to correct NT measurement occurs when the umbilical cord wraps around the neck of the fetus; this occurs in 5-10 % of all cases and can increase NT measurements by 0.8mm (Nicolaides *et al*, 2004).

Using NT measurements combined with maternal age, successful detection of trisomy 21 can be as high as 75-80% with a false positive rate of 5% (Avgidou *et al*, 2005; Nicolaides, 2004; Nicolaides *et al*, 2005 b). Another test based on ultrasonography is the measurement of fetal nasal bone, which is also measured in relation to the fetal crown to rump length. Although, nasal bone measurement alone is not an absolute marker for chromosomal abnormalities, as 2.2% of all Caucasian, 9.0% of all Afro-Caribbean and 5.0% of all Asian fetuses have no nasal bone but are chromosomally normal (Cicero *et al*, 2004). However, using the measurement of nasal bone in addition to other measurements such as NT can reduce the overall false positive rate of a screen (Filkins and Koos, 2005).

1.2.2.2 Maternal serum screening

Maternal serum screening for fetal biomarkers is another non-invasive method for prenatal diagnosis. Screening for Down syndrome (DS) by means of maternal serum analytes in the second trimester was first described in 1988 (Wald *et al*, 1988). This first small study included 17 cases of DS and 74 unaffected cases, in the DS population (65%) were found to have a human gonadotropin (hCG) level greater than 20IU/ml, whereas only 1 of the unaffected population (1.4%) exhibited the

same level (Wald *et al*, 1988). As serum testing continued to develop another marker was discovered which, was termed pregnancy-associated plasma protein A (PAPP-A) and the measurement of human gonadotropin was refined to detect the β subunit of human gonadotropin (βhCG). Optimal times for gestational testing have since been established, and it has been shown that measurement of PAPP-A is optimal between 8-11 weeks (Cuckle and Van Lith, 1999).

1.2.2.3 Combined non-invasive testing

Combining NT measurement and serum biochemical testing has often been used as a way of increasing trisomy detection rates, whilst also reducing the false positive rate. One strategy described (Wright *et al*, 2006) has been successfully developed to give a high detection rate of 90% for trisomy 21 and a low false positive rate of 2.0%. This strategy is called the three stage contingent test. The first stage of the test is biochemical testing of PAPP-A and βhCG at 10 weeks. Cut off values for PAPP-A and βhCG levels determine if women need to proceed to the next stage which is NT measurement. NT measurement is offered to women at 12 weeks but only if the cut-off levels are exceeded for the first stage test, or if there is an additional risk factor such as higher maternal age. Only women who exceed the cut-off value for NT testing proceed to the third and final test. The third stage is quadruple serum biomarker testing; markers included are α-fetoprotein, βhCG, unconjugated estriol and inhibin-A. Information on maternal age and the three tests are combined to assess risk (figure 1.3) (Wright *et al*, 2006).

This test is regarded as being one of the best, and has a high detection rate for the low false positive rate of 2%. However, this means that potentially women are subjected to three assessments before a diagnosis can be made, and even after these tests women may still be referred for invasive prenatal testing such as amniocentesis and CVS. A more efficient method for non-invasive prenatal diagnosis is sought, such a test would need to provide a much more direct method of fetal assessment and also eliminate the requirement for invasive prenatal testing.

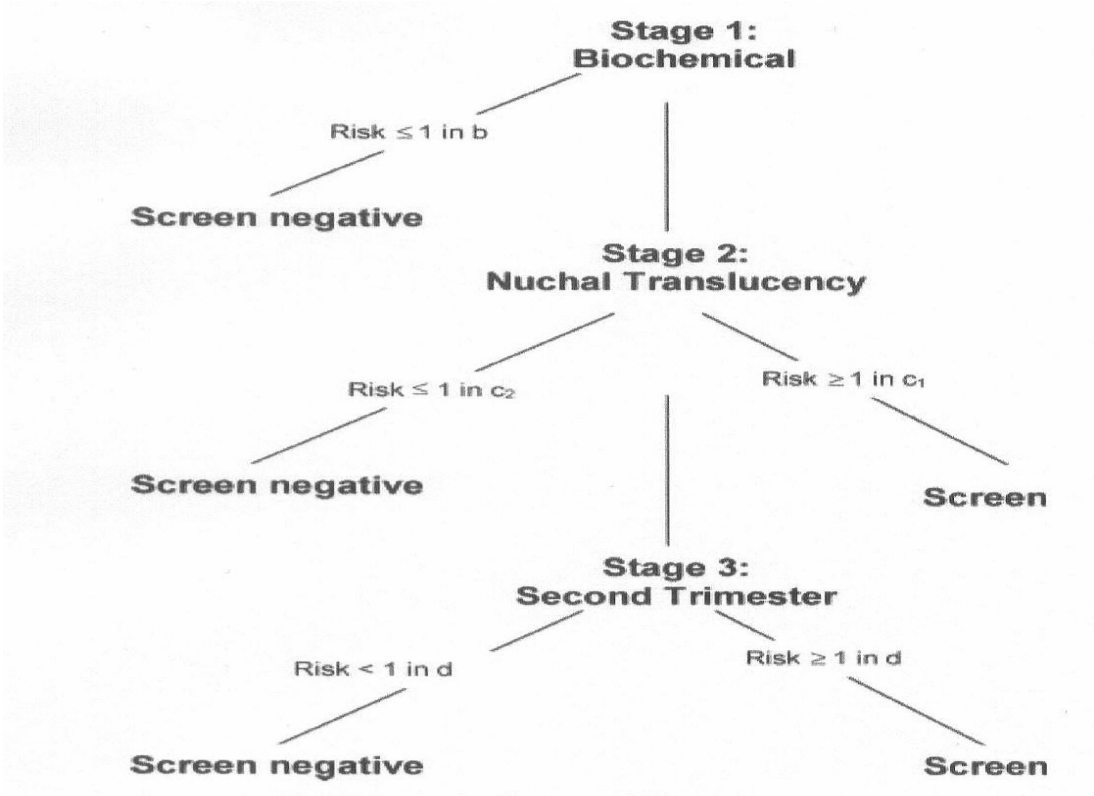


Figure 1.3: Schematic showing the three stage contingency policy
At stage one PAPP-A and β-hCG are measured at approximately 10 weeks, the first stage risk cut off 1/b determines the proportions of pregnancies that go on to have nuchal translucency (NT) measurements in the second stage. NT is measured at approximately 11 weeks. The cut off value $1/c_1$ determines the early detection and early false positive rate, while the lower cut off value $1/c_2$ determines the proportion of women who continue to the third stage, and undergo second trimester-testing with biochemical biomarkers. The quadruple biochemical biomarker test measures amounts of: α-fetoprotein, β-hCG, unconjugated estriol and inhibin-A. On the basis of this risk, the pregnancy is screened as negative or positive according to the final cut off value 1/d. (Taken from Wright *et al*, 2006).

1.3 Fetal cells

The first paper to document the description of feto-maternal cellular trafficking, was published in 1893 by Georg Schmorl, the title of the paper translates as 'Pathological and anatomical examinations of puerperal-eclampsia' (Schmorl G, 1893). Schmorl described the results obtained from the examination of 17 deceased pregnant women, in relation to the liver as this was the original focus of his work. In autopsied lungs Schmorl observed liver cells which make up part of the thrombi and other peculiar cells, which were observed in the capillaries of the lung (Lapaire *et al*, 2007).

The cells were described as ‘giant’ and contained between 6-15 nuclei with rounded nucleoli. These cells either originated from the bone marrow or the placenta; and since none of the cases exhibited internal bleeding, it seemed unlikely that the cells originated from bone marrow (Lapaire *et al*, 2007). Schmorl postulated that these multinuclear cells were originally from the surface of the villi, and several observations supported his hypothesis:

1) The morphological pattern corresponded to that of placental giant cells
2) In all of the cases the giant cells were detectable in all parts of the body, including the uterine veins
3) In one case cubic cells with the same morphological features as villi were found in association with the ‘giant’ cells

Schmorl was the first physician to describe the presence of fetal cells in the maternal body, and since that discovery the placenta has been studied by scientists around the world. The presence of fetal cells in the maternal circulation offers a prime opportunity for the development of non-invasive prenatal diagnosis, if it is possible to isolate these cells from the maternal blood.

1.3.1 Placental development

Development of mammals depends on the formation of a placenta, this process involves uterine attachment and acquisition of the maternal blood supply. This connection to the mother allows gaseous exchange, supply of nutrients to the fetus and the removal of fetal waste products.

Cytotrophoblasts (CTBs) are specialised epithelial cells, which function to aggressively invade maternal tissues (Fisher *et al*, 1989; Librach *et al*, 1991; Weier *et al*, 2005). During early pregnancy CTBs invade the maternal uterine wall and associated arterial network, this process takes place during the first trimester and is completed rapidly thereafter. CTBs have been found to express several gelatine degrading proteinases, the activities of which are inhibited by the use of tissue inhibitor of metalloproteinases (TIMPs). This evidence suggests that CTBs are

specifically regulated to produce metalloproteinases in order to aid trophoblast invasion of the uterus (Fisher *et al*, 1989). CTBs have also been shown to invade and migrate through extracellular matrix (ECM) *in vitro* (Glass *et al*, 1983).

Many elements of trophoblast invasion are similar to events which occur during tumour cell invasion. After a brief adherent stage CTBs penetrate the basement membrane of uterine epithelial cells and invade the stroma and its associated arterioles. However, unlike tumour invasion, CTB invasion is regulated and confined spatially and temporally, to placental development in early pregnancy (Librach *et al*, 1991). CTBs up-regulate the production of type IV collagenase during the first trimester, indicating a role in maternal invasion (Fisher *et al*, 1989; Librach *et al*, 1991).

CTBs arise by differentiation of a progenitor population that is anchored to the trophoblast basement membrane surrounding the mesochymal cores of the chorionic villi. CTBs at the anchoring chorionic villi leave the trophoblast basement membrane and form columns of non-polarised cells, which attach to and penetrate the uterine wall (figure 1.4). As well as allowing the development of the essential route for nutrient and gaseous exchange, the formation of the placenta also allows fetal cells to enter the maternal circulation (Weier *et al*, 2005).

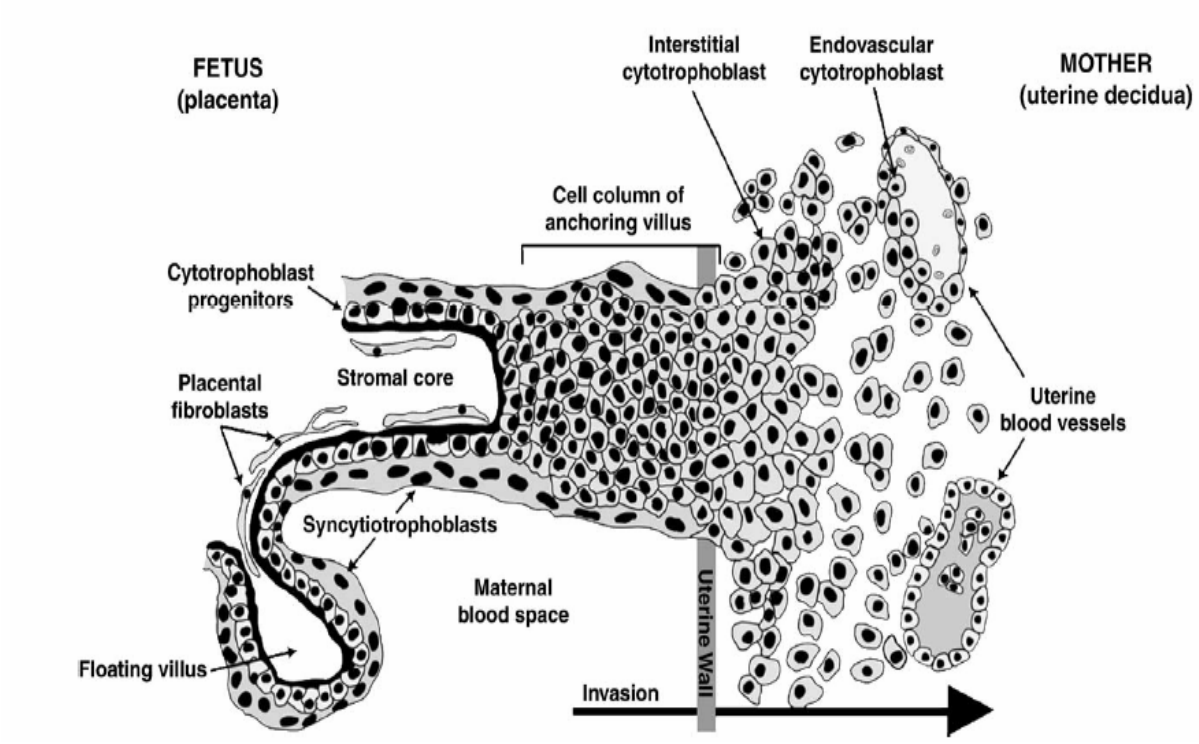


Figure 1.4: Placenta invasion at the fetal-maternal interface at approximately 10 weeks of gestation
The anchoring villi, which function as a bridge between the fetal and maternal compartments, form cell columns that give rise to the subpopulation of cytotrophoblasts (CTBs) that invade the uterine interstitium and maternal vasculature, thereby anchoring the fetus to the mother and accessing the maternal circulation. (Taken from Weier *et al*, 2005).

1.3.2 Fetal cells in the mother

Transfusion of cells from the fetus to the mother may be the cause of some 'auto-immune diseases'; most notably systemic sclerosis (SSc). Cells determined to be of fetal origin by fluorescent *in situ* hybridisation (FISH) analysis, by the detection of a Y chromosome, have been observed in the tissue of women with SSc in at least tissue type (Johnson *et al*, 2001). Fetal cells are predominantly found in the spleen, as it filters the contents of the peripheral blood (Johnson *et al*, 2001). The detection of fetal cells in the spleen indicates the presence of fetal cells in the maternal circulation, and highlights their potential use in non-invasive prenatal diagnosis (NIPD).

Many other studies have also documented cases of SSc, not only in cases of pregnancy, SSc can also been linked with blood transfusion, with cells detected in the skin, lungs, and lymph nodes. The observation of fetal nucleated red blood cells (FNRBCs) in skin lesions in pregnant women with SSc indicates that fetal cells are able to migrate and survive in the maternal circulation (Arlett *et al*, 1998).

1.3.2.1 Fetal nucleated red blood cells

In addition to cytotrophoblasts, FNRBCs are also present in the maternal circulation (Bianchi *et al*, 1990; Chen *et al*, 2004; Elias *et al*, 1996; Krabchi *et al*, 2006; Wachi and Kitagawa 2004; and many others). FNRBCs were first detected in the maternal circulation in 1990 (Bianchi *et al*, 1990).

Using a monoclonal antibody for the CD71 cell surface receptor, transferrin (TfR), candidate fetal cells were identified from the peripheral blood of pregnant women. Y chromosome specific sequences were then detected by PCR, this was followed with karyotyping analysis. In 7/19 cases fetal cells were detected in the blood of pregnant women, with cells from 6/7 cases being confirmed as fetal by Y chromosome sequence detection (Bianchi *et al*, 1990).

One issue with the use of FNRBCs for NIPD is the paucity of these cells in the maternal circulation; on average there are approximately 1.9-6 cells detectable per ml of maternal blood, depending on which study is observed (Bianchi *et al*, 1990; Chen *et al*, 2004; Kolvraa *et al*, 2005; Krabchi *et al*, 2001; Guetta *et al*, 2003).

1.3.2.1.1 Fetal cell enrichment techniques

The isolation and enrichment of fetal cells in the maternal circulation is a goal that has been pursued by many research groups for many years, for the development of non-invasive prenatal diagnosis (NIPD). If sufficient cells can be isolated and recovered from the maternal blood they could be used directly in tests for fetal aneuploidy disorders or other genetically inherited diseases. FNRBCs are an attractive candidate for enrichment, as 10% of fetal blood cells are nucleated red blood cells and these cells are rare in adults (Elias *et al*, 1996).

Developing markers for the enrichment of FNRBCs has long been a goal in NIPD, the first such marker was the transferrin receptor (Bianchi *et al*, 1990; Elias *et al*, 1996). Another technique developed relied on slides stained with linked α galactose residues, to bind to lectin receptors on fetal cells. This enabled the enrichment of fetal blood cells before further sorting by filtration. The enrichment achieved was 4.7-10.9 FNRBCs per ml of maternal blood (Wachi and Kitagawa 2004).

Other techniques have not been so efficient. The use of CD34+ progenitor cells was also targeted as a candidate for fetal cell enrichment (Guetta *et al*, 2003). Under the correct conditions CD34+ cells proliferate in culture and are more abundant in the fetal circulation than the maternal circulation (Guetta *et al*, 2003). However this technique yielded on average 1 fetal cell/ml of maternal blood, and so did not give the level of enrichment required for NIPD.

Another study reported the utility of an erythroblast scoring system, as a method for the detection of FNRBCs and their use in detecting fetal aneuploidies. FNRBCs were isolated by density gradient separation, followed by CD15 and CD45 depletion, and γ haemoglobin positive selection.

Each cell was then designated a score ranging from 0-12, with cells scoring 5 or less deemed to be of maternal origin (94.3% accurate) and cells scoring 9 or more deemed to be of fetal origin (86.8% accurate). The method did not improve the enrichment of fetal cells, but used blood samples of ample volume so as to give enough cells for analysis (Cha *et al*, 2005). Overall the technique was labour intensive and costly, requiring the use of expensive cell sorting equipment. In addition this method offered no additional benefits for prenatal diagnosis above that of the three stage contingency testing currently available (Wright *et al*, 2006).

In addition to the difficulties of obtaining sufficient FNRBCs from the maternal circulation, some studies have suggested that FNRBCs are not the optimal targets for use in NIPD (Kolvraa *et al*, 2005). A large multi-centre collaborative study funded by the National Institute of Child Health and Human development (NICHD) has concluded; and results from the first part of the study have been published (Bianchi

et al, 2002). The study was named NIFTY and the objectives were to identify optimal methods for fetal cell enrichment from maternal blood and to demonstrate the utility of these cells for the determination of fetal sex and fetal aneuploidy, in comparison with the standard techniques of CVS and amniocentesis (Bianchi *et al*, 2002). Isolated fetal cells were assessed by FISH analysis on interphase nuclei of cells. After the first 5 years of the study 2744 maternal blood samples were analysed, making this study one of the largest studies of fetal cells so far (Bianchi *et al*, 2002). Of 3302 blood samples available for analysis approximately 17% failed to yield any fetal cells for analysis (Bianchi *et al*, 2002). Of the remaining samples available for analysis the detection rate for aneuploidies was 74.4% which is lower than the current detection rates of 90% for the three stage contingency test (Wright *et al*, 2006).

Further studies have shown that the nuclei of FNRBCs disintegrate some time before they are expelled from the cell; this makes them an unsuitable target for FISH analysis and an unattractive prospect for NIPD (Babochkina *et al*, 2005). Other fetal cells have also been detected in the maternal circulation; however these cells have been shown to persist in the maternal circulation for years and even decades, thus making them unsuitable for NIPD (Bianchi *et al*, 1996; Guetta *et al*, 2003).

1.3.2.1.2 Acquisition of aneuploidy in cytotrophoblasts

As well as the issues effecting the enrichment and use of FNRBCs, CTBs are also unsuitable for use in NIPD. During the process of differentiation into the invasive phenotype CTBs acquire aneuploidies, which can lead to the cells becoming either hyperdiploid or hypodiploid (figure 1.5) (Weier *et al*, 2005). This differentiation is thought to enable increased expression of the proteinases required for invasion and possibly as a mechanism for ensuring a short life span (Weier *et al*, 2005).

There are new technologies in development, for the isolation and enrichment of fetal cells from maternal blood; such as a cell-sorting biochip (Biotech international USA); and other techniques which rely on the use of physiological cell properties to culture FNRBCs. However the prospect of developing NIPD using fetal cells in the

maternal circulation is unlikely at present. Diagnostics which make use of the cell-free nucleic acids released by fetal cells are now sought after.

1.4 Presence of cell-free fetal nucleic acids in the maternal circulation

Since the discovery of fetal cells in the maternal circulation it has been hypothesised that fetal nucleic acids may well also circulate in the maternal peripheral blood. However what is not certain is whether or not it is possible to utilise them in clinical practice for NIPD. The presence of cell-free fetal DNA (cff DNA) in the maternal circulation was first discovered, by identification of a Y chromosome-specific sequence *SRY* (Lo *et al*, 1997). This was followed later by the detection of cell-free fetal RNA (cff RNA) in the maternal circulation (Poon *et al*, 2000), with the detection of *ZFY*-specific mRNA sequences, which relate to a zinc finger protein which is present on the Y chromosome (Poon *et al*, 2000).

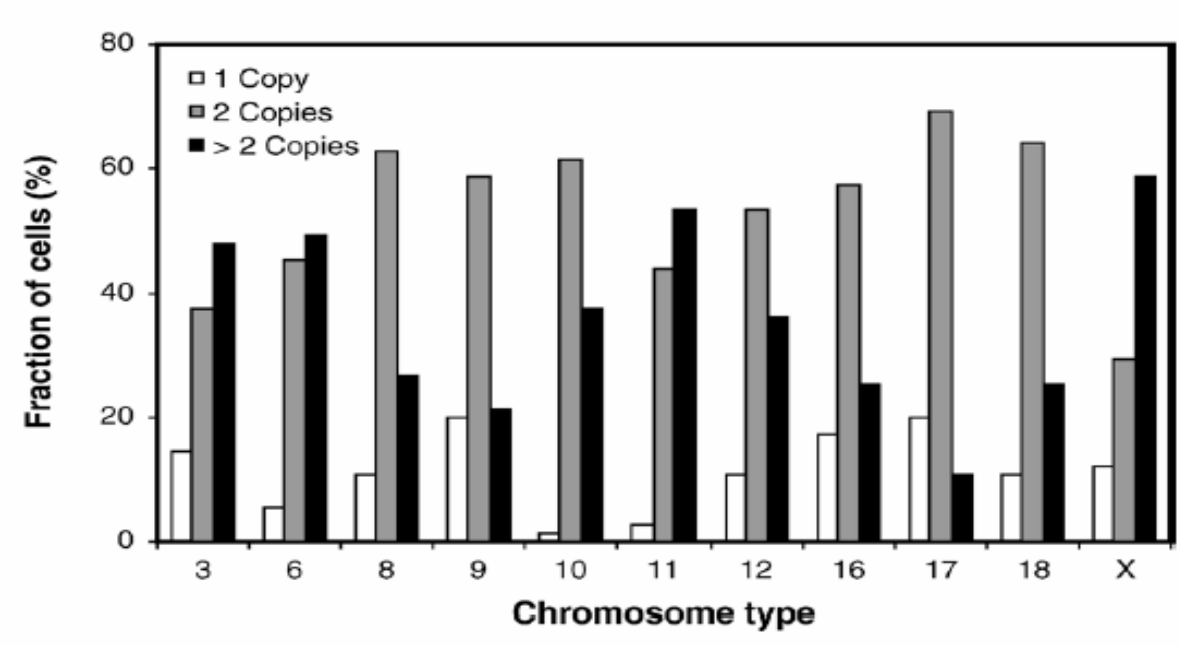


Figure 1.5: Common aneuploidies detected in a cytotrophoblast population
The analysis of a population of 75 interphase cytotrophoblasts taken from a second trimester placenta shows that aneuploidies are common in CTBs. Monosomies are shown in white, normal disomies are shown in grey and trisomies and other gains are shown in black, chromosomes 3, 6, 8, 9, 10, 11, 12, 16, 17, 18 and X are assessed. (Taken from Weier *et al*, 2005).

1.4.1 Cell-free fetal RNA in the maternal circulation

Cell-free RNA (cf RNA) was first discovered in the plasma and serum of cancer patients, it was proposed that cf RNA could be used as a diagnostic marker for tumour/cancer outcome in patients after chemotherapy, as in the case examples of malignant melanoma and nasopharyngeal carcinoma, respectively (Kopreski *et al*, 1999; Lo *et al*, 1999). Since the discovery of cf RNA in plasma and serum of cancer patients, the detection and quantification of cell-free fetal RNA (cff RNA) has been sought for the development of non-invasive prenatal diagnosis (NIPD).

Cff RNA was first detected in the plasma of pregnant women, by use of the first fetal-specific RNA marker *ZFY*; a gene encoding transcripts for a zinc finger protein. *ZFY* is on the Y chromosome and therefore a male specific gene, and can only be detected in male bearing pregnancies (Poon *et al*, 2000). The existence of cell-free plasma RNA (cfp RNA) is remarkable given the highly unstable nature of RNA and the fact that ribonuclease is present in the circulation (Reddi *et al*, 1976; Tsui *et al*, 2002). One study (Tsui *et al*, 2002), investigated the stability of endogenous cfp RNA extracted from plasma in comparison with commercially available human RNA. Unfiltered blood stored at room temperature exhibited the greatest difference over three time points of 0, 6 and 24 hours after collection, with total mRNA levels increasing over time; blood stored at 4 ^{0}C did not vary significantly in its plasma RNA levels. Non-particle-associated RNA in plasma was assessed by filtering plasma before extraction. Non-particle-associate RNA levels did not vary significantly in either blood stored at room temperature or at 4 ^{0} C. The concentration of mRNA was found to be highly stable in samples stored in ethylene diamine tetra acetic acid (EDTA) (figure 1.6). Commercially available RNA was also added to plasma samples and assessed in a time course experiment, after incubation with the plasma for 5, 10 and 15 seconds, Trizol LS reagent was added (Invitrogen) to stop any reaction with ribonuclease in the plasma (figure 1.7). Exogenous RNA was presumed to be rapidly broken down by ribonuclease present in the plasma, demonstrating that cfp RNA is somehow protected from ribonuclease action. The stability of cfp RNA indicates that it may be of diagnostic use in NIPD (Tsui *et al*, 2002).

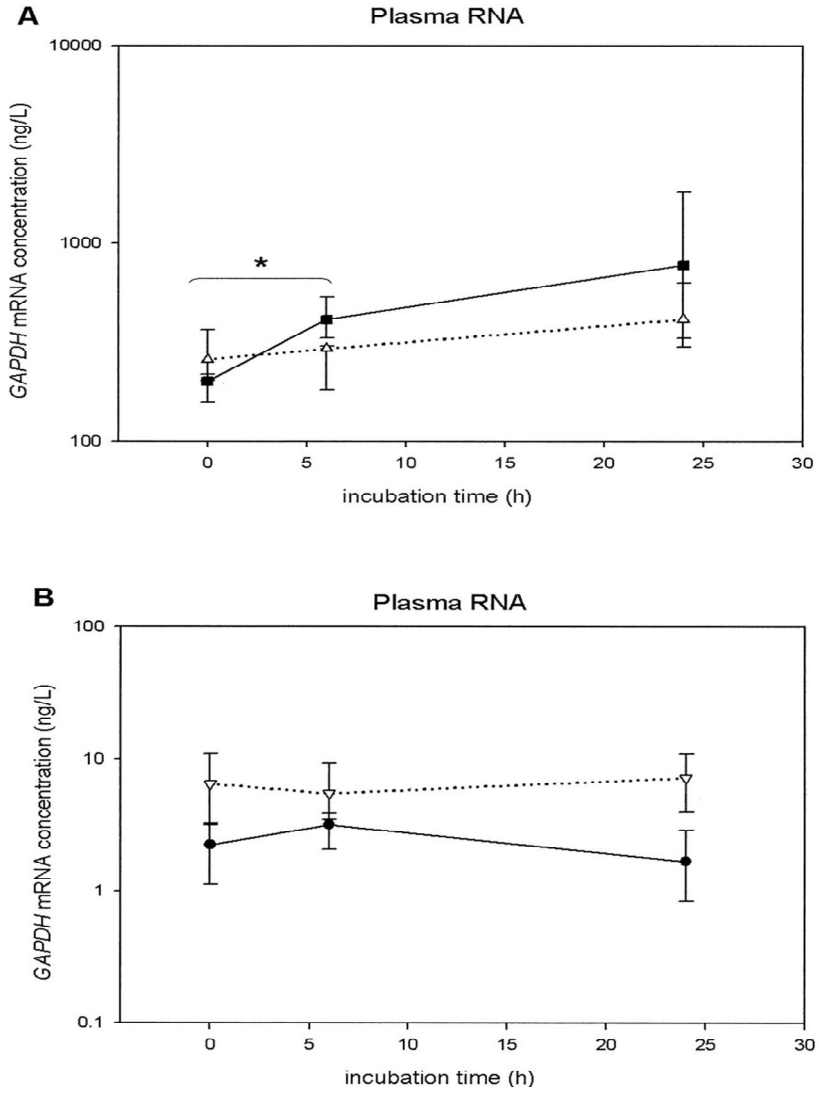


Figure 1.6: Effect of delayed processing of EDTA-stored blood on RNA plasma concentration

(A) Shows unfiltered plasma RNA concentrations as measured at 0, 6 and 24 hours after blood collection. All samples were stored with EDTA before processing at either room temperature (black line) and at 4 ^{0}C (dotted line). (B) Shows filtered RNA in plasma as measured at intervals of 0, 6 and 24 hours after blood collection. All samples were stored with EDTA before processing at either room temperature (black line) or at 4 ^{0}C (dotted line). (Taken from Tsui *et al*, 2002).

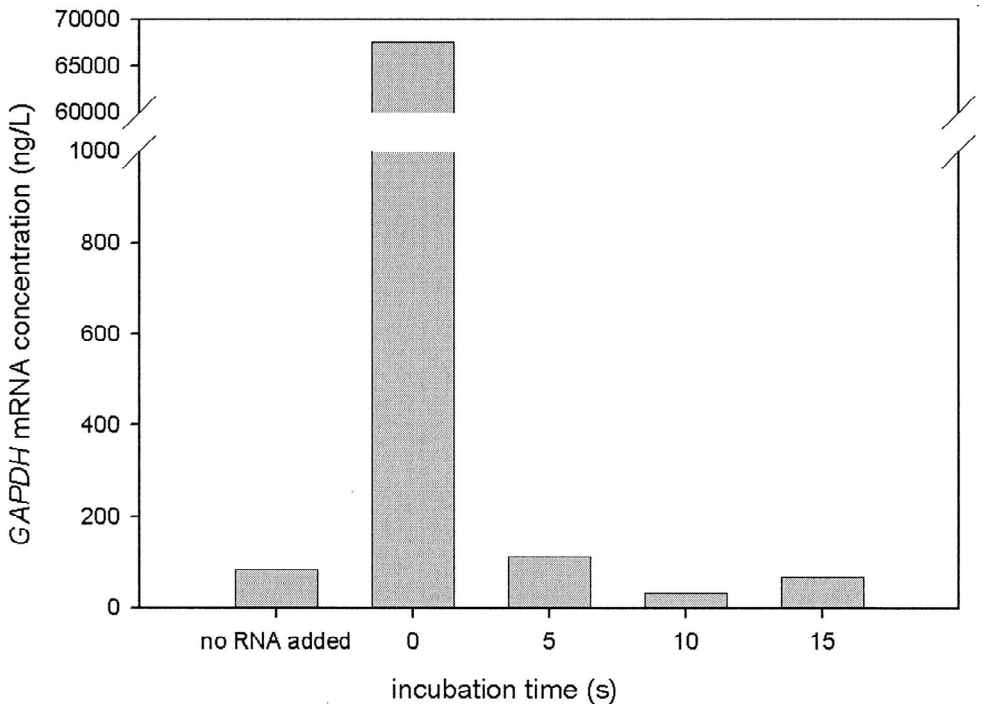


Figure 1.7: Time course for clearance of exogenous free-RNA from plasma
Exogenous free-RNA (as measured by *GAPDH* mRNA levels) is cleared rapidly from plasma, with levels returning to normal after 5 seconds; this is presumably due to the action of ribonuclease in the plasma. The small but finite levels of mRNA detected are assumed to be protected from ribonuclease action, by association with sub-cellular particles. (Taken from Tsui *et al*, 2002).

Cfp RNA has been shown in other studies to be protected by apoptotic bodies and other smaller sub-cellular particles (Hassleman *et al*, 2001; Ng *et al*, 2002). The presence of RNA in apoptotic bodies has been reported, fluorescence microscopy has shown the presence of typically granulated structures, representing apoptotic bodies containing RNA (Hassleman *et al*, 2001). A more advanced study showed the detection of cell-free mRNA in plasma after filtration with filters of various pore sizes of 5, 0.45 and 0.22 μm before quantitative analysis, showing that there is a stable background amount of RNA present in the plasma, which is associated with small sub-cellular particles.

The next incremental discovery was the detection of cff RNA of placental origin in the maternal circulation (Ng *et al*, 2003). It has been determined that the placenta is the major source of cff RNA in the maternal circulation, and that cff RNA can be

used for NIPD (Ng *et al*, 2003). This was shown by the detection and measurement of two placentally expressed mRNAs throughout gestation, in a cohort of pregnant women. The mRNAs measured were derived from the placental genes human placental lactogen (hPL) and the beta subunit of human chorionic gonadotropin (βhCG). The measurement of each cff RNA marker correlated to plasma protein levels of each gene product (figure 1.8). In addition *hPL* mRNA was shown to clear rapidly after delivery, and by 24 hours post delivery was undetectable (Ng *et al*, 2003). This further supports the utility of cff RNA as a useful resource in developing NIPD.

Further characterisation studies have shown that circulating cfp RNA has a propensity toward being fragmented, with a bias toward a higher absolute quantity of 5' end mRNA detected in the plasma. One study has shown that by designing primers which span mRNA transcripts, at the 5' end the middle and the 3' end that not all regions of the same transcript are equally abundant in plasma (Wong *et al*, 2005). The measurement of several transcripts with primers spanning the different amplicon regions, detected a higher absolute quantity for 5' end fragments. In each of the transcripts assessed the absolute amount of 5' ended molecules as greater than that of the 3' ended molecules; this is an important factor when choosing primers to assess transcripts for NIPD (Wong *et al*, 2005).

1.4.2 Cell-free fetal RNA markers on Chromosome 21

The first cff RNA marker discovered on Chromosome 21 is *PLAC4* mRNA, as a placentally expressed message it is ideal as a marker as it is not contaminated by background maternal nucleic acids (Lo *et al*, 2007). The *PLAC4* marker was discovered by screening mRNA obtained from pregnant and non-pregnant women by microarray analysis (Tsui *et al*, 2004) *PLAC4* is readily detectable in plasma of pregnant women, and undetectable in non-pregnant individuals, it is also cleared rapidly after delivery, with no sequences detected 24 hours after delivery (figure 1.9).

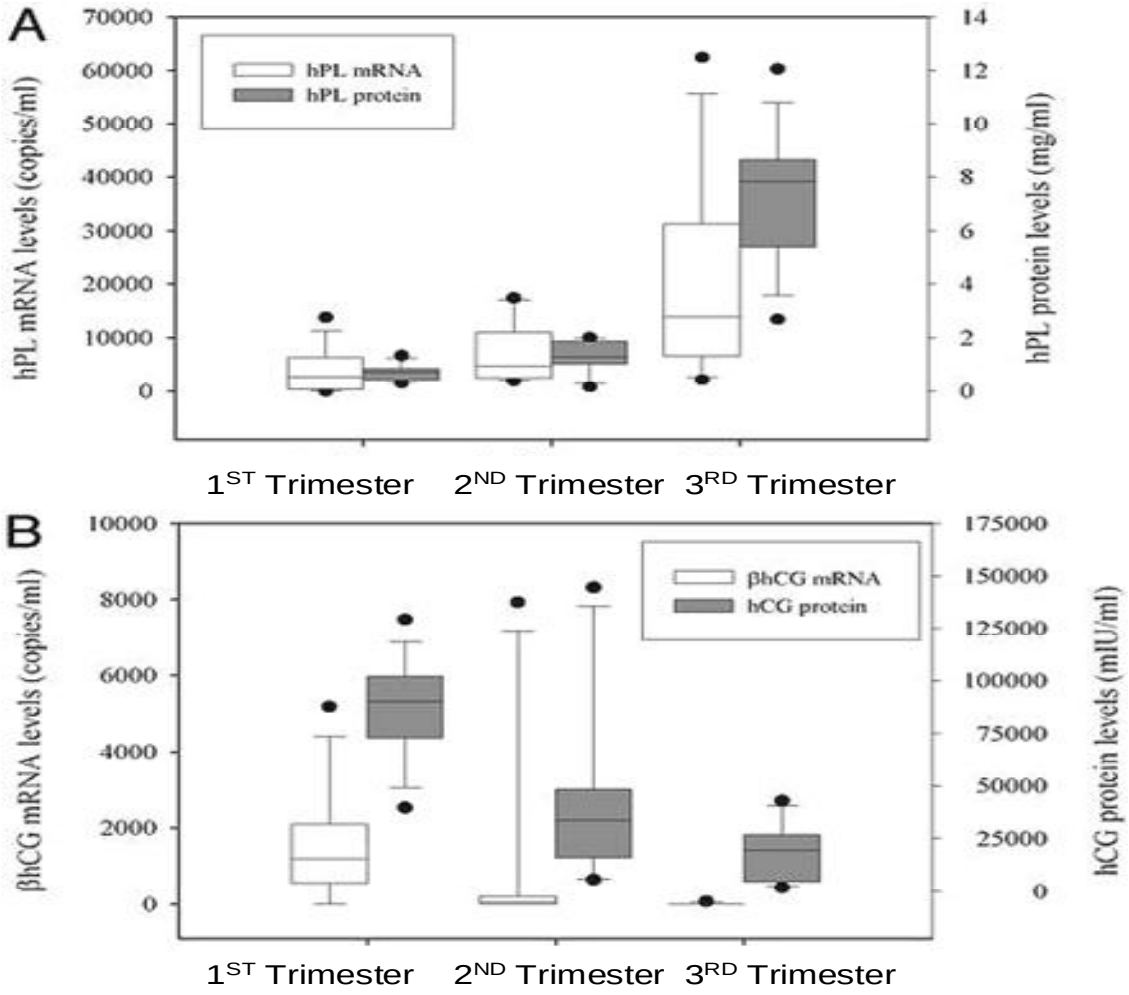


Figure 1.8: Levels of placenta derived mRNA and proteins in normal pregnancy
(A) *hPL* mRNA and hPL protein levels in the maternal serum at all three trimesters correlate with each other increasing throughout gestation, *hPL* mRNA data are shown by clear boxes and hPL protein data are shown by grey boxes. (B) *βhCG* mRNA and βhCG protein levels also correlate with each other, decreasing throughout gestation. *βhCG* mRNA data are shown by clear boxes and βhCG protein data are shown by grey boxes. The lines inside each box represent median values, the boxes mark the interval between the 25th and 75th percentiles, the whiskers denote the intervals between the 10th and 90th percentiles, and filled circles mark the data points outside of the 10th and 90th percentiles. (Taken from Ng *et al*, 2003).

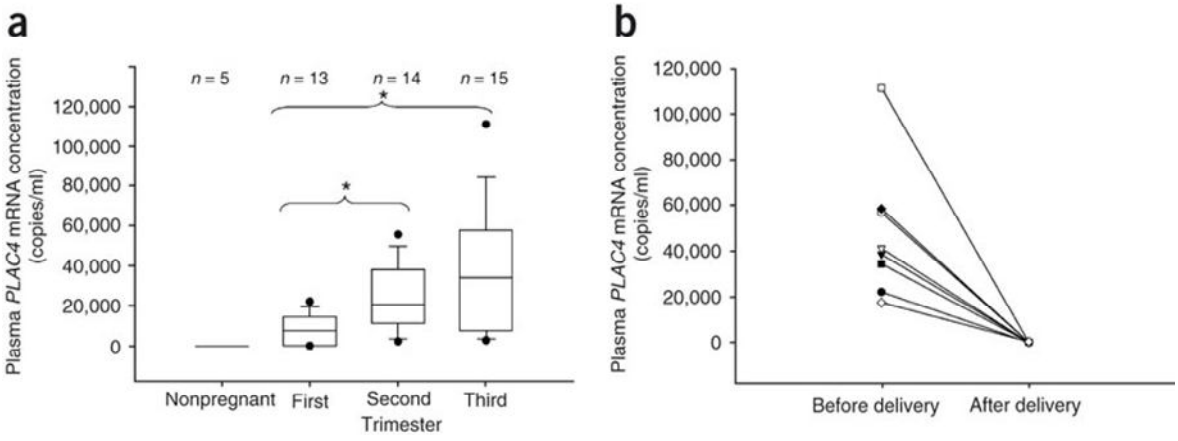


Figure 1.9: Detection of *PLAC4* mRNA in maternal plasma
(A) *PLAC4* mRNA levels increase throughout gestation. The lines inside each box represent median values, the boxes mark the interval between the 25th and 75th percentiles, the whiskers denote the intervals between the 10th and 90th percentiles, and filled circles mark the data points outside of the 10th and 90th percentiles. (B) Circulating *PLAC4* mRNA is cleared rapidly after-delivery, to undetectable levels 24 hours post delivery. (Taken from Lo *et al*, 2007).

In order to detect trisomy of Chromosome 21 the allelic ratio of single nucleotide polymorphisms (SNPs) must be detected in order to give the ratio of maternal alleles to paternal alleles (Lo *et al*, 2007). In the case of the *PLAC4* mRNA transcript the specific SNP (rs8130833 NCBI Build 36.2) is an A → G transition. The allelic ratio of cfp RNA is presumed to reflect the gene dosage in cases of euploidy and trisomy (figure 1.10). The detection of the ratio of SNPs can be used to detect an allele ratio of 2:1 in cases of trisomy 21 and a ratio of 1:1 in normal diploid cases (see section 1.5.4.2). Due to the number of molecules of RNA being in the range of 10^4 - 10^6 in the first trimester, *PLAC4* mRNA is readily detectable, in the case of the specific SNP tested only 69/119 pregnancies were heterozygous and therefore assessable by the technique (Lo *et al*, 2007); the test detected 90% of trisomy 21 pregnancies in the sub group studied with 0% false positive results (Lo *et al*, 2007).

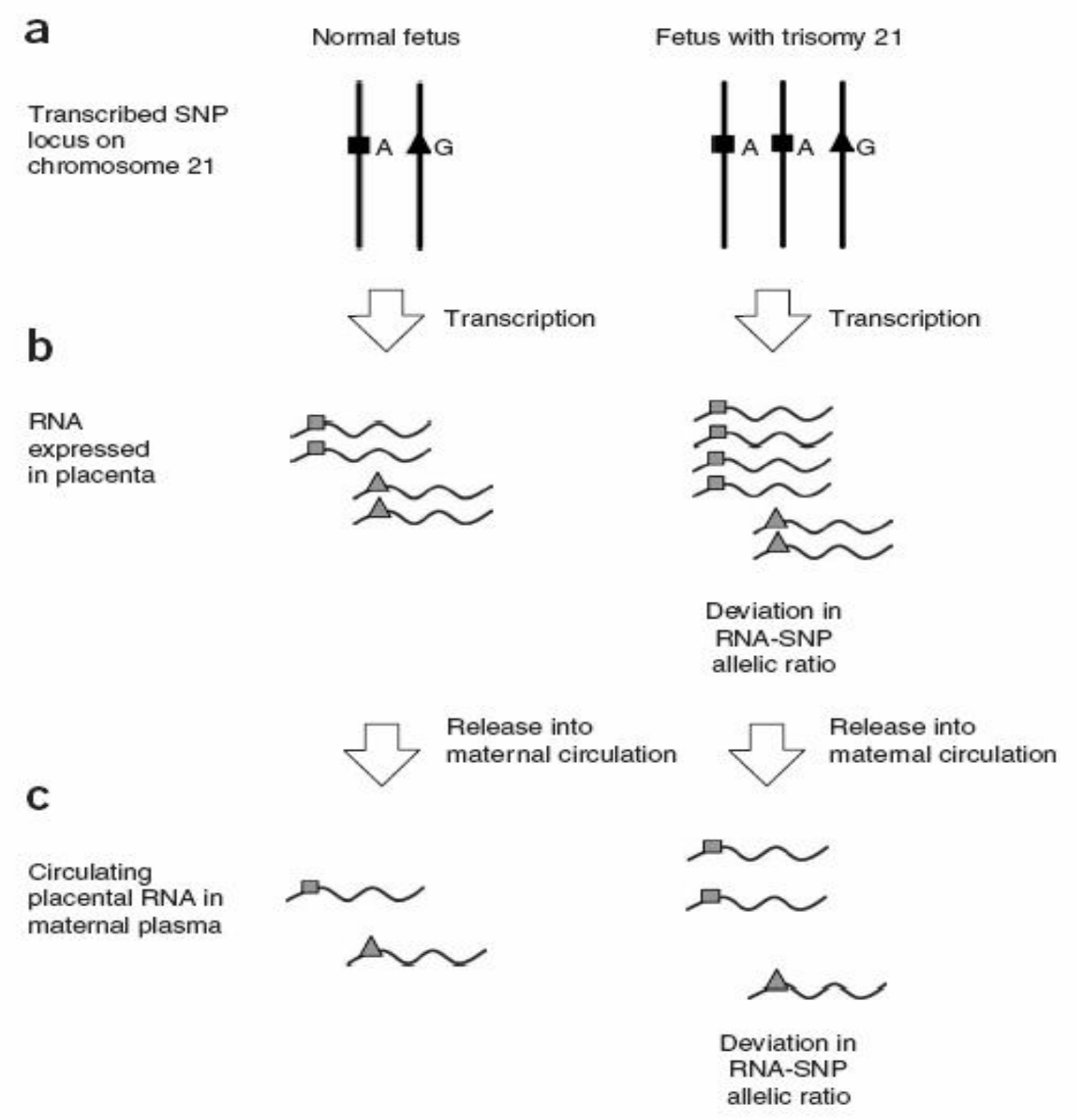


Figure 1.10: Schematic of the allelic ratio strategy for the measurement of fetal trisomy using RNA
(A) Normal and trisomic fetuses heterozygous at the transcribed SNP for *PLAC4*. "A" and "G" denote the SNPs for each respective allele. (B) The gene is expressed in placental tissue; the ratio of the two RNA alleles in trisomy 21 is expected to differ from that of a normal placenta, as the result of the extra copy of the gene. (C) The RNA transcripts are released into maternal blood and their relative abundance is reflective of that in the placenta. In trisomy 21 the ratio of the two RNA alleles is expected to deviate from that of a normal pregnancy. (Taken from Lo *et al*, 2007).

Using the allelic ratio method with a number of informative (heterozygous) SNPs such as the one described (Lo *et al*, 2007); could provide enough coverage of the population to be used in routine clinical NIPD. As the rate of heterozygosity of specific SNPs varies in different populations and the maximum theoretical

percentage of heterozygosity for any given SNP is 50%, a number of SNPs are required to ensure that at least one is informative (heterozygous) for the chromosome of interest in each test.

1.5 Cell-free fetal DNA in the maternal circulation

The passage of fetal cells into the maternal circulation has been well documented (Bianchi, 2004; Fisher *et al*, 1989; Librach *et al*, 1991; Weier *et al*, 2005). Given the rarity of fetal cells in the maternal circulation and the technical difficulties of fetal cell enrichment (Bianchi *et al*, 1990; Chen *et al*, 2004; Kolvraa *et al*, 2005; Krabchi *et al*, 2001; Guetta *et al*, 2003); other avenues have been sought for the development of non-invasive prenatal diagnosis (NIPD), such as the use of cell-free DNA in the maternal circulation.

The potential use cell-free DNA in plasma or serum was first developed as a tool for molecular diagnosis of cancer risk; as tumour DNA has been detected in the plasma and serum of cancer patients. One example is presence of hypermethylated DNA in plasma from the *APC* gene, which may be a useful biomarker of aggressive disease in oesophageal adenocarcinoma patients (Kawakami *et al*, 2000). Numerous studies have been undertaken in order to develop diagnostic tests based on the detection of cell-free tumour DNA in the circulation of cancer patients.

The placenta has been well characterised as an invasive tissue (Fisher *et al*, 1989). Based on the hypothesis that the placenta has some cancer like properties Lo and colleagues began work on attempting to identify cell-free fetal DNA (cff DNA) in the circulation of pregnant women. If cff DNA could be identified and isolated from maternal peripheral blood samples, it could potentially be used in diagnostic tests for NIPD of fetal aneuploidy.

The first major breakthrough in NIPD came when cff DNA was detected in maternal plasma (Lo *et al*, 1997). A PCR assay was developed for the detection of the "sex determining region Y" (*SRY*) gene, the product of which is the "testis determining factor" (TDF) which is responsible for male sex determination. *SRY* is present on

the Y chromosome (Yp11.3). Although *SRY* is a marker for cff DNA it can only be applied to male bearing pregnancies.

In a study of 30 pregnant women (all bearing a male fetus) plasma and serum from maternal peripheral blood collected were analysed by PCR assay of the *SRY* gene. Fetal (*SRY*) DNA was detected in 24/30 cases using maternal plasma as a source of cff DNA and in 21/30 cases using maternal serum as a source of cff DNA. In the control group of 13 pregnancies in which the fetus was female and 10 non-pregnant women, serum and plasma samples were also assessed by *SRY* PCR assay; no *SRY* sequences were detected in these samples (Lo *et al*, 1997).

1.5.1 Origin of cell-free fetal DNA in the maternal circulation

In contrast to the traditional teaching that the placenta forms an impermeable barrier between the mother and fetus, many studies have shown that intact fetal cells circulate within the maternal blood. There has been much speculation as to the tissue source of circulating cff DNA. Fetal cells circulating in the maternal circulation include: cytotrophoblasts (CTBs) which invade the maternal circulation, fetal lymphocytes, and fetal nucleated red blood cells (FNRBCs). Another possible source is direct transfer of cff DNA from the amniotic fluid (Bianchi, 2004).

Evidence suggests that the amount of cff DNA molecules present in the maternal circulation is up to 25 times greater than that which could be obtained from the DNA present in circulating fetal cells. This indicates that circulating FNRBCs are not a major contributor to cff DNA in the maternal circulation (Lo *et al*, 1998 a).

Cff DNA has also been detected in several other bodily fluids including: amniotic fluid, maternal urine, and maternal cerebrospinal fluid (Bianchi *et al*, 2001; Botezatu *et al*, 2000; Angert *et al*, 2004). It is possible that the source of cff DNA circulating in the maternal circulation may originate from the fetus by leakage of the amniotic fluid; however conclusive studies have not been conducted to test this hypothesis so far. The most likely source for the majority of cff DNA in the maternal circulation is the placenta (Bianchi 2004).

The definitive feto-placental circulation establishes approximately 28-30 days after conception. By day 28 cff DNA is detectable in (80%) of all pregnancies. The increase of cff DNA throughout gestation shows strong correlation with hCG levels in the maternal circulation (Ohashi *et al*, 2002; Sekizawa *et al*, 2001).

These studies imply that the placenta is the major source of cff DNA in the maternal circulation; however they do not prove this conclusively. In 2007 Alberry and colleagues demonstrated that the placenta is the major source of cff DNA in the maternal circulation, by testing the plasma of women undergoing anembryonic pregnancies (where an embryo fails to develop or terminates at an early stage of development). In 5 anembryonic pregnancies cff DNA was detected using a *DYS14* gene PCR assay, the *DYS14* gene is also present on the Y chromosome (Yp11.2) (Alberry *et al*, 2007).

The amount of cff DNA detected was comparable to that of other normal pregnancies, indicating that the placenta is the major contributor of cff DNA in the maternal plasma (figure 1.11). The range of cff DNA detectable in normal controls was 88.9-201.2 ge/ml and the range of detectable cff DNA in the anembryonic pregnancies was 80.1-770 ge/ml (Alberry *et al*, 2007).

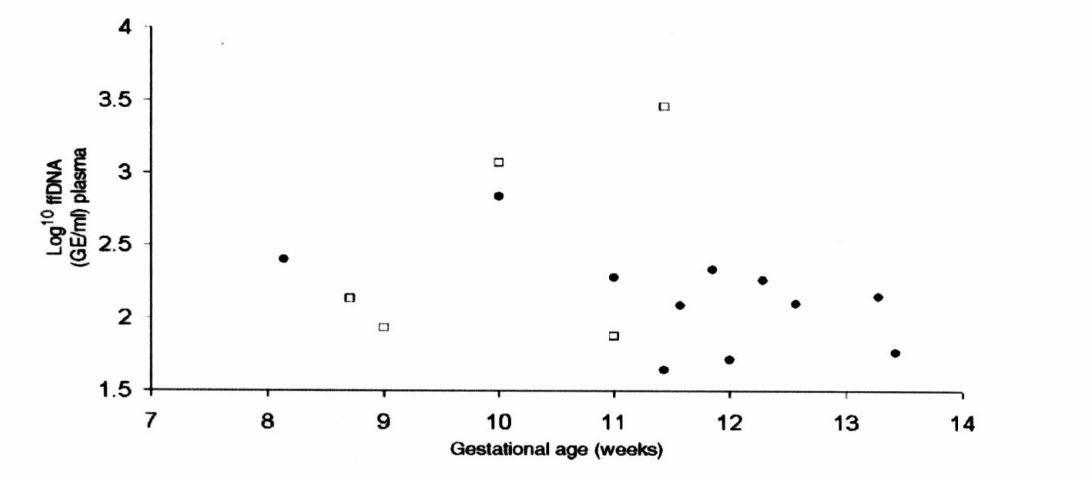


Figure 1.11: Log distribution of cff DNA in anembryonic and normal pregnancies throughout gestation
Cff DNA sequences from the *DYS14* gene are detected in anembryonic pregnancies throughout "gestation" (□); at amounts which are comparable to that of regular pregnancies (●). (Taken from Alberry *et al*, 2007).

One issue with the study is that the *DYS14* assay was used to detect cff DNA; *DYS14* is a variable-multi copy gene, with copy numbers in males ranging from approximately 30-60 copies (Vodicka *et al*, 2007). Therefore exact quantification of cff DNA cannot be compared between individuals; however the fact still remains that cff DNA was detected in anembryonic pregnancies; showing that the placenta does have a major role in the release of cff DNA into the maternal circulation.

FNRBCs are also thought to contribute a small proportion of cff DNA and have been shown to undergo apoptosis in the maternal circulation. FNRBCs have been shown to undergo apoptosis by (terminal UdTP nuclear end labelling (TUNEL). TUNEL detects apoptotic cells by catalysing the binding of biotin labelled UdTP to nicks in double stranded DNA (Gavrieli *et al*, 1992). In one study results showed approximately half of all FNRBCs were undergoing apoptosis (Sekizawa *et al*, 2000).

1.5.2 Haematopoietic origin of cell-free maternal DNA

Many studies have used markers such as *SRY* and *DYS14*, which are specific to male DNA. This is because cff DNA in the maternal circulation is present in a high background of cell-free maternal DNA (cfm DNA). If DNA can be released from fetal cells circulating in the maternal circulation, it is logical to assume that maternal circulating blood cells can also release DNA into the circulation. Lo and colleagues developed a sex-mismatch bone marrow allotransplantation model; in order to identify the source of cell-free plasma DNA (cfp DNA) in otherwise healthy individuals (Lui *et al*, 2002).

In the study blood was collected from male patients who had been the recipient of a bone marrow transplant from a female donor; and female patients who had been the recipient of a bone marrow transplant from a male donor. Samples were collected from patients who had demonstrated engraftment of bone marrow by fluorescent *in situ* hybridization (FISH) analysis. Assay results for male DNA (*SRY*) and total DNA (*β-GLOBIN*) were used to calculate the percentage of cfp DNA which was male in origin. Analysis of the percentage of *SRY* sequences detected was variable between the two patient groups. Male recipients of bone marrow from female donors had a

much lower amount of detectable *SRY* DNA than female recipients who received bone marrow from a male donor. This indicates that the higher levels of *SRY* DNA sequences detected in females were of haematopoietic origin (figure 1.12).

The interpretation of that data is that the majority of cfp DNA is predominantly derived from cells originating from the haematopoietic lineage. Although a small proportion of cfp DNA may originate from other tissues and organs, it has been determined that endothelial cells which line veins and arteries contribute very little if any cell-free DNA to the circulation (Jahr *et al*, 2001).

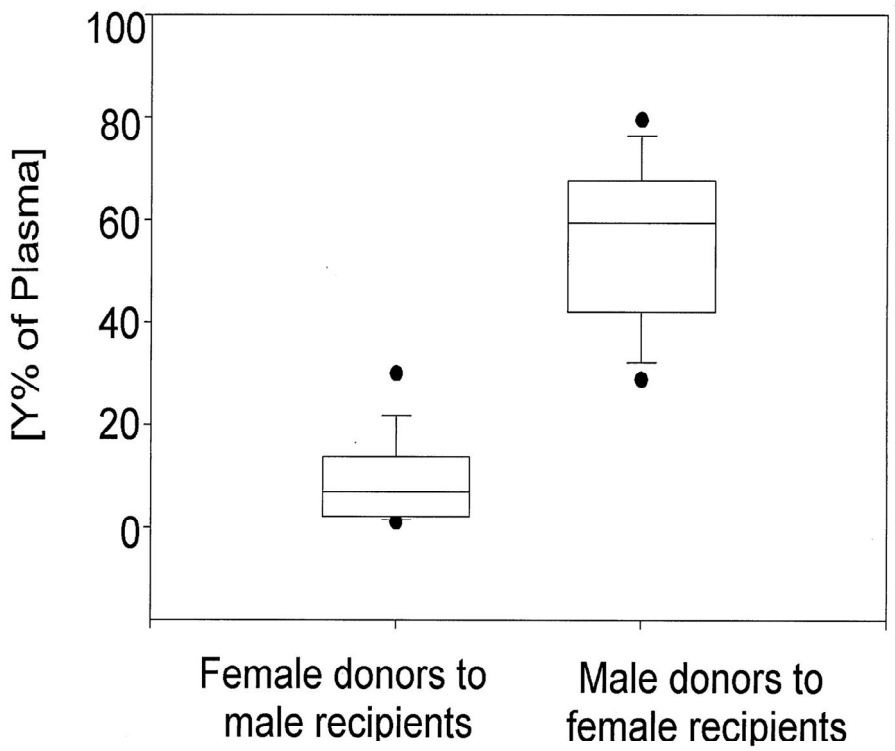


Figure 1.12: Percentage of Y chromosome DNA present in the plasma of bone marrow transplantation patients
The subject categories are shown on the x axis and the percentage of Y chromosome DNA in plasma is on the y axis. *SRY* sequences originating from the Y chromosome are detected by real-time PCR. The upper and lower horizontal bars denote 90th and 10th percentiles respectively. The upper and lower line in the boxes represent the 75th and 25th percentiles and the line within each box represent the median. Outliers are also shown (●). (Taken from Lui *et al*, 2002).

1.5.3 Stability and size distribution of cell-free fetal DNA in the maternal circulation

Cff DNA has been shown to be extremely stable when isolated from the maternal circulation. This may be due to association with sub-cellular particles after release from the cell by apoptosis (Sekizawa *et al*, 2000). TUNEL staining has been used on FNRBCs to show that a proportion of FNRBCs are undergoing apoptosis. In one study by Sekizawa and colleagues 42.7% of FNRBCs were shown to be undergoing apoptosis compared to only 3.5% of maternal cells; whereas in another study the percentage of FNRBCs undergoing apoptosis was much higher at 95% (Kolialexi *et al*, 2004; Sekizawa *et al*, 2000).

Due to the nature of these studies (collecting cells and assessing them *in vitro*) it is not possible to determine if the cells undergoing apoptosis would be doing so *in vivo*; as sample collection and procedures performed may unduly stress cells and cause them to undergo apoptosis. What is also unknown is the rate at which cells in the circulation are undergoing apoptosis and if there are any external factors in the circulation which may affect this.

The size distribution of cff DNA molecules when compared to that of cfm DNA molecules is relatively small; this suggests that fetal cells in the maternal circulation are undergoing apoptosis. A number of studies have been conducted with the aim of detecting apoptosis in fetal cells (Gavrieli *et al*, 1992; Kolialexi *et al*, 2004; Sekizawa *et al*, 2000). However, whilst detecting apoptosis of fetal cells; these studies overlooked their role in the production and release of fragmented fetal DNA.

Later studies assessed the size of cff DNA molecules in the plasma of pregnant women (Chan *et al*, 2004; Chan *et al*, 2005 a; Koide *et al*, 2005; Li *et al*, 2004). Evidence has shown that cell-free DNA which is released from apoptotic cells is likely to be fragmented by nucleosomal cleavage (Giacona *et al*, 1998). By separating plasma DNA from pregnant women by gel electrophoresis, it has been shown that cell-free DNA molecules are present in a wide range of sizes. The study also showed that cff DNA is generally shorter than 300 bp in length (Li *et al*, 2004).

Another study by Chan and colleagues used a real-time PCR assay to amplify different sized amplicons of the *SRY* gene, to assess the size distribution of cff DNA. Each primer pair had the same forward primer and a different reverse primer to give an amplicon size distribution panel of 107 bp, 137 bp, 193 bp, 313 bp, 392 bp and 524 bp. These primers were used to amplify plasma DNA from pregnant women bearing a male fetus, and were compared to total cell-free DNA by a similar assay for the *LEPTIN* gene (7q31.3) (figure 1.13). The data show that cff DNA is fragmented, and that the majority (< 99 %) of cff DNA molecules are smaller than 313 bp in length (Chan *et al*, 2004).

It is important to bear in mind the fact that cff DNA molecules are very small. When developing primers for NIPD it is important to design primers which target small regions so as to maximize the potential for amplification of cff DNA fragments. Additionally it is important to note that cfm DNA molecules are on average much longer than cff DNA, with some overlap in size distribution. This provides scope to use size separation techniques in order to enrich cff DNA for NIPD (Clausen *et al*, 2007; Legler *et al*, 2007).

1.5.3.1 Effects of sample collection and handling on cell-free fetal DNA levels

Before samples can be analysed, they must first be collected and processed and in most cases stored before use. One study has shown that the long-term storage of plasma DNA at -20 ^{0}C does not significantly degrade the DNA molecules when compared to plasma DNA that has only been stored for a month (Koide *et al*, 2005). Additional studies have also provided valuable information on the collection and storage of plasma DNA before use. One study analysed a number of factors including: allowing blood to clot before plasma separation; the use of ethylene diamine tetra acetic acid (EDTA) in the storage of blood samples; the effect of repeat freeze-thawing of samples; and the effects of long term storage at -80 ^{0}C (Chan *et al*, 2005 a).

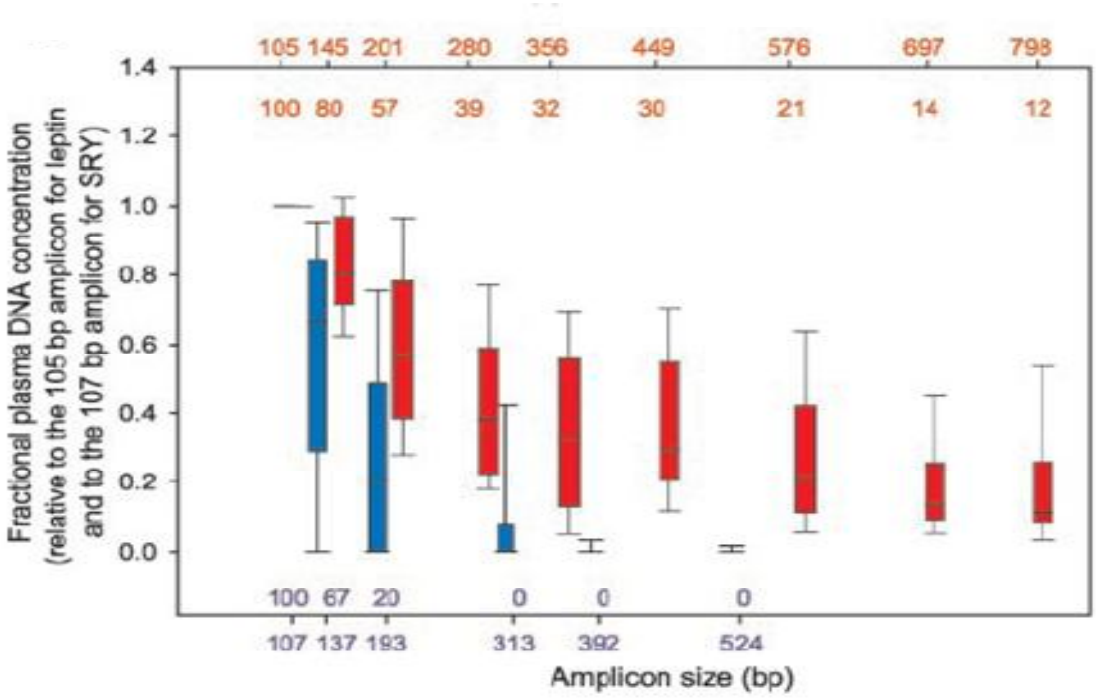


Figure 1.13: Fractional concentrations of fetal derived and maternal derived DNA in maternal plasma, as assessed by PCR of differently sized amplicons
The numbers at the top of the plot (red) represent the amplicon sizes of the maternal fragments 105-798 bp, and underneath these values is the median fraction (red) represented by these amplicons 100-12 %. The numbers the bottom of the plot (blue) represent the amplicon sizes of the maternal fragments 107-524 bp, and above these values is the median fraction (blue) represented by these amplicons 100-0 %. The top middle and bottom lines on each bar represent the 75th, 50th, and 25th percentiles respectively; and the error bars represent the 95th and 5th percentiles. (Taken from Chan *et al*, 2004).

The results of the study offer insight into how samples should be collected and stored, for the maximum preservation of cff DNA whilst preventing maternal lymphocyte lysis. Allowing blood to clot before the separation of the plasma fraction increases the amount of DNA in the samples; this is primarily due to the lysis of maternal lymphocytes and should be avoided in order to reduce the background of maternal DNA. Storing plasma at -20 ^{0}C and thawing it once before use has little effect on the amount of cff DNA present in the sample, however repeat freeze-thaw cycles degrade DNA and cause further fragmentation. Therefore, when DNA is extracted it should be aliquoted into separate tubes for individual usage. The general consensus is that upon obtaining a blood sample, it should be stored in an EDTA tube at 4 ^{0}C and the plasma fraction should be separated no longer than 6 hours after collection to minimise maternal cell lysis (Chan *et al*, 2005 a).

1.5.4 Kinetics of circulating cell-free fetal DNA

Cff DNA is detectable in the maternal circulation only 5 weeks after conception. To date a number of small studies have shown that transfer of cff DNA into the maternal circulation is a dynamic process and that cff DNA is cleared rapidly after delivery (Ariga *et al*, 2001; Birch *et al*, 2005; Guibert *et al*, 2003; Ingargiola *et al*, 2003; Smid *et al*, 2003). The absolute amount of cff DNA and its ratio to cfm DNA increases throughout gestation (Ariga *et al*, 2001). Cff DNA is also more readily detectable in the plasma fraction than the serum fraction, and is rapidly cleared *post partum* (Ariga *et al*, 2001).

In strong agreement with the body of evidence in the literature, the increase in cff DNA throughout gestation has been shown to correlate with the increase of βhCG amounts, which is a product of the placenta. Cff DNA is generally only detectable in the maternal circulation once the feto-maternal circulation is established (Guibert *et al*, 2003).

Studies have also been undertaken to assess total cff DNA levels throughout gestation, one study assessed 246 male bearing pregnancies from time points ranging from 6-40 weeks. The main aim of the study was to assess cff DNA (*SRY*) levels at each time point during gestation, to establish the feasibility of using cff DNA in NIPD (Galbiati *et al*, 2005). Median values for cff DNA in the first trimester were found to be 9.1 diploid genome equivalents per ml ge/ml and in the second trimester 16.6 ge/ml and in the third trimester 51.8 ge/ml. Another study conducted by Birch and colleagues on a similar number of samples was found to agree with the findings of the Galbiati study, with the exception of slightly elevated cff DNA levels in the first and second trimesters (figure 1.14) (Birch *et al*, 2005).

One issue with both studies is that they assess single time points in different pregnancies and do not follow single cases from early gestation to term. This is likely to be due to the difficulty of regularly testing pregnant women and that

regularly collecting blood from pregnant women is unadvisable due the association of anaemia in some pregnancies.

What both studies do show is that the amount of cff DNA increases throughout gestation, but cff DNA increases do not correlate exactly with gestational age, there is a small increase from the first trimester to the second and then a larger increase in the third trimester. This may be due to placental breakdown in the third trimester before delivery (Birch *et al*, 2005). Cff DNA levels also vary between individual pregnancies, with many pregnancies having much lower levels than the median levels detected. Cff DNA levels are expected to be the product of multivariate factors including: placental size, extravastion of cells and the rate of fetal cell apoptosis (Birch *et al*, 2005).

In the study by Birch and colleagues, calibration curves were generated from commercially available human DNA (Promega, USA). However, when generating calibration curves from genomic DNA, it is not shown that the DNA used was sonicated, and is used directly in the assessment of the plasma DNA samples. Due to the fragmented nature of cfp DNA it is essential to fragment genomic samples, in order to more closely mimic the situation in plasma, to more accurately assess the amount of cfp DNA.

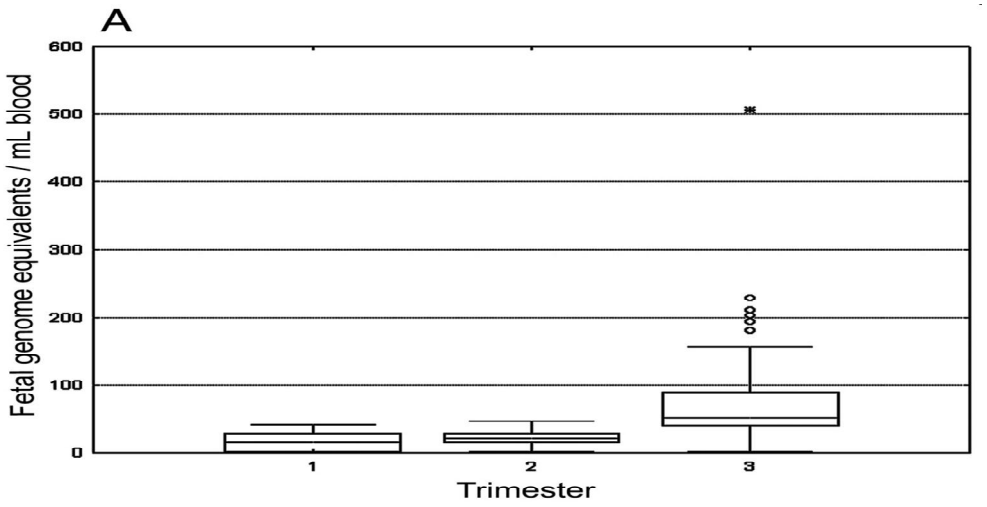


Figure 1.14: Concentrations of cff DNA during gestation as determined by *SRY* assay
Fetal genome copy number as extrapolated from *SRY*-PCR assay, increases throughout gestation, with a negligible increase from the first to second trimesters. The greatest increase is observed between the second and third trimesters. The upper and lower limits of each box denote the 75th and 25th percentiles respectively and midline represents the median. Median values are 15.89, 21.53 and 52.03 ge/ml at the first, second and third trimesters respectively. The error bars represent the 10th and 90th percentiles. Outliers (○) and extreme outliers (*) are also shown. (Taken from Birch *et al*, 2005).

1.5.4.1 Clearance of cell-free fetal DNA from the maternal circulation

One aspect of cff DNA kinetics is that cff DNA is cleared rapidly after delivery. Lo and colleagues demonstrated that the average half-life of cff DNA in the maternal circulation is 16.3 minutes in pregnancies delivered by caesarean section (Lo *et al*, 1999 b). They also demonstrated that cff DNA levels surge during delivery and are higher immediately after delivery than at any time point during the pregnancy, this is most likely due to placental breakdown and fetal cells being released into the maternal circulation (Lo *et al*, 1999 b). Another study assessing cff DNA levels after conventional vaginal delivery also broadly agreed that cff DNA is cleared rapidly in most cases (Ingargiola *et al*, 2003). However, both studies showed cases where cff DNA is not completely cleared 24 hours after delivery (figure 1.15) (Lo *et al*, 1999 b; Ingargiola *et al*, 2003).

Data from these studies indicate that the release of cff DNA into the maternal circulation is a dynamic process, requiring the continuous transfer of fetal DNA from the placenta in order to remain present throughout gestation. One likely route of clearance is via the kidneys as cff DNA can be detected in urine (Botezatu *et al*, 2000). This means that cff DNA in the maternal circulation from one pregnancy does not persist for long periods after delivery, and therefore cannot be detected in subsequent pregnancies.

The persistence of some fetal cells *post partum* is the only caveat by which cff DNA may be released into the maternal circulation outside of pregnancy, these cells are generally stable in the maternal circulation and can persist for a number of years (Bianchi *et al*, 1996; Guetta *et al*, 2003). It is only upon sample collection and DNA extraction procedures that these cells may release cff DNA from a previous pregnancy. This potential difficulty of detecting cff DNA from a previous pregnancy is overcome by the use of a specific protocol for sample collection and processing developed by Santacroce and colleagues, which completely prevents cff DNA from previous pregnancies from contaminating the sample from the current pregnancy (Santacroce *et al*, 2006).

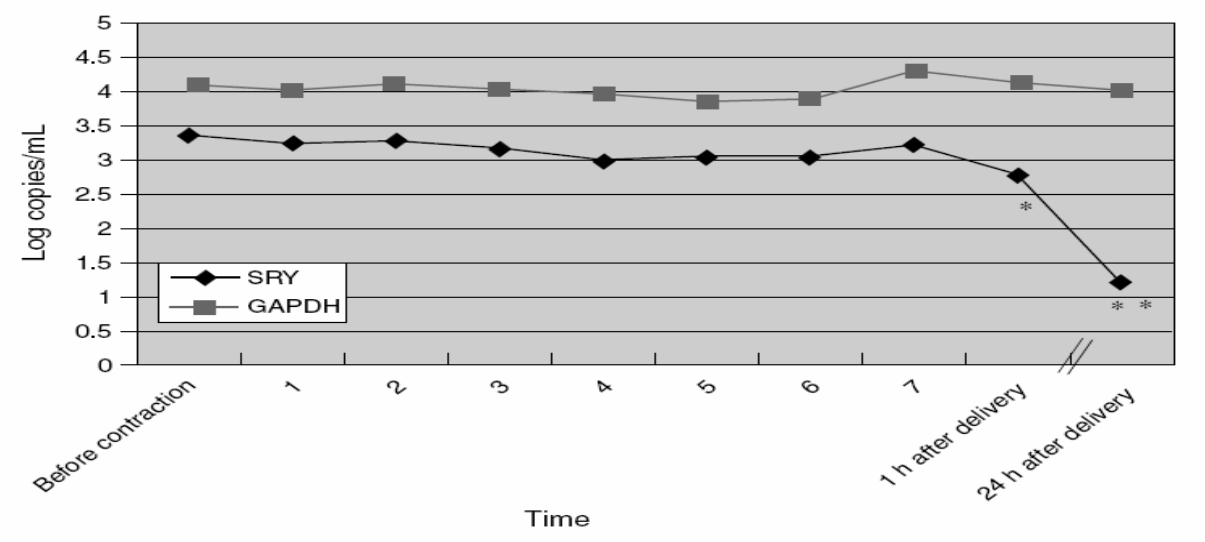


Figure 1.15: Log copies of cell-free DNA in the maternal plasma before and after delivery
Log copies of cff DNA (*SRY*) and total DNA (*GAPDH*) per millilitre in maternal plasma. Samples were collected before contraction, each hour during labour, 1 hour after delivery and 24 hours after delivery. Most fetal (*SRY*) sequences are cleared post delivery. (Taken from Inagargiola *et al*, 2003).

Blood collected is stored in EDTA tubes with formaldehyde to prevent any cell lysis from either maternal cells or cells from a previous pregnancy. The blood is then centrifuged twice at 13,000 g to obtain plasma and the plasma is finally filtered (with a 0.22 μm pore filter). This procedure stabilises any fetal cells present and removes them carefully to remove any potential for cff DNA from a previous pregnancy to be analysed (figure 1.16) (Santacroce *et al*, 2006). Assessment of cff DNA was made by the detection of *AMELOGENIN* DNA marker which has a 6 bp difference in size between the X chromosome copy and the Y chromosome copy in the 3rd intron, with the X chromosome copy being 102 bp in length and the Y chromosome copy being 108 bp in length, the presence of fetal DNA was measured in plasma DNA from women who had previously carried a male fetus but were not pregnant at the time of testing (figure 1.16) (Santacroce *et al*, 2006).

One study stated that formaldehyde could be used for the stabilisation of maternal blood samples and greatly increases the amounts of cff DNA detected in each sample; these observations were attributed in part to the properties of formaldehyde, which caused two effects. First, it was surmised that the cell membrane stabilising properties of formaldehyde prevented the lysis of circulating maternal blood cells, thus lowering the maternal background of cfm DNA. Second, formaldehyde's nuclease inhibitory effect presumably helped preserve the cff DNA already present and prevented its loss by degradation (Dhallan *et al*, 2004). Dhallan's group claimed to be able to increase the proportion of cff DNA present from a maximum of about 6% (Lo *et al*, 1998 a) to mean values of 20.2 and 25% from samples collected during various stages of gestation (Dhallan *et al*, 2004); another group supported these findings (Costa *et al*, 2004). However, several other groups followed this work in an attempt to reproduce the results, as this discovery would represent a major breakthrough for NIPD. One such study reported a moderate increase of cff DNA of between 1 and 3% (Chung et al, 2005); whilst other studies conducted by other researchers failed to reproduce the results reported by Dhallan's group (Chinnapapagari et al, 2005).

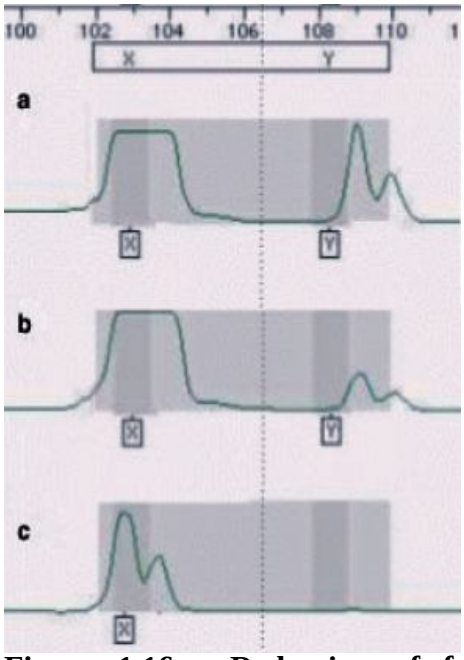


Figure 1.16: Reduction of fetal DNA contamination from previous pregnancy

Panels A-C show the effect of various treatments on plasma samples from non-pregnant woem whom had previously carried a male fetus. (A) Shows the amount of Y DNA contamination after gentle centrifugation at 3,000 g. (B) Shows the amount of Y DNA contamination after double centrifugation at 13,000 g. (C) Shows the amount of Y DNA contamination after double centrifugation at high speed and filtration. Only by double centrifugation of plasma samples at 13,000 g followed by filtration can DNA from previous pregnancies be eliminated from current samples. (Taken from Santacroce *et al*, 2006).

1.5.4.1.1 Clearance of methylated and unmethylated DNA

One factor which may affect the clearance of cell-free plasma DNA is the role of the immune system in pregnancy, depending on whether cell-free plasma DNA is methylated or unmethylated there may be bias of clearance of specific sequences. It is well documented that bacterial unmethylated CpG DNA acts as an antigen which drives the humoral T_H1 immune response (Agren *et al*., 2006; Yoshikawa *et al*., 2006). It is therefore possible that unmethylated CpG DNA originating from maternal leukocytes or placental tissue may also elicit a response. In the case of systemic lupus erythematosus (SLE) it has been shown that a lack of methylation in peripheral blood mononucleocytes causes an autoimmune response as the unmethylated CpG DNA mimics the antigenic properties of bacterial CpG DNA (Januchowski *et al*., 2004). It has also been suggested that methylated CpG DNA may be preferentially present in the circulation, due to the re-modelling of chromatin in nucleosomal DNA, perhaps biasing the levels of cff DNA present in the maternal circulation (Herman *et al*., 2004). The situation regarding the immune response to cell-free DNA during gestation is unclear, and requires further study (Puszyk *et al*, 2008).

1.5.4.2 Cell-free fetal DNA kinetics and diagnostic testing

The kinetics of cff DNA transfer into the maternal circulation is an important factor in the development of assays for NIPD, it is crucial to take into account the low levels of cff DNA obtained from maternal plasma; at the lower end of the scale the amplification of low copy number of cff DNA molecules may cause stochastic effects similar to that of allelic drop-out, which is the failure of an allele in a sample to amplify during PCR.

A careful study by Old and Stallard analysed the sampling errors inherent with the low copy numbers of cff DNA. By tabulating the expected numbers of cff DNA molecules from a standard 3ml sample of blood and accounting for standard errors, the minimum sample size for accurate NIPD was calculated. The two main methods of cff DNA detection were also compared and the minimum sample requirement for each test was calculated (Old and Stallard, accepted for publication).

The simplest method of detecting an aneuploidy is to measure the relative amount of the target chromosome by selecting a region located on the chromosome and amplifying it by real time PCR assay, and comparing the product with that of another region on another chromosome. For example a Chromosome 21 specific PCR product could be compared to a Chromosome 1 PCR product, this method is termed relative chromosome dosage (RCD). In a normal pregnancy the RCD is 2:2, however in a trisomy pregnancy the RCD is 3:2 (Chim *et al*, 2005; Chan *et al*, 2006; Old *et al*, 2007; Old and Stallard, accepted for publication).

The alternative method of detecting Chromosome 21 trisomy is to utilise single nucleotide polymorphisms (SNPs), which are single base sequence changes which may or may not differ depending on whether the allele is maternal in origin or paternal in origin. In a normal pregnancy the allelic ratio (AR) of SNPs is expected to be 1:1 in a normal pregnancy, however in a trisomic pregnancy the ratio would be 2:1 or 0.5:1 (see section 1.4.2). This is the allelic ratio method. (Tong *et al*, 2006; Old and Stallard, accepted for publication).

Both techniques are capable of being applied in a single tube test on the same sample, thereby reducing the error rates associated with product detection in real time PCR platforms, which may be as much as 0.6 cycle thresholds (Ct) (BIORAD, USA).

The low numbers of cff DNA molecules present, pose a difficult technical challenge, one that can only be overcome by the development of a panel of markers for cff DNA (table 1.1). Using the data from table 1.1, the volume of blood required for each test is calculated for different confidence intervals. Setting a relatively high standard for a diagnostic test, with a false positive rate of 0.01% and a false negative rate of 0.01%, the number of cff DNA genome equivalents (ge) required for the AR test is 72, and the number of cff DNA molecules for an RCD test is 216.

Assuming that there is a mean number of molecules present of 16.36 ge/ml (Birch *et al*, 2005) and that each (ge) genome equivalent contains 2 DNA molecules, the volume of blood required for the RCD test is calculated to be 216/(16.36 x 2)= 6.6mls. In the AR test the number of DNA molecules for each allele is counted separately, the volume of blood required for the AR test is calculated to be 72/16.36= 4.4 mls (Old and Stallard, accepted for publication).

One potential solution to the problem posed by sampling error caused by low cff DNA copy number is to use multiple biomarkers in an assay. By assaying marker using multiplex techniques, each biomarker is independent and subject to sampling error. However, independent biomarker quantifications can be combined to give a better estimate of RCD or AR. (Old and Stallard, accepted for publication).

False positive rate	**Number of DNA molecules required for AR (and RCD) method**				
	False negative rate = 0.05	**False negative rate = 0.025**	**False negative rate = 0.01**	**False negative rate = 0.005**	**False negative rate = 0.001**
0.05	39 (108)	46 (130)	55 (158)	62 (178)	77 (224)
0.025	45 (130)	53 (153)	63 (184)	70 (206)	85 (255)
0.01	53 (158)	62 (184)	72 (216)	80 (240)	96 (293)
0.005	59 (178)	68 (206)	79 (240)	87 (266)	104 (321)
0.001	73 (224)	83 (255)	95 (293)	103 (321)	122 (382)

Table 1.1: Number of fetal DNA molecules required for non-invasive prenatal diagnostic tests
Numbers of cff DNA molecules required to be assayed for the indicated false positive and false negative rates, for both tests are given. The number for the allelic ratio test equates to copy numbers required, whereas the numbers for the relative chromosome dosage test (in parenthesis) equate to the number of cff DNA molecules required.

1.5.5 Non-invasive prenatal diagnosis of paternally inherited genetic disorders

NIPD using cff DNA is already offered in some countries for specific genetic conditions. In most cases these are conditions where paternally inherited DNA sequences which are not present in the maternal genome can be used as fetal specific DNA markers, some examples follow.

Fetal sexing can be used to assess risk for certain sex-linked genetic disorders. Haemophilia is an X chromosome linked genetic disease where males are affected with defective blood clotting factors and females are carriers of the disease.

In a small study fetal sexing was offered to women who were carriers of haemophilia, using an *SRY* PCR assay; successful fetal sexing by the detection or absence of *SRY* sequences was 100%. This means that in the case of haemophilia diagnosis, non-invasive fetal sexing can be offered in the first instance, instead of invasive tests such as amniocentesis or CVS (Santacroce *et al*, 2006).

As part of routine prenatal care, NIPD for the Rhesus D blood group is offered in the U.K, France and the Netherlands (Lo *et al*, 1998, Sekizawa *et al*, 2007). This test provides the basis for administration of anti-D immunoglobulin as prophylaxis in pregnancies where the fetus is confirmed as *RHD*+ and the mother is *RHD*- (Sekizawa *et al*, 2007). The detection of *RHD*+ sequences can be used to predict fetal *RHD* status either by conventional or by real-time PCR. This is important for the prevention of fetal haemolytic disease, where an *RHD*- mother becomes sensitised to an *RHD*+ fetus, which initiates a maternal immune response to produce IgG anti-D antibodies, which can cross the placenta and destroy fetal blood cells, giving rise to fetal anaemia and jaundice (Van der Schoot *et al*, 2006).

In the Caucasian population the *RHD*- phenotype is caused by a deletion of the *RHD* gene, whereas in over 80 % of the African population the phenotype is caused by a non-functional copy of the *RHD* gene, *RHD* ψ the *RHD* pseudogene. Using a PCR assay for exons 4, 5 and 6 it is possible to distinguish *RHD*- from *RHD* ψ, which enables detection of all *RHD*- cases with 100 % accuracy and provides the basis of the non-invasive test for *RHD* (Finning *et al*, 2002). A non-invasive test has also been developed for *β-Thalassemia* major, the test is based on a PCR assay designed to detect a paternally inherited –CTTT deletion in the gene which is the most common mutational cause of *β-Thalassemia* (Chiu *et al*, 2002 b).

Adrenal congenital hyperplasia is detectable by assay of intragenic polymorphic markers in intron 2 of the *CYP21* gene. Used alongside the analysis of extragenic minisatellite sequences, successful detection of normal paternal sequences excludes healthy female fetuses, and allows the administration of the therapeutic drug dexamethasone, which reduces or prevents virilisation of affected females fetuses *in utero*, preventing the development of external genitalia (Chiu *et al*, 2002 a; Rjinders *et al*, 2001).

Myotonic dystrophy is an autosomal dominant disorder, associated with the unstable expansion of a CTG trinucleotide repeat in the 3' untranslated region of the *DM KINASE* gene. Affected individuals are heterozygous for the expanded allele. Myotonic dystrophy is characterised by the wasting of muscles (muscular dystrophy), heart defects, endocrine defects and difficulty in relaxing muscle (myotonia). Most notably, the highly variable age of onset decreases with successive generations (Amicucci *et al*, 2000).

The current scope of non-invasive prenatal diagnosis is so far limited to the detection of paternally inherited DNA sequences, which are not present in the maternal genome, for these limitations to be overcome the development of fetal-sex and paternal-mutation independent cff DNA specific biomarkers is required. Universal fetal-specific DNA markers would be of great benefit, in the diagnostic tests mentioned above, as they could be exploited as controls for the verification of the presence and quantity of cff DNA.

1.6 Discovery of putative fetal-specific epigenetic markers for use in non-invasive prenatal diagnosis

In 2005 Lo and colleagues hypothesised that it may be possible to detect cff DNA in the maternal circulation, by detecting epigenetic differences between fetal and maternal peripheral blood leukocyte DNA. Epigenetic differences are those differences that result in a change of phenotype without a change in DNA sequence (see pages 49 -52). They developed a novel model system, using placenta DNA and maternal blood DNA in order to measure epigenetic differences between the two tissues (Chim *et al*, 2005).

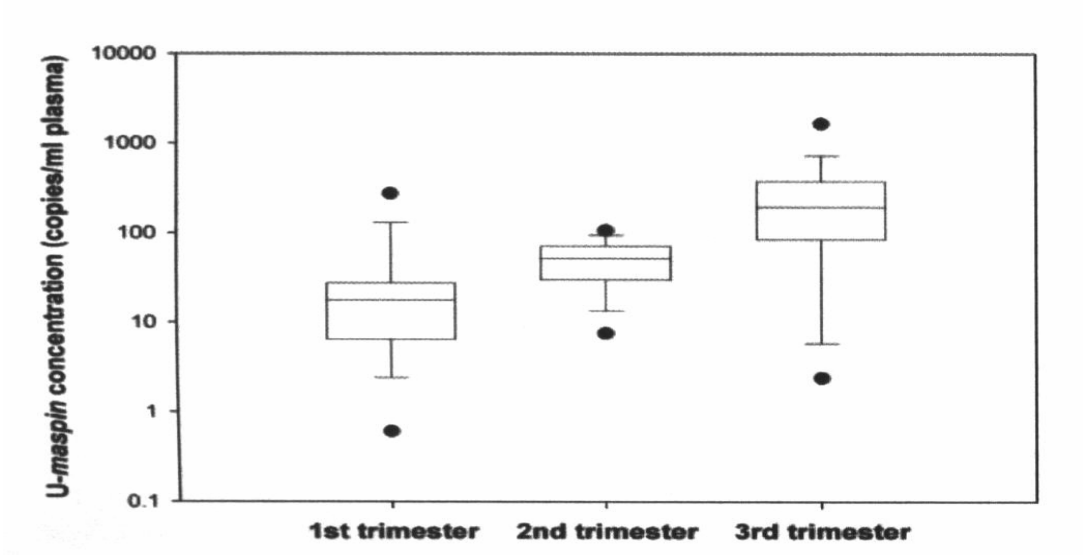
Applying bisulfite conversion of DNA to detect the methylation status of the *MASPIN* gene promoter (18q21.3), differential methylation was detected between placental DNA and maternal leukocyte DNA. *MASPIN* was proposed to be the first epigenetic biomarker of cff DNA. DNA with the placenta methylation signature was detected in the plasma of pregnant women (Chim *et al*, 2005).

Hypomethylated *MASPIN* or (U *MASPIN*) is the first potential epigenetic biomarker of cff DNA, levels of U *MASPIN* increase throughout gestation and U *MASPIN* is cleared rapidly after delivery (figure 1.17).

The development of a novel model system for the discovery of epigenetic biomarkers appeared to be an important breakthrough in NIPD. This potentially provided a much needed method for the identification of epigenetic fetal biomarkers, which could be used for the detection of aneuploidies (Chim *et al*, 2005).

An additional putative epigenetic biomarker was discovered in 2006, using the same model system and by detecting DNA with placental methylation signature a novel marker on the *RASSF1A* gene promoter (3p21.3) (Chan *et al*, 2006). The *RASSF1A* biomarker is the opposite of the *MASPIN* biomarker, in that it is methylated in placenta DNA and unmethylated in maternal leukocyte DNA. Markers whereby the fetal DNA is methylated are most useful as this allows the removal of the hypomethylated maternal DNA by methylation-sensitive restriction assay, leaving hypermethylated fetal *RASSF1A* undamaged (Chan *et al*, 2006).

A



B

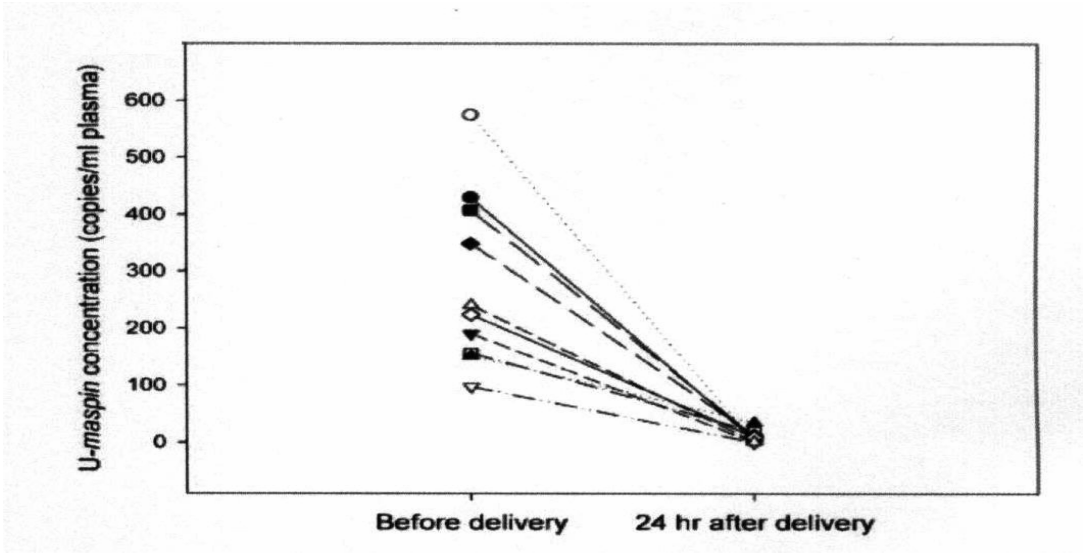


Figure 1.17: U-*MASPIN* concentrations throughout gestation and clearance after delivery

(A) The box plots of U-*MASPIN* concentrations in the first, second and third trimesters are shown in copies/ml of maternal plasma. U-*MASPIN* increases throughout gestation. The line within each box denotes the median, the limits of each box denote the 75th and 25th percentiles and error bars denote the 90th and 10th percentiles. Outliers are also shown (●). (B) Plots show levels of U-*MASPIN* in copies per ml of maternal plasma, before delivery and 24 hours after delivery in paired samples. U-*MASPIN* is cleared rapidly after delivery. (Taken from Chim *et al*, 2005).

The assessment of *RASSF1A* in the maternal circulation involves the digestion of plasma DNA by the methylation–sensitive restriction enzyme *BstU* I (Tsui *et al*, 2007). *RASSF1A* levels increase throughout gestation and are not detectable in non-pregnant women (figure 1.18) additionally *RASSF1A* sequences are cleared post delivery (Chan *et al*, 2006).

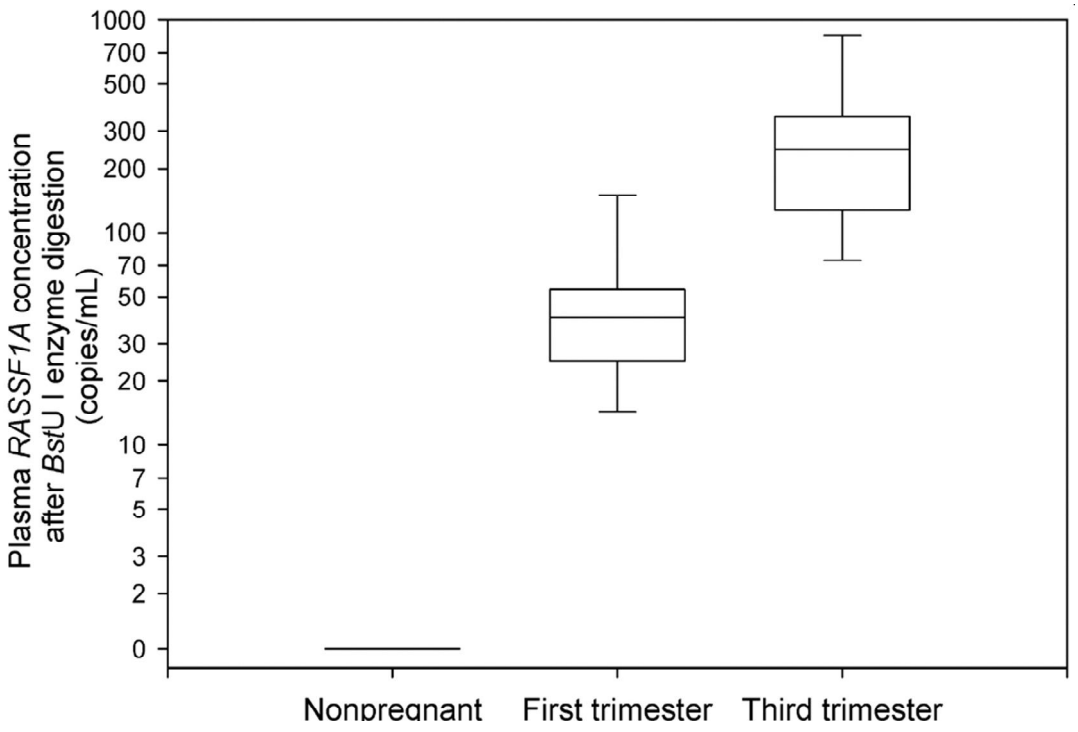


Figure 1.18: *RASSF1A* concentrations in maternal plasma during pregnancy
The whiskers represent the 90th and 10th percentiles and the upper, middle and lower box limits represent the 75th, 50th and 35th percentiles respectively. The Y axis shows copy numbers on a logarithmic scale. (Taken from Chan *et al*, 2006).

1.7 Epigenetics

Epigenetics is the term which refers to the changes in gene expression that are stable between cell divisions and sometimes between generations that do not involve changes in the underlying DNA sequence of the organism (Bird, 2007). In a broader context epigenetics is the term given to all of the processes and events occur around the genome, but do not alter the DNA sequences, even where stable inheritance has not been demonstrated (Ptashne, 2007). Epigenetic mechanisms include DNA methylation, histone methylation, histone deacetylation, and chromatin remodelling.

1.7.1 DNA methyltransferases methylate DNA

DNA methylation in mammalian cells occurs at the 5' position of Cytosine nucleotides which are followed by a Guanine in a CpG dinucleotide (p denotes the phosphate backbone, indicating that the nucleotides are on the same linear DNA strand rather than in a complementary pair). Approximately 70% of all CpGs are methylated and although CpG sequences are under-represented in the genome they can be found in relative abundance in the promoter region and first exon of genes in regions termed CpG Islands (CGIs) (Robertson and Wolffe, 2000).

5-methylCytosine does not exist in the cellular environment as a free nucleotide; in normal somatic cells the addition of a methyl group to Cytosine occurs after DNA replication and is catalysed by DNA methyltransferase (DNMT)(Bestor, 2000; Dahl and Guldberg 2003). DNMTs are proteins which catalyse the addition of methyl groups to Cytosine nucleotides in CpG dinucleotides (figure 1.19) (Chen *et al*, 2005).

There are two types of DNA methyltransferases: *de novo* methyltransferases and maintenance methyltransferases, the *de novo* methyltransferases DNMT3a and DNMT3b are responsible for methylating previously unmethylated cytosines, and the maintenance DNA methyltransferase DNMT1 copies pre-existing patterns onto new DNA strands during DNA replication (Klose and Bird, 2006). Another DNA methyltransferase has been identified, DNMT2 shows weak DNA methyltransferase activity, and its deletion from embryonic stem cells causes little detectable effect on DNA methylation (Hermann *et al*, 2003).

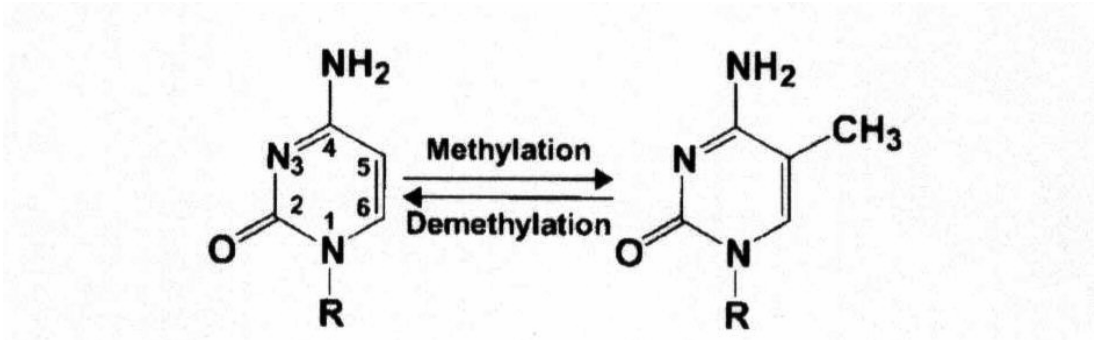


Figure 1.19: Schematic of Cytosine and 5-MethylCytosine
The structure of Cytosine (left) and 5-MethylCytosine (right) are shown. The methyl group is added to the 5 carbon on the ring structures of the Cytosine molecule. (Taken from Shiraishi *et al*, 2002).

An additional protein has also been included in the DNA methyltransferase family, DNMT3L (DNA methyltransferase like protein), on the basis that this protein has a highly conserved N-terminal PHD-like zinc finger domain with corresponding domains on DNMT3a and DNMT3b (Chen *et al*, 2005).

1.7.2 Role of DNA methylation in transcriptional silencing

DNA methylation produces a reversible protrusion from the major groove of DNA and is capable of altering local recognition signals (i.e. transcription factors) in the same way as it can block the action of methylation-sensitive restriction enzymes (preventing cleavage site recognition) (Eden and Cedar, 1994).

Most methylated CpGs are found within parasitic DNA elements in the human genome, as well as retrotransposons such as endogenous retroviruses, L1 elements and ALU elements. Parasitic DNA elements represent a significant threat to the structural integrity of the genome because they can mediate recombination between non-allelic repeats, which can cause chromosome rearrangements or translocations. Also active retrotransposons can integrate into and disrupt genes (Montagna *et al*, 1999; Kazazian, 1998; Robertson and Wolffe 2000). Local Cytosine methylation can interfere with transcription factor binding, and can also interfere with transcription when the first exon is methylated preventing the action of DNA polymerase (Robertson and Wolffe 2000), it is thought that this occurs in the same way as methylation-sensitive enzymes are blocked (Eden and Cedar, 1994). Methylated Cytosine can also recruit methyl-binding proteins (MBPs) which also function to repress transcription. MBPs contain a methyl-binding domain (MBD) which functions to recognise the methyl group on the Cytosine residue, Methyl-CpG binding domain protein 2 (MeCP2) was the first MBP characterised; it contains 2 structural domains, an MBD and a transcriptional repressor domain (TRD) (Nan *et al*, 1997).

In addition to MeCP2 there are other methyl-binding domain family members, proteins included in this family are: KAISO, MBD1, MBD2 and MBD3. How these proteins repress transcription is not fully understood, however it has been shown that

MBPs can bind repressors and histone deacetylases, which leads to a change in chromatin conformation to a condensed inactive state (Miranda and Jones, 2007). One theory is that DNA methylation causes the recruitment of MBPs which in-turn interact with histone deacetylases to form histone deacetylation complexes (HDACs), which is followed by the conformational change of chromatin to heterochromatin and the methylation of histone proteins (figure 1.20).

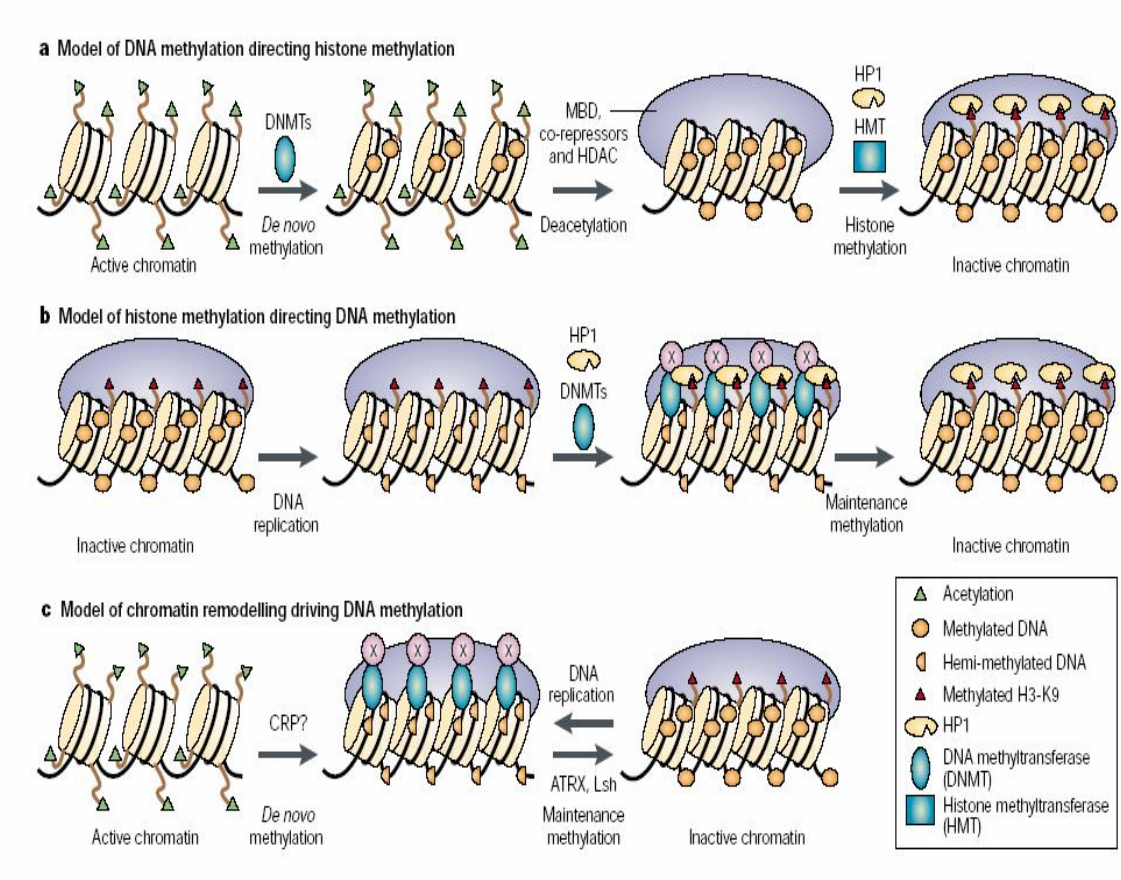


Figure 1.20: Links between DNA methylation, histone modification and chromatin remodelling

(A) A model of DNA methylation directing histone methylation. DNA methylation patterns are established through *de novo* methylation by DNA methyltransferases *DNMT3a* and *DNMT3b* and are maintained by *DNMT1*. Methyl binding proteins (MBPs) and histone deacetylase complexes (HDACs) are then recruited to the methylated CpG DNA in order to induce histone deacetylation and silencing. (B) A model of histone methylation directing DNA methylation. Methylated histone H3K9 acts as a signal for the inactive state chromatin, by recruiting histone protein 1 to methylated histones, which may in turn recruit DNA methyltransferases directly or indirectly, in order to stabilise inactive chromatin. (C) A model for chromatin remodelling driving DNA methylation. The ATP-dependent chromatin – remodelling and DNA helicase activities of proteins such as ATRX and Lsh, might facilitate DNA methylation and histone modification by unwinding nucleosomal DNA to increase its accessibility to DNMTs, HDACs and histone methyltransferases (HMTs). (Taken from Li, 2002).

1.7.3 Role of DNA methylation in embryonic development and imprinting of gametes

During female embryogenesis one of the X chromosomes is randomly chosen to be inactivated in each cell, leading to transcriptional silencing of thousands of genes on the X chromosome (Clerc and Avner, 1998). The X inactivation process converts one of the X chromosomes from the active euchromatin form to the transcriptionally silent heterochromatin form (Avner and Heard, 2001). DNA methylation is involved in this inactivation.

DNA methyltransferases are essential for *de novo* methylation of DNA during development. In mice the fertilised egg undergoes a wave of demethylation during pre-implantation development, which erases part of the inherited parental methylation pattern. After implantation the embryo undergoes a wave of *de novo* methylation, which establishes a new embryonic methylation pattern (Okano *et al*, 1999). The biological function of this *de novo* methylation is unknown; however it is essential for successful development.

Demethylation accompanied by *de novo* methylation occurs in gametogenesis, this is mediated by the DNA methyltransferases DNMT3a and DNMT3b, this process is essential for the establishment of parental imprinting of gametes before fertilisation (Tremblay *et al*, 1995). Genomic imprinting is a form of epigenetic gene regulation, which determines parent-of-origin-specific gene expression during embryonic development. The oocyte and sperm cell each contribute the same DNA to form a zygote; however, epigenetic imprinting of specific genes ensures that only 1 copy of the parent's alleles is expressed. Epigenetic paternalisation occurs in mitotically dividing spermatogonial stem cells and in meiotically dividing spermatocyte progeny, in order to endow sperm with imprinted alleles. Epigenetic maternalisation is restricted to the oocyte growth phase of folliculogenesis and takes place not during DNA replication, but during dictyate arrest in meiosis I (Kierszenbaum, 2002).

Inactivation of DNMT3a and DNMT3b disrupts *de novo* methylation. In one study mouse embryonic stem cells (ES cells) were raised to determine the function of *de novo* methylation, and whether or not the function of either protein overlaps.

DNMT3a +/- and DNMT3b +/- mice were grossly normal, however DNMT3a -/- mice died after 4 weeks of birth and DNMT3b -/- embryos did not survive to term. DNA methylation was measured by the methylation of endogenous retroviral DNA, it was also found that DNA was methylated to normal levels in DNMT3a -/- mice and slightly under-methylated in DNMT3b -/- mice. However, mice with both DNMT3a -/- and DNMT3b -/- embryos were highly under-methylated (Okano *et al*, 1999). In addition to the overlapping functions of DNMT3a and DNMT3b it was found that DNMT3b is required for the methylation of centromeric minor satellite DNA, which consists of tandem repeats of copy numbers between 50,000 and 100,000 (Okano *et al*, 1999). These data show that knock out of one copy of either or both genes does not have a deleterious effect on development, indicating an overlapping in the function of these proteins; however, if both copies of either gene are deleted normal development does not occur.

DNMT1 is the maintenance methyltransferase, the function of which is coupled to DNA replication. DNMT1 methylates hemimethylated DNA after DNA replication and, the knock out of the DNMT1 gene leads to global demethylation and embryonic lethality after a few rounds of replication (Howell *et al*, 2001; Li *et al*, 1992; Okano *et al*, 1999).

Differences in the methylation profiles of placenta and peripheral blood leukocyte tissues, due to the action of DNA methyltransferases are expected to form the basis of epigenetic markers by which cell-free fetal DNA can be isolated from the background of maternal DNA in the circulation.

1.7.4 Epigenetic study of DNA methylation

Methylation of DNA can modify gene expression by silencing transcription without changing the function or primary nucleotide sequence of a gene (Dahl and Guldberg, 2003). During the process of DNA methylation, methyl groups added to the Cytosine residues provide a chemical marker by which identical DNA sequences can be distinguished.

1.7.4.1 Genomic methylation mapping by methylated DNA immunoprecipitation

Monoclonal antibodies raised against 5-methylCytosine provide a powerful tool for the enrichment of methylated DNA sequences (Dahl and Guldberg, 2003). Methylated DNA immunoprecipitation (MeDIP) was developed as a tool for genomic methylation analysis as it does not require specific sites such as those required for methylation-sensitive restriction assays and is not as laborious as bisulfite sequencing (Weber *et al*, 2005).

This approach for the immunocapturing of fragmented methylated DNA is capable of providing up-to a 90 fold enrichment of methylated DNA depending on the proportion of methylated CpGs in the sequence (figure 1.21). The use of MeDIP in combination with comparative genomic hybridisation (CGH) or DNA microarrays, can provide 'methylation maps' of whole chromosomes, showing regions where methylated DNA is over-represented and regions where methylated DNA is under-represented (Wilson *et al*, 2006).

Before immunoprecipitation can be performed, DNA must be fragmented to approximately 200-300 bp in length. Oligonucleotides on microarray platforms are between 60 -70 bp in length, this means that any methylated CpGs detected on an array may be up-to 200 bp upstream or downstream from the position of the oligonucleotide signal. The inaccuracy of detection by the array means that higher resolution techniques such as methylation-sensitive restriction enzyme (MSRE) analysis or bisulfite sequencing are required to determine the exact sequence location of any methylated CpGs.

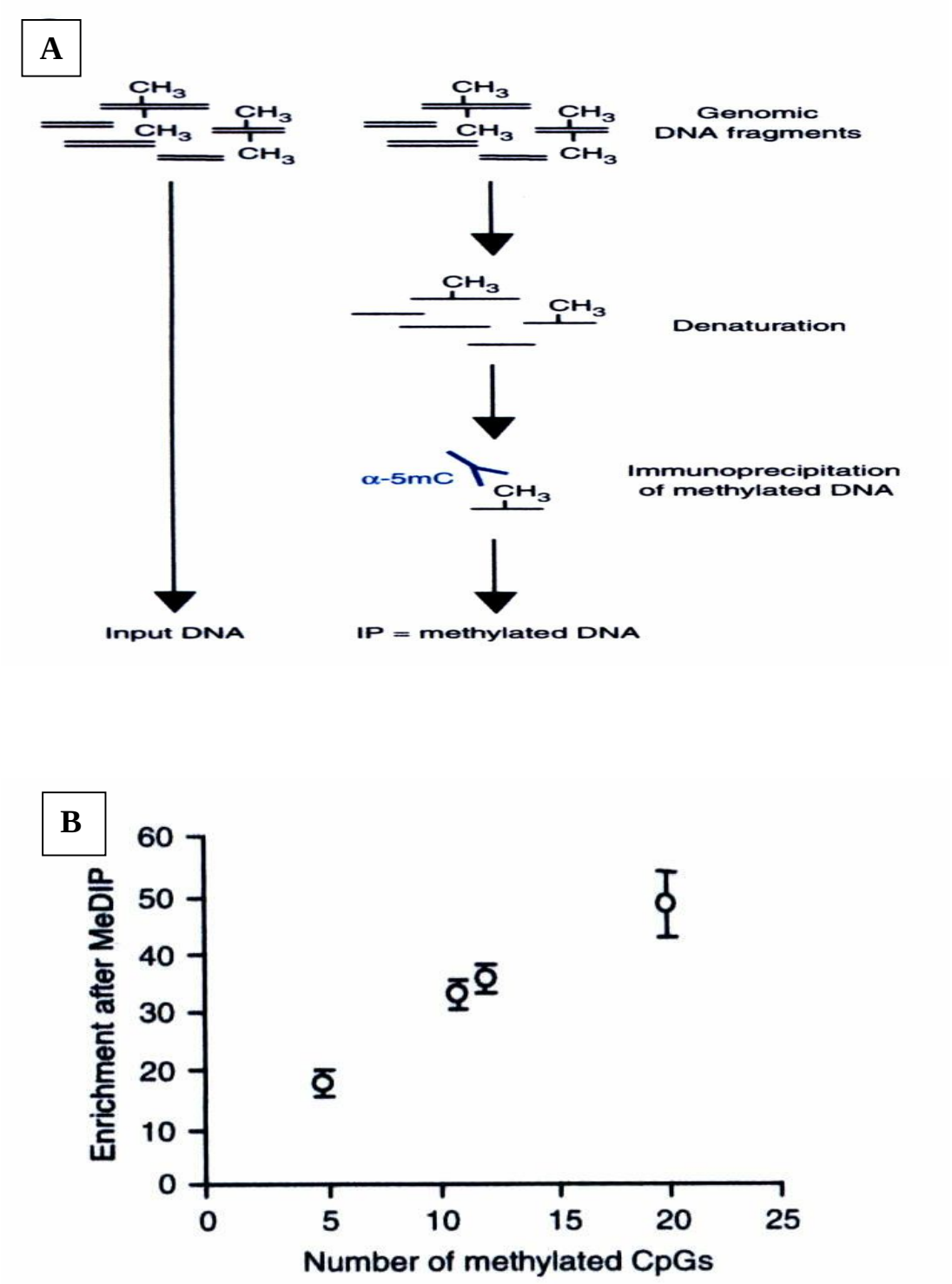


Figure 1.21: Methylated DNA enrichment by MeDIP
(A) Denatured genomic DNA is hybridised with the anti-5-methylCytosine antibody (α-5mC) and methylated DNA is isolated by immunoprecipitation. (B) Correlation between enrichment and the number of methylated Cytosines on four DNA sequences, created by the restriction of *ALU* repeat DNA sequences. Enrichment increases linearly with the increasing number of methylated Cytosines. (Taken from Weber *et al*, 2005).

1.7.4.2 DNA methylation profiling by (MSRE) analysis

MSRE analysis is the simplest method of DNA methylation analysis; this is achieved by use of MSREs to cleave at specific restriction sites, which contain CpG dinucleotides. The most commonly used MSRE is *Hpa* II, which has the restriction site CCGG, and does not cleave at its restriction site where the internal Cytosine is methylated. MSRE digestion of DNA can be coupled with either conventional or real-time PCR to detect the methylation of specific CpG sites (figure 1.22).

A comprehensive study of allelic methylation was carried out by Yamada and colleagues, they assessed 149 CGIs on Chromosome 21, containing the *Hpa* II restriction site and measured the methylation status of peripheral blood cell DNA (Yamada *et al*, 2004). The study showed that it is possible to use a methylation-sensitive restriction enzyme to detect differentially methylated sequences, and to that the technique is adaptable for high through put analysis.

One limitation of MSRE analysis is that the CpG in question must lie within a restriction enzyme site and in the case of *Hpa* II this is only the applicable to 3.9% of CpGs on non-repetitive sequences in the human genome (Weber *et al*, 2005). Another limitation is that MSRE analysis can only be used to identify completely methylated or completely unmethylated DNA sequences, and any incompletely methylated sequences cannot be assessed accurately.

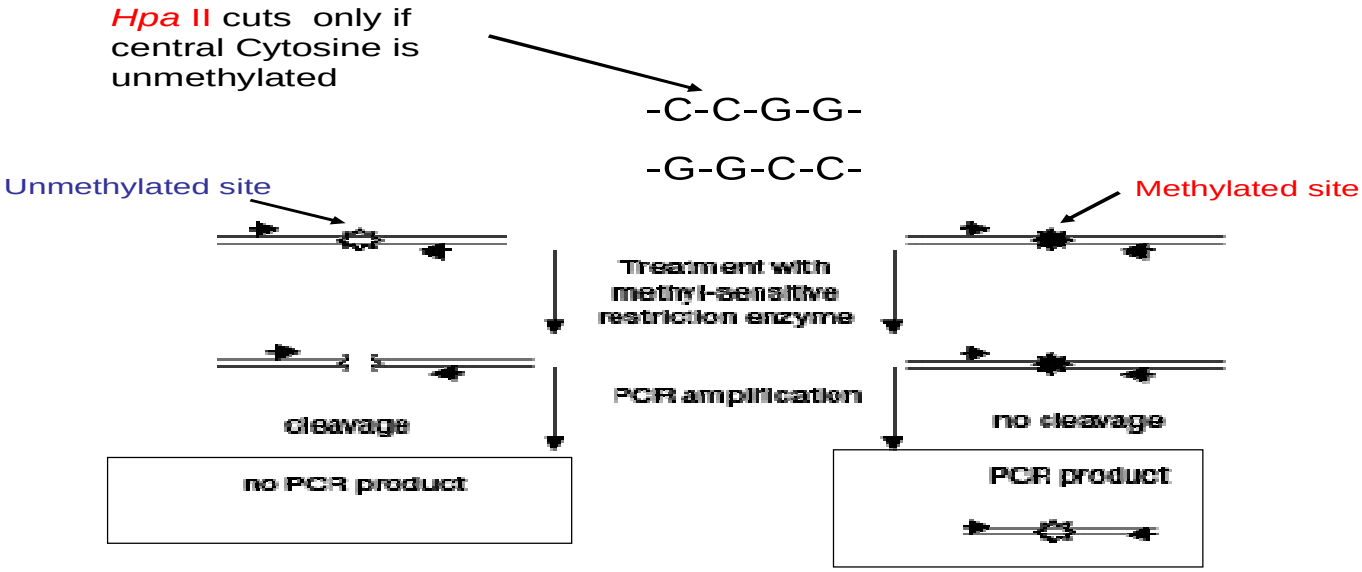


Figure 1.22: Methylation-sensitive restriction endonuclease analysis of DNA
Shows the methylation-sensitive restriction enzyme *Hpa* II and its cleavage site CCGG. In the presence of an unmethylated Cytosine *Hpa* II cleaves its restriction site. However, in the presence of a 5-MethylCytosine *Hpa* II does not cleave its restriction site. Following PCR of methylated and unmethylated DNA fragments only methylated DNA fragments will produce a product

1.7.4.3 Bisulfite conversion of DNA and associated techniques

There are a number of techniques which rely on the chemical modification of DNA in order to detect methylation. The *Sss* I acceptance assay was one of the first techniques developed for the measurement of methylated DNA. *Sss* I DNA methyltransferase is an orphan DNA methyltransferase, which recognises any CpG and uses S-adenosylmethionine (SAM) as a methyl donor in CpG methylation. Using tritium labelled SAM the number of unmethylated CpGs can be directly assessed by scintillation measurement of the pooled sample of DNA. However this technique does not provide an accurate measurement of individual CpGs (Wu *et al*, 1993).

The *Sss* I acceptance assay was followed by the development of chloroacetaldehyde assay, which was an assay for genome wide methylation, which relied on the detection of the intensely fluorescent ethenocytosine derivative of 5-methylCytosine, after bisulfite conversion of DNA. Methylation was quantified by fluorimeter readings. However, the chloroacetaldehyde assay, still lacked the sensitivity to assess individual CpGs (Dahl and Guldberg, 2003).

These methods have been superseded by bisulfite conversion of DNA. Treatment of genomic DNA with sodium bisulfite effectively deaminates unmethylated Cytosine residues to Uracil under specific conditions, where 5-methylCytosines are deaminated at a very slow rate (Frommer *et al*, 1992). Bisulfite converted DNA differs from genomic DNA, as unmethylated Cytosine is converted to Uracil, whereas methylated Cytosines are retained; this causes the 2 strands of bisulfite DNA to be non-complementary.

The bisulfite conversion of DNA is based on the deamination of Cytosine to Uracil. Genomic DNA is prepared by sonication, shearing, or digestion with restriction enzymes. The DNA is then denatured with sodium hydroxide and treated with a concentrated solution of sodium bisulfite at pH 5. The DNA is then desalted and desulfonated with sodium hydroxide, before being neutralised, desalted, and finally dissolved in water for use (figure 1.23) (Grunau *et al*, 2001).

Bisulfite converted DNA can then be amplified by bisulfite-specific PCR primers for (BS-PCR); or by methylation or unmethylation-specific PCR (MS-PCR or US-PCR). Bisulfite-specific primers (BSPs) hybridise with bisulfite converted DNA, the primers do not include CpG sites, and the assumption is made that single Cytosines are transformed to Uracil and incorporated as Thymine during PCR (Li and Dahiya 2002).

Methylation specific primers (MSPs) hybridise with methylated bisulfite converted DNA; primers include CpG sites which are assumed to retain their Cytosines. Unmethylation-specific primers (USPs) hybridise with unmethylated bisulfite converted DNA (i.e. the transformed DNA); primers include CpG sites in which the

Cytosines are assumed converted to Thymine. This is because after PCR with Taq DNA polymerase Uracil is incorporated into the PCR product as Thymine.

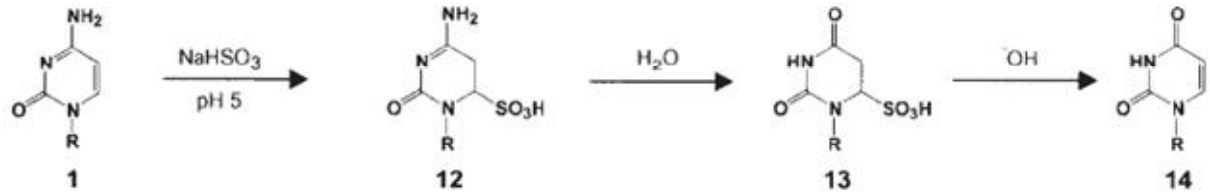


Figure 1.23: Sodium bisulfite conversion of Cytosine to Uracil
Treatment of Cytosine (1) with sodium bisulfite under acidic conditions (pH 5) yields 5,6-dihydroCytosine-6-sulfonate (12). This product is susceptible to hydrolytic deamination at C4 and is easily converted to 5,6-dihydroUracil-6-sulfonate (13). After desulfonation by treatment with NaOH, the final product is Uracil (14). (Taken from Shiraishi *et al*, 2002).

Bisulfite conversion can be combined with a number of techniques in order to effectively assess DNA methylation. Techniques include bisulfite sequencing, combined bisulfite restriction analysis (COBRA) and melt curve analysis. COBRA relies on the detection of a fluorescent tag on one of the primers after bisulfite-specific PCR (BS-PCR); restriction of the BS-PCR product allows the detection of different fragments from that product which indicate its methylation profile. One of the most comprehensive studies on DNA methylation profiling was undertaken at the end of 2006. Using bisulfite sequencing of genomic DNA, high resolution methylation profiles were obtained for 12 different tissue types including: leukocyte subsets CD4+ and CD8+ and placenta on chromosomes 6, 20 and 22. This was one of the first major attempts to assess DNA methylation at the genomic level. In total 2524 amplicons on 873 genes were analysed on chromosomes 6, 20, and 22; including 5' untranslated regions, intronic regions, exonic regions and transcription factor binding sites (Eckhardt *et al*, 2006).

Approximately 27% of loci were hypomethylated (<20% methylation), and 42% of loci were hypermethylated (>80% methylation), with 30% of CpGs methylated in a heterogeneous pattern (20-80% methylation). This study is an important landmark, showing that bisulfite sequencing can be used to produce high throughput methylation profiles of genomic regions at the resolution of single nucleotides. The data generated from this study is an important resource, which in the context of

NIPD could be used to provide targets for candidate biomarkers (Eckhardt *et al*, 2006).

Restriction enzyme digestion can be used to reveal DNA methylation dependent sequence differences in bisulfite-specific PCR products. This procedure is termed COBRA and can be used to analyse the fractions of methylated and unmethylated DNA from a BS-PCR product (Xiong and Laird, 1997). The sequence conversion of a C → T can lead to creation of a new restriction site, as well as the retention of the Cytosine can lead to the retention of a restriction site. In a mixed sample of methylated and unmethylated DNA molecules, the methylated and unmethylated fraction can be determined by COBRA, to give a direct measurement of the percentage of each fraction, by the detection of a fluorescent tag in the primer of each product by a genetic analyser (Xiong and Laird, 1997).

Another technique termed melt curve analysis (MCA) can be performed after BS-PCR. The PCR product can be melted over a temperature gradient in order to assess the level of DNA methylation. Because G:C base pairs between complementary DNA strands contain 3 hydrogen bonds and A:T base pairs have 2 hydrogen bonds, DNA sequences with a higher G:C content have a higher melting temperature (denaturing temperature) Tm than sequences with a higher A:T ratio. Thus BS-PCR is followed by temperature gradient melting of DNA. Methylated sequences have a higher Tm than unmethylated sequences as they retain their G:C content (Guldberg *et al*, 2002). Most real-time PCR machines such as the BIORAD I-cycler carry out post-amplification melt profile analysis as standard.

In review MeDIP is a useful technique when combined with CGH or microarray analysis, it can provide methylation patterns for entire chromosomes in a single experiment, which can be used to target regions of interest. MSRE analysis is a simple technique, which is amenable to high throughput of techniques such as 96 well PCR and can also be used to provide information on methylation at the level of a single CpG. Bisulfite conversion is a laborious technique, as is not easily amenable to high through put analysis; however, its use in high throughput methods can yield high resolution information on selected regions of the genome.

1.8 Summary of current biomarker discovery for NIPD

1.8.1 Recent developments in cell-free DNA markers for non-invasive prenatal diagnosis

Putative epigenetic markers for cff DNA have been discovered on chromosomes 18, 3 and 21, using the model system first proposed by Chim and colleagues in 2005 (Chim *et al*, 2005; Chan *et al*, 2006 and Chim *et al*, 2008). On the basis of the early discovery of the *MASPIN* marker, and by use of the model system I discovered several novel candidate epigenetic biomarkers on Chromosome 21 (Old *et al*, 2007). However it is unlikely that any of the cff DNA markers discovered so far either by the Lo group or by my own research could form part of a non-invasive diagnostic test.

So far I have discovered that after testing CEBs on plasma DNA from pregnant women (Old *et al*, 2007), the methylation signature of placenta DNA is not maintained throughout pregnancy (see chapter 8). In addition to this the markers published by the Lo group are unlikely to be used in diagnostic test as they are not true biomarkers for first trimester plasma DNA. The methylation signature of placenta DNA for the *MASPIN* marker is significantly different to that of maternal blood cell DNA at the third trimester. However, the two tissue types are less easily distinguished during the first trimester, with completely unmethylated placenta DNA making up a smaller proportion of circulating cff DNA in the first trimester (Chim *et al*, 2005).

The discovery of an additional potential marker of cell-free fetal DNA was published in 2006. Hypermethylation of the *RASSF1A* gene is associated with an increased risk of many types of cancer and is located on chromosome 3. *RASSF1A* was published as a universal fetal DNA marker epigenetic biomarker by Chan and colleagues (Chan *et al*, 2006); but has been shown to be variable in it's specificity for cff DNA (data presented by Aitcha Ait Soussan at SAFE PhD Workshop, Bologna 2008).

The latest proposed markers on Chromosome 21 are *PDE9A* and *U-CGI137* markers have been also only been shown to match the epigenetic profile of placenta DNA in the third trimester. So far they have not been shown to be concordant during either the first or second trimesters, also the paucity of methylated CpG sites in the *U-CGI137* putative marker is noted by the authors (Chim *et al*, 2008).

Having conducted my own research using the model system and by testing candidate epigenetic biomarkers with plasma DNA from pregnant women (see chapter 8), one conclusion which can be drawn is that the model system is not appropriate for the discovery of epigenetic biomarkers as the peripheral blood leukocyte DNA and term placenta DNA do not accurately represent the DNA populations present in the plasma of pregnant women (see chapter 9).

1.8.2 Recent developments in cell-free fetal RNA markers for non-invasive prenatal diagnosis

On 04/06/2008 Sequenom a US company, which has purchased the intellectual property rights with regard to cell-free fetal nucleic acids, issued a press release. The release stated that from a small study of 200 plasma samples from pregnant women that 100% of trisomy 21 cases were detected with 0% false positive results. The company also reported that the test offered coverage of 93% (for the US Caucasian population). In a blind study out of 180 samples 3 trisomy 21 samples were detected with 0% false positive results, and of a smaller blind study 7/21 trisomy 21 cases were detected with 0% false positive results (www.sequenom.com).

These results look promising, however it is still desirable to develop DNA based NIPD as it relies on much simpler and cost effective technology, to give an even more rapid response from sample collection to result; as the measurement of cff DNA does not rely on reverse transcription to be assessed. Also using cff DNA would eliminate the small risk of misdiagnosis caused by cases of undetected confined placental mosaicism.

1.9 Aims

The main aim of the project is to develop fetal specific DNA markers for the detection of cff DNA sequences originating from fetal Chromosome 21 in the maternal circulation. This will require the use molecular techniques in order to distinguish epigenetic modifications (CpG methylation) between cff DNA and cfm DNA. This will be achieved by the identification of a number of sites of differential methylation, between term placenta genomic DNA and adult female leukocyte genomic DNA as candidates for epigenetic biomarkers.

Once candidate epigenetic biomarkers have been identified, further testing and validation using plasma DNA samples from pregnant and non-pregnant women will be required.

To determine the utility of analysing DNA methylation in the circulation, the ratio of methylated to unmethylated CpG DNA will be assessed in plasma DNA. This is in order to end any speculation as to the preferential clearance of unmethylated CpG DNA plasma. This work will require the measurement of the allelic ratio, of an imprinted DNA sequence, in order to determine if either methylated or unmethylated CpG DNA are cleared preferentially.

Further to the suitability of using DNA in the circulation as a biomarker, the relative abundances of specific DNA sequences will be assessed by real-time PCR, in order to determine if the general assumption that all DNA sequences are represented equally in the circulating genome (Chan *et al*, 2005 b). The PCR efficiency of each sequence will be analysed to give quantitative data on the relative abundances of each sequence in plasma DNA.

Materials and Methods

Chapter 2

2.1 Reagents and suppliers

Applied Biosystems (Warrington, Yorkshire, UK)

Universal PCR Mastermix (5ml) Cat No. 4304437

BioChain (Hayward, California USA)

Peripheral blood leukocyte genomic DNA (100 µg) Cat No. D1234148
Placenta genomic DNA (100 µg) Cat No. D1234200

Fermentas (Vilnius, Lithuania)

BstU I restriction enzyme (2500 U/ml) Cat No. ER0922
Restriction site 5' CG/CG 3'
pUC19 plasmid DNA (50 µg) Cat No. SD0061
Taq I restriction enzyme (10,000 U/ml) Cat No. ER0671
Restriction site 5' T/CGA 3'

Taq DNA polymerase (5U/ml) Cat No. EP0402

First Link (Birmingham, UK)

Raw human female plasma (HIV/HCV Negative) (25-10ml) self-declared non-pregnant (special order)
Raw human male plasma (HIV/HCV Negative) (10 ml) (special order)

Fisher Scientific (Leicester, UK)

Sodium Hydroxide, analytical grade reagent (500g) Cat No. S/4920/53

Helena Biosciences (Tyne and Wear, UK)

Agarose, molecular biology grade reagent (500g) Cat No. 8201-07

Invitrogen (Abingdon, Oxfordshire, UK)

DEPC-treated H_2O, pyrogen-free (100ml) Cat No. 75-0024

Platinum Taq polymerase kit (5u/ml) Cat No. 10966-034

New England Biolabs Ltd (Hitchin, Hertfordshire, UK)

BamH I (20,000 U/ml) Cat No. R0136L
Restriction site 5' G/GATCC 3'

Hha I (20,000 U/ml) Cat No. R0139L
Restriction site 5' GCG/C 3'

Hpa II (10,000 U/ml) Cat No. R0171L
Restriction sites 5' C/CGG 3'

Promega Ltd (Southampton, UK)

pGEM-T Easy Vector System Kit (20 reactions) Cat No. A1380

QIAGEN Ltd (Crawley, West Sussex, UK)

QIAamp DNA Blood Maxi kit (50 reactions) Cat No. 51194

QIA-quick Gel Extraction kit (50 reactions) Cat No. 28704

QIAquick PCR purification kit (50 reactions) Cat No. 28104

QIAprep spin miniprep kit (50 reactions) Cat No. 27104

Sigma-Aldrich Ltd (Gillingham, Dorset, UK)

Ethidium bromide (10mg/ml) Cat No. E1510-10ml

Hydroquinone 99% pure (100g) Cat No. H9003-100G

Sodium metabisulfite ACS reagent 97% pure (100g) Cat No. 255556-100G

2.1.1 Department of Biological Sciences media preparation

10 M Ammonium acetate (770g of ammonium acetate in 1 litre H_2O).

LB Ampicillin Agar plate (15g bacto-agar/litre, 100µg Ampicillin/ml) 30ml per plate

LB Medium (10:5:10) (10g bact-tryptone + 5g bacto-yeast extract + 10g NaCl in 1 litre H_2O, adjusted to pH 7.0 with 5M NaOH)

SOC Medium (20g bact-tryptone + 5g bacto-yeast extract +0.5 g NaCl in 950 ml H_2O) shake until dissolved + 10ml 250mM KCL + adjust to pH 7.0 with 5M NaOH + 20ml 1M glucose +5ml 2M $MgCl_2$.

3M Sodium Acetate (408.1g sodium acetate in 800ml H_2O, adjusted to pH 6.5 with dilute glacial acetic acid, adjusted to 1 litre with H_2O.

5x TBE (54g Tris Base + 27.5g boric acid + 20ml 0.5 M EDTA pH 8.0). Adjusted to 1 litre with H_2O.

2.2 General laboratory equipment

Most PCR assays were run using the BIORAD iCycler IQ5 Multicolor Real-Time PCR detection system. In some cases the ABI prism 7000 was used for TAQMAN Real-Time PCR.

Horizon 11.14 and Thermo Classic tanks were used for gel electrophoresis. Gene Genius Bio-imaging system (SYNGENE) was used to view gels.

Plasma samples were separated by centrifugation using a Beckman Ultracentrifuge and DNA was extracted using the Heraeus Megafuge 20R.

2.3 DNA samples

Genomic placenta DNA was supplied by (BioChain, USA) Cat No. D1234200.

Genomic peripheral blood DNA was supplied by (BioChain, USA) Cat No. D1234148.

Plasma DNA from non-pregnant (self-declared) individuals was supplied by (First Link Ltd, Birmingham, UK) by special order.

Plasma DNA from pregnant women was supplied by Dr Manu Vatish (Warwick Medical School). Samples were collected before and after delivery from women undergoing elective surgery for delivery at Warwick University Hospital, under local ethical approval 07/H1210/150.

2.3.1 Plasma storage

Blood samples collected were stored in EDTA vacutainers at 4 ^{0}C until separation of plasma. Plasma was obtained by centrifugation at 1,600g for 10 minutes and at 16,000g for 10 minutes. Plasma was stored at -20 ^{0}C before DNA extraction.

2.3.2 Primers

Primer Name	Chromosome location (NCBI 36.3)	Sequence	Details
GNAS_1_1F	20q13.3	5’ ctt cca aaa agg gac cca tc 3’	Genomic DNA
GNAS_1_1R	20q13.3	5’ agc ctc ctc tct tcc cac tc 3’	Genomic DNA
GNAS_1_5F	20q13.3	5’ gat ttt ttt tgt ttt tat gga ttt agg 3’	Bisulfite DNA
GNAS_1_5R	20q13.3	5’ cta acc aac taa acc tcc tct ctt c 3’	Bisulfite DNA
Chr_14_500_F	14q22	5’ gca aca agg tgc att tag ca 3’	Genomic DNA
Chr_14_500_R	14q22	5’ tta ggc caa tgc aaa gta agc 3’	Genomic DNA
MPH 1F (MASPIN SEQUENCE 1)	18q21.3	5’ tac ttt ttg tgc cac caa cg 3’	Genomic DNA
MPH 1R (MASPIN SEQUENCE 1	18q21.3	5’ cag gac cgg aag gtg aaa ta 3’	Genomic DNA
MPH 2F (MASPIN SEQUENCE 2	18q21.3	5’ tat ttc acc ttc cgg tcc tg 3’	Genomic DNA
MPH 2R (MASPIN SEQUENCE 2)	18q21.3	5’ ctt cca aaa ggc ctc caa c 3’	Genomic DNA
PDE9A F	21q22.3	5’ ggc gtc tga cgt ctc cag 3’	Genomic DNA
PDE9A R	21q22.3	5’ tgg agt tgc agt act tgc tga 3’	Genomic DNA
RASSF1A F	3p21.3	5’ ctc att gag ctg cgg gag ctg gca 3’	Genomic DNA
RASSF1A R	3p21.3	5’ tgg ggt tgc acg cgg tgc cc 3’	Genomic DNA
CpG Negative F	21q22	5’ tcc caa aac cca ata aaa gc 3’	Genomic DNA
CpG Negative R	21q22	5’ ttt ccc cct gac aat ctc tc 3’	Genomic DNA
Beta Actin F	7p15-p12	5’ gcg ccg ttc cga aag tt 3’	Genomic DNA
Beta Actin R	7p15-p12	5’ cgg cgg atc ggc aaa 3’	Genomic DNA
21-58-F	21q22.3	5’ ttc tgt gca act ttc gct tg 3’	Genomic DNA
21-58-R	21q22.3	5’ cgt ctc tca ctc cct gca ct 3’	Genomic DNA
21-58 SHORT F	21q22.3	5’ agg aaa ccc agc gag cag 3’	Genomic DNA

Primer Name	Chromosome location (NCBI 36.3)	Sequence	Details
21-210 SHORT F	21q22	5’ aag gca gga ggt ttc tgt cc 3’	Genomic DNA
21-210 SHORT R	21q22	5’ gaa cac act gcg cgc ttt a 3’	Genomic DNA
1-16 SHORT F	1q21	5’ aga cct ccg gtt tct ggt tt 3’	Genomic DNA
1-16 SHORT R	1q21	5’ gag agc ctg gct aga aaa agg 3’	Genomic DNA
2-28 SHORT F	2p14	5’ ttg gag cta aca gct tct gtc tt 3’	Genomic DNA
2-28 SHORT R	2p14	5’ gaa ggc tcg ggg tga gtc 3’	Genomic DNA
1-50 SHORT F	1q31.2	5’ act gtg gag ccc ggg tct gt 3’	Genomic DNA
1-50 SHORT R	1q31.2	5’ tgg aag ttt cat ccc tgg ag 3’	Genomic DNA
12-42 SHORT F	12q24	5’ gac caa cca gcc tga caa ac 3’	Genomic DNA
12-42 SHORT R	12q24	5’ tca ccg caa cct ccg cct 3’	Genomic DNA
18-29 FV1	18p11.2	5’ gtt agc ttt cct tca gcc gg 3’	Genomic DNA
18-29 RV2	18p11.2	5’ gca ctt ggc gac agc cgg tg 3’	Genomic DNA
21-37-1F	21q21	5’ agt cgc agg aag ctc taa cg 3’	Genomic DNA
21-37-1R	21q21	5’ agt ttg gct gaa agg tca gg 3’	Genomic DNA
21-43-3F	21q22	5’ gcc agg aac tcc agt gag ac 3’	Genomic DNA
21-43-3R	21q22	5’ tct ttg gtt tat ggc gct gt 3’	Genomic DNA
21-107 SHORT F2	21q22.3	5’ cgc att ggt ttc tga tgg a 3’	Genomic DNA
21-107 SHORT R2	21q22.3	5’ gtg gtc ggg tgc tag atg tt 3’	Genomic DNA
18-CR SHORT F	18q21	5’ aag agc gct cca tga agc ca 3’	Genomic DNA
18-CR SHORT R	18q21	5’ gtc ctc act tcc ttg cgt gt 3’	Genomic DNA
22-09 F	22q12.2	5’ ttc tta gga cgg gaa cga tg 3’	Genomic DNA
22-09 R	22q12.2	5’ ggg tac agt tac cgc tca cc 3’	Genomic DNA

Primer Name	Chromosome location (NCBI 36.3)	Sequence	Details
Beta-Globin F	11p15.5	5’ gtg cac ctg act cct gag gag a 3’	Genomic DNA
Beta-Globin R	11p15.5	5’ cct tga tac ca acc tgc cca g 3’	Genomic DNA
Beta Globin 5’ VIC Probe 3’ TAMRA	11p15.5	5’ aag gtg aac gtg gat gaa gtt ggt gg 3’	Genomic DNA
SRY F	Yp11.3	5’ tcc tca aaa gaa acc gtg cat 3’	Genomic DNA
SRY R	Yp11.3	5’ aga tta atg gtt gct aag gac tgg at 3’	Genomic DNA
SRY 5’ VIC Probe 3’ TAMRA	Yp11.3	5’ cac cag cag taa ctc ccc aca acc tct tt 3’	Genomic DNA
21-128 F	21q22.3	5’ tga gag gca ttg caa act aga tg 3’	Genomic DNA
21-128 R	21q22.3	5’ cca gcg tct ttt atc tgc agt tg 3’	Genomic DNA
21-58 MSP F	21q22.3	5’ agg gga gag ggt agg aaa ttt agc gag tag tag ctc 3’	Methylated DNA
21-58 MSP R	21q22.3	5’ gaa aaa aaa acc ccc tac cta cat acc tac ctt acc g 3’	Methylated DNA
AIRE A1 F	21q22.3	5’ tag ttt ttg tta ggg ttt tga gat t 3’	Bisulfite DNA
AIRE A1 R	21q22.3	5’ acc cac caa aac aac tac ctt aa 3’	Bisulfite DNA
MCM3AP F	21q22.3	5’ gaa gtt agg tgg gag gag att tt 3’	Bisulfite DNA
MCM3AP R	21q22.3	5’ act att cca cct tac aac tta aaa ac 3’	Bisulfite DNA
H19 DMR F	11p15.5	5’ gta tag tat atg ggt att ttt gga ggt ttt 3’	Bisulfite DNA
H19 DMR R	11p15.5	5’ taa ata tcc tat tcc caa ata acc c 3’	Bisulfite DNA
MASPIN F	18q21.3	5’ gaa tgg aga tta gag tat ttt ttg tgt tat 3’	Bisulfite DNA
MASPIN R	18q21.3	5’ act tcc aaa aaa cct cca aca tat 3’	Bisulfite DNA
LM18-58A F	18q21	5’ aga cag aca ctt gtc ttg cag aa 3’	Genomic DNA
LM18-58A R	18q21	5’ gtc ctc act tcc ttg cgt gt 3’	Genomic DNA

Table 2.1: DNA oligonucleotide primers
Shows all of the primers used throughout the project, most oligonucleotides were supplied by (VH-BIO, Newcastle, UK) Probes were supplied by (Applied Biosystems California, USA). Primers were designed using the web-based program primer 3 input (Rozen and Skaletsky, 2000). Bisulfite specific primers and methylation specific primers were designed using the web-based program MethPrimer (Li and Dahiya, 2002). Bisulfite specific primers amplify all bisulfite DNA sequences and primer sequences do not contain any CpGs, although the internal amplicon may contain CpG sites. Methylation specific primers amplify bisulfite converted DNA only contain methylated CpGs in their sequences, primers are designed to contain CpGs, which are retained after bisulfite conversion. For clarification only the current primers for each region are given.

2.4 Methods

2.4.1 Extraction of DNA from plasma

DNA was extracted from plasma samples using the QIAamp Blood Maxi kit (QIAGEN, UK); using the purification of DNA from whole blood spin protocol, according to the manufacturer's instructions.

The following alterations were made:

Up to 10 mls of plasma was used as the input as opposed to whole blood.

Step 13 b was chosen to gain maximum yield.

Samples were ethanol precipitated after elution from the column.

2.4.1.1 Ethanol precipitation of plasma DNA

Samples from the DNA Maxi kit are eluted in 1ml.

Each sample was split equally into 2x 1.5 ml eppendorf tubes and labelled accordingly. 50 µl of 3M sodium acetate was added to each tube (to give a final concentration in solution of 3mM). 1ml of (96-100%) ethanol was added to each tube, the tubes were incubated overnight. The tubes were then centrifuged at 13,000 rpm at 4 ^{0}C for 30 minutes and the supernatant was carefully removed with a 1ml pipette. The centrifugation step was repeated for an additional 5 minutes and any remaining supernatant was removed with a 200 µl pipette. The samples were then allowed to air dry in a fume cupboard for 5-10 minutes, and both samples were re-hydrated and recombined in 30 µl in a single tube.

2.4.2 Enzyme restriction

Taq I restriction digestion of bisulfite-specific PCR products for COBRA:

Enzyme mix:

Taq I enzyme 10 µl (100 units)

10 x *Taq I* buffer 15 µl

DDH_2O 25 µl

Final reaction volume Σ 50 µl

(5 µl of enzyme mix is used to digest each PCR product) a mix without enzyme is used as a negative control for digestion. Samples are incubated at 65 ^{0}C for 1 hour).

BamH I digestion for the fragmentation of commercial genomic DNA prior to bisulfite conversion:

Input genomic DNA **x** µl (10 µg)

BamH I enzyme 3.5 µl (70 units)

10 x *BamH* I buffer 10 µl

BSA 1 µl

DDH_2O **y** µl

Final reaction volume Σ 100 µl (sample concentration is 100ng/µl).

Samples are incubated at 37 ^{0}C for 2 hours.

Hpa II or *Hha* I digestion of commercial genomic DNA for screening of biomarkers:

Input genomic DNA **x** µl (10 µg)

Hpa II enzyme 10 µl (100 units) **OR** *Hha* I enzyme 10 µl (200 units) **or both**

10 x buffer 4 10 µl

DDH_2O **y** µl

Final reaction volume Σ 100 µl (sample concentration is 100ng/µl).

Samples are incubated at 37 ^{0}C for 2 hours and the enzyme is inactivated by incubation at 65 ^{0}C for 20 minutes.

Hpa II or *Hha* I digestion of plasma DNA for screening of biomarkers:

Plasma DNA 10 µl
Hpa II enzyme 1 µl (10 units) **OR** *Hha* I enzyme 1 µl (20 units) **or both**
10 x buffer 4 1 µl

Σ 12 µl

Samples were incubated at 37 ^{0}C for 40 minutes and the enzyme was inactivated by incubation at 65 ^{0}C for 20 minutes, negative controls with water instead of enzyme were also used.

2.4.3 PCR

Reagent	Volume (final concentration)
H_2O	12.5 µl
10 x PCR buffer	2.5 µl (1 x PCR buffer)
10 mM dNTPs	1.0 µl (0.1mM)
25 mM $MgCl_2$	3.0 µl (3mM)
2.5 µm Forward primer	2.0 µl (0.2 µM)
2.5 µm Reverse primer	2.0 µl (0.2 µM)
Taq DNA polymerase (Fermentas)	1.0 µl (5 units)
Total volume of reagents	24.0 µl
Genomic DNA input (100 ng/µl)	1.0 µl
Final total volume	**25.0 µl**

Table 2.2: Standard PCR protocol for biomarker screening
Standard PCR cycle used was 35 cycles of (95 ^{0}C 30s, 60 ^{0}C 30s, 72 ^{0}C 1min) followed by a holding step at 4 ^{0}C.

Reagent	Volume (final concentration)
2 x SYBR green mastermix	12.5 µl
2.5 µm Forward primer	2.0 µl (0.2 µM)
2.5 µm Reverse primer	2.0 µl (0.2 µM)
H_2O	3.5 µl
Total volume of reagents	20.0 µl
Genomic bisulfite converted DNA input (20 ng/µl)	5.0 µl
Final total volume	**25.0 µl**

Table 2.3: Nested PCR conditions for methylation-specific PCR of the *AIRE* 58 biomarker
PCR cycle conditions for the pre-amplification step: 1 cycle of 95 ^{0}C for 15 minutes, 32 cycles of (95 ^{0}C 30s, 60 ^{0}C 30s, 72 ^{0}C 30s) 1 cycle of 55 ^{0}C for 1 min and (melting gradient 80 cycles of 55 ^{0}C + 0.5 ^{0}C every 10s). Samples were diluted 200 fold and the PCR amplification step was repeated for the second step of the nested PCR.

Reagent	Volume (final concentration)
2 x SYBR green mastermix	12.5 µl
2.5 µm Forward primer	1.0 µl (0.2 µM)
2.5 µm Reverse primer	1.0 µl (0.2 µM)
H_2O	5.5 µl
Total volume of reagents	20.0 µl
Genomic bisulfite converted DNA input (20 ng/µl)	5.0 µl
Final total volume	**25.0 µl**

Table 2.4: Methylation-specific PCR conditions for optimal *AIRE* 58 biomarker MSPs
PCR cycle conditions: 1 cycle of 95 ^{0}C for 15 minutes, 40 cycles of (95 ^{0}C 30s, 69 ^{0}C 30s, 72 ^{0}C 30s) 1 cycle of 55 ^{0}C for 1 min and (melting gradient 80 cycles of 55 ^{0}C + 0.5 ^{0}C every 10s).

Reagent	Volume (final concentration)
2 x TAQMAN Universal mastermix	12.5 µl
10 µm Forward primer	0.75 µl (0.3 µM)
10 µm Reverse primer	0.75 µl (0.3 µM)
10 µm VIC Probe- TAMRA	0.5 µl (0.2 µM)
H_2O	5.5 µl
Total volume of reagents	20.0 µl
Plasma DNA input (20 ng/µl)	5.0 µl
Final total volume	**25.0 µl**

Table 2.5: TAQMAN Real-time PCR on ABI PRISM 7000 platform for plasma DNA testing
PCR cycle conditions: 1 cycle of 50 ^{0}C for 2 minutes, 1 cycle of 95 ^{0}C for 15 minutes 45 cycles of (95 ^{0}C 15s, 60 ^{0}C 1 min).

2.4.4. Gel electrophoresis

All gels were 1.5 % agarose with 10 µl of 10mg/ml ethidium bromide (Sigma, UK) per 100 ml of 1x TBE solution. All gels were run in either 11.14 Horizon gel tanks or Thermo Classic tanks.

2.4.5 Bisulfite conversion of plasma DNA

Due to the problematic nature of performing bisulfite conversion of DNA on small amounts of DNA, our laboratories standard operating protocol for the bisulfite conversion of plasma DNA is given in detail below.

DNA is first extracted from plasma using the QIAamp DNA Blood maxi kit (QIAGEN, UK). The final volume of each plasma DNA sample was assessed (30 µl) and 1/9th of the total volume of 3M sodium acetate was added to each sample (3.3 µl). 5.3 µl of 0.471ng/µl pUC 19 plasmid DNA (**2.5 µg**) was added to each sample for use as carrier DNA. 2 volumes of ethanol plus an additional 30 µl were added to each sample, each sample was mixed by inverting the tube 6 times. Each sample was incubated at -20 ^{0}C overnight.

NEXT DAY

Before starting one heating block was set to 65 ^{0}C and one programmable heating block was set to 97 ^{0}C. Samples were centrifuged at 13,000 rpm at 4 ^{0}C for 30 minutes. The supernatant was carefully removed with a 1ml pipette and re-spun to remove any remaining supernatant. DNA pellets were allowed to air dry in a fume cupboard for 10-20 minutes; and were re-suspended in 30 µl of H_2O. Small volumes of 10M NaOH and 3M NaOH were prepared. 3.66 µl of 3M NaOH were added each sample, and each sample was incubated on the hot bars at 65 ^{0}C for 20 minutes.

SODIUM BISULFITE PREPARATION (IN ABSENCE OF LIGHT)

1x 50 ml falcon tube and 1 x 15 ml falcon tube are covered in foil. In the 15 ml falcon: 0.22g of hydroquinone was diluted in 10mls of dd H_2O and dissolved by inverting gently about 50 times. The final concentration of hydroquinone was 0.2 M. In the 50 ml falcon: 5.7g of Sodium metabisulfite was diluted in 8mls of dd H_2O, and mixed by inverting 6 times. 500 µl of fully dissolved 0.2 M hydroquinone was added to the sodium metabisulfite; add 400 µl of 10M NaOH to the sodium metabisulfite solution to equilibrate the solution to pH 5.0.

The solution was mixed by inverting twice, and the pH using an indicator strip. The solution was allowed to stand for 5 minutes to allow a precipitate to form; the precipitate must not be disturbed when aliquoting. Sample tubes were covered with foil and 400 µl of sodium bisulfite solution was added to each tube, the solution was mixed by inverting each tube 6 times. Samples were placed on a heating block at 95 ^{0}C for 7-10 minutes, and switched to 55 ^{0}C for a total incubation time of 4 hours. 1.66mls of buffer PB were added to labelled 15 ml Falcon tubes. An Eppendorf with dd H_2O is laced on hot bars now at 95 ^{0}C for use later.

When the bisulfite conversion reaction was completed the DNA solution was added to the falcon tubes with PB buffer. Each sample was then processed by QIAquick PCR column according to the manufacturer's protocol. Each sample is eluted in 100 µl Hot H_2O. 11 µl of 3M NaOH was added to each sample and mixed by pipetting; each sample was then incubated at 37 ^{0}C for 20 minutes to denature the DNA. 167 µl of 5M ammonium acetate was added to each sample and mixed by pipetting and incubated at room temperature for 5 minutes. 834 µl of absolute ethanol was added

to each sample and mix by inverting the tubes. Samples were incubated overnight at -20 ^{0}C.

NEXT DAY

Each sample was centrifuged at 13,000 rpm for 15-30 minutes at 4 ^{0}C to remove the supernatant and the centrifugation was repeated for 5 minutes to allow any further supernatant to be removed.

All DNA pellets were re-suspended in 30 µl of hot H_2O.

2.4.6 Cloning and sequencing

PCR product cloning was done using the pGEM –T Easy vector systems (Promega, USA) according to the manufacturer's instructions.

2.4.6.1 Analysis of clones using Clustal

Samples were analysed in the applied biosciences sequence scanner program (ABI sequence scanner 1.0) are converted to text files. The clones added to master notepad files and save in text format, any anti-sense strand clones were converted to the sense strand by use of (Clone Manager suite software). Alignments for each set of clones were made and written as postscript files using the Clustal X programme v1.81.

2.4.7 Sonication of DNA

DNA was pulse sonicated in 4 rounds of 15 seconds with rest periods on ice to ensure DNA did not over heat, using a standard sonicator (MSE, UK).

Results

Chapter 3

Imprinting of cell-free plasma DNA

3.1 Aims

The main routes by which cell-free plasma DNA (cfp DNA) enters the circulation is by apoptosis of leukocytes and exocytosis of nuclei by maturing erythrocytes. In order to use DNA methylation as a tool for the detection of specific DNA, it must first be determined that the methylation profile of cellular DNA is maintained in cfp DNA. The data presented in this chapter describes the relative abundance a specific sequence of cfp DNA. The sequence lies within the *GNAS*1 gene promoter. *GNAS*1 is an imprinted DNA sequence, in that it contains sequences with parent-of-origin-specific allelic methylation patterns. The region studied also contains a single nucleotide polymorphism (SNP) to allow allelic identification (figure 3.1). Work on the cloning and sequencing presented in this chapter was completed in collaboration with Kamalit Chatha.

3.2 Introduction

DNA is released by cells in the circulation during the process of apoptosis of leukocytes and by exocytosis of nuclei by maturing erythrocytes (Bianchi *et al*, 2004; Bischoff *et al*, 2005; Koide *et al*, 2005; Kolialexi *et al*, 2004 and Sekizawa *et al*, 2000). In order for DNA methylation to be a viable tool for the identification of fetal specific DNA sequences; methylated and unmethylated CpG DNA sequences present in circulation must maintain the same methylation profiles as their cellular counter parts. One factor which may affect the levels of methylated and unmethylated CpG DNA in the circulation is the immune system. It has been well documented that in certain pathological conditions bacterial unmethylated CpG DNA acts as a pathogen associated molecular pattern (PAMP) and can stimulate an immunogenic response (Agren *et al*, 2006; Ahmad-Nejad *et al*, 2002; Cornelie *et al*, 2004 and Yoshikawa *et al*, 2006). Bacterial CpG motifs are short sequences consisting of a CpG flanked by two purines at the 5' end and two pyrimidines at the 3' end. These sequences occur 20 times less frequently in mammalian genomes (Hemmi *et al*, 2000). The presence of unmethylated CpG sequences in the plasma can also stimulate auto-immune responses, as is the case in the auto-immune disease systemic lupus erythematosus (SLE) (Januchowski *et al*, 2004). Bearing these publications in mind it is a reasonable hypothesis that unmethylated CpG DNA

sequences may stimulate an immune response and interact with immune-cells in the circulation.

It has also been suggested that due to chromatin remodelling in nucleosomal DNA, methylated DNA may persist in the circulation for longer periods than unmethylated DNA (Herman, 2004). Taken together these two scenarios may mean that there is a bias towards the presence of methylated CpG DNA and clearance of CpG unmethylated DNA in the circulation. Very little research has been carried out to assess the levels of methylated CpG and unmethylated CpG DNA in the human circulation. The work here has carefully assessed the levels of the methylated maternal allele of *GNAS*1 and the unmethylated paternal allele of *GNAS*1 (figure 3.1).

The experiments presented here attempt to address the situation of methylated and unmethylated CpG DNA sequences for unique single copy genes and does not address the special circumstances of high repeat sequences such as ALU, which is heavily methylated or the inactivated X Chromosome.

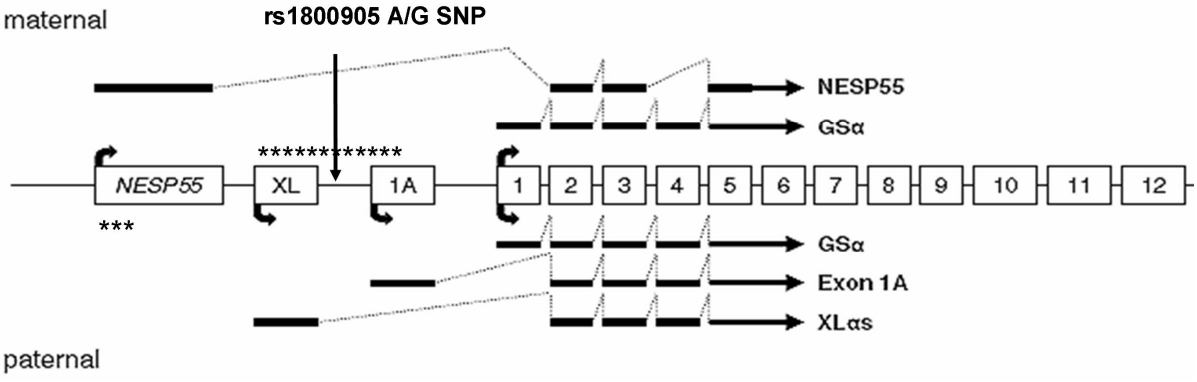


Figure 3.1: Imprinted XL region of the *GNAS1* gene 20q13
Shows a diagram of the various overlapping *GNAS1* gene transcripts. The imprinted XL region is shown; which is methylated on the maternal allele and unmethylated on the paternal allele, methylated regions are indicated by asterisks. Also shown is the SNP rs1800905 (NCBI 36.3) (Taken from Eggermann *et al*, 2007).

Using the XL imprinted region of the *GNAS1* gene it is possible to assess the level of methylated and unmethylated CpG DNA in plasma, by measuring the allelic ratio of the maternal (methylated) and paternal (unmethylated) alleles. The SNP rs1800905 can be used to distinguish between the maternal and paternal alleles.

The data from these imprinting experiments will provide a measurement of CpG methylated and CpG unmethylated DNA, which is present in plasma, if either DNA form is present in excess, there could be detrimental implications for the use of epigenetic DNA biomarkers.

3.3 Status of imprinted cell-free DNA in plasma

DNA was extracted from the plasma of normal healthy individuals (First Link, U.K) using the QIAamp DNA blood Maxi Kit (QIAGEN, U.K). The DNA was amplified by PCR and then each PCR product was divided equally into two samples and one sample was digested with *FnuD* II (Fermentas, USA) the other was not digested. This allowed each sample to be identified as either a heterozygote or a homozygote for the SNP re 1800905, by comparison of digested and undigested OCR products by gel electrophoresis. (figure 3.3). In the presence of the G SNP the site CGCG is completed allowing digestion of the PCR product; however in the presence of the A SNP the sequence CGCA does not cleave when digested with *FnuD* II, in the case of heterozygote the digested PCR product should show two bands of approximately equal intensity (figure 3.3). The full length of the *GNAS1* amplicon is 192 bp, with fragments of 134 bp and 58 bp when the SNP is cleaved, the amplicon sequence and SNP site are shown (figure 3.2).

CAAAGAAGCCCACCCGCCGTGACGCGTCCCCGGAGTCCC**CTTCCAAAAAGGGACCCATC**CCCATCCGG
CGTCACTAATGGAGGACGCCGTCCAGATTCTCCTTGTTTTCATGGATTCAGGTTAGTTGCCCACCGCT
AAACTGGGGAGCCTGAGGGCGGTGTGGGAGCAGCGC**R**GGTGGAAAGGAGGTGAGAAGGAAAGGCAGGT
CAGGGGCG**AGTGGGAAGAGAGGAGGCT**CAGCTGGTCAGCCTGGGATCGGGGGTCAGGGTGAGGCGGCG

Figure 3.2: *GNAS1* amplicon sequence map
Showing the *GNAS1* amplicon as defined by the genomic primers (in bold), all CpGs are highlighted in red, with CpGs internal to the bisulfite-specific PCR amplicon in yellow. The SNP rs1800905 is an A to G transtion marked as **R** in the sequence and is adjacent to a 5' C which may form an additional SNP -dependent CpG. The bisulfite converted amplicon is shorter, the position of the bisulfite primers are given (underlined).

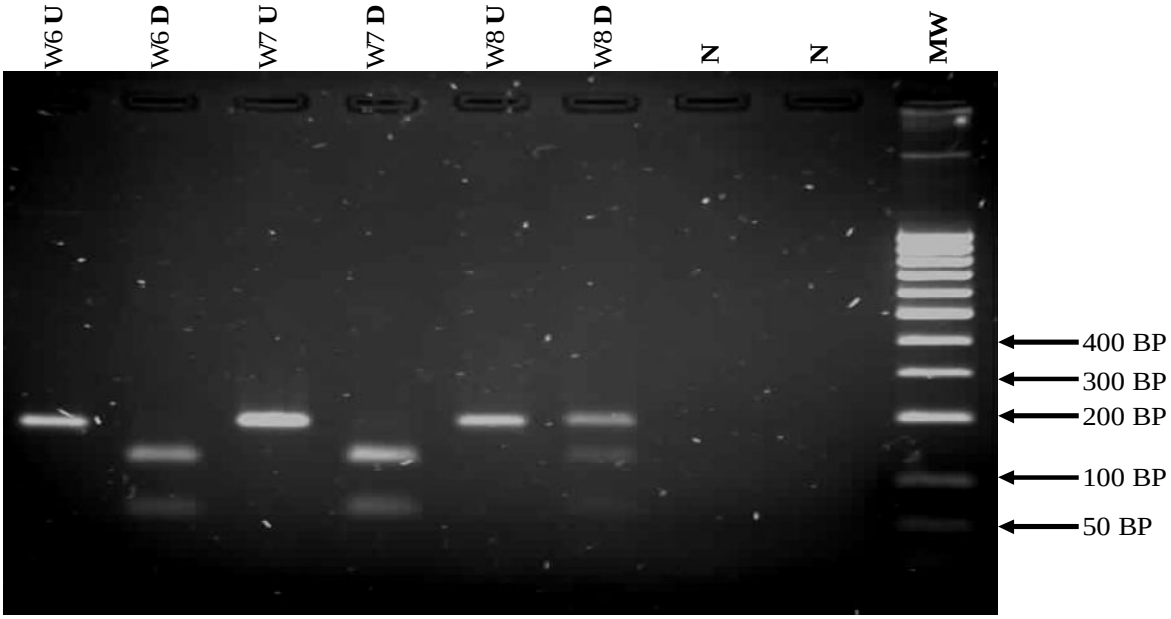


Figure 3.3: Gel electrophoresis image of digested and undigested *GNAS1* PCR products of plasma DNA
Sample W6 is homozygous for the G allele; this is shown by the complete digestion of the PCR product by *FnuD* II in lane W6 **D** when compared to the undigested product labelled W6 **U**. Sample W7 is also homozygous for the G allele; this is shown by the complete digestion of the PCR product by *FnuD* II in lane W7 **D** when compared to the undigested product labelled W7 **U**. Sample W8 is heterozygous containing both an A allele and a G allele; as it is partially digested as shown in lane W8 **D**, when compared to W8 **U.** This is indicated by the presence of two bands, were the G allele is cleaved, but the A allele remains undigested. The lane marked MW shows 100 BP molecular weight DNA ladder (Geneflow, UK).

Cellular controls of placenta DNA and peripheral blood leukocyte DNA were selected. Leukocyte DNA was used to represent the cellular fraction of blood as a broader indicator of the situation of cells in the peripheral blood, and could be used as a direct comparison to plasma DNA. Placenta DNA was chosen as a cellular control in order to identify the methylation status of DNA in a solid tissue. By coincidence all of the four control samples assessed were heterozygotes (figure 3.4).

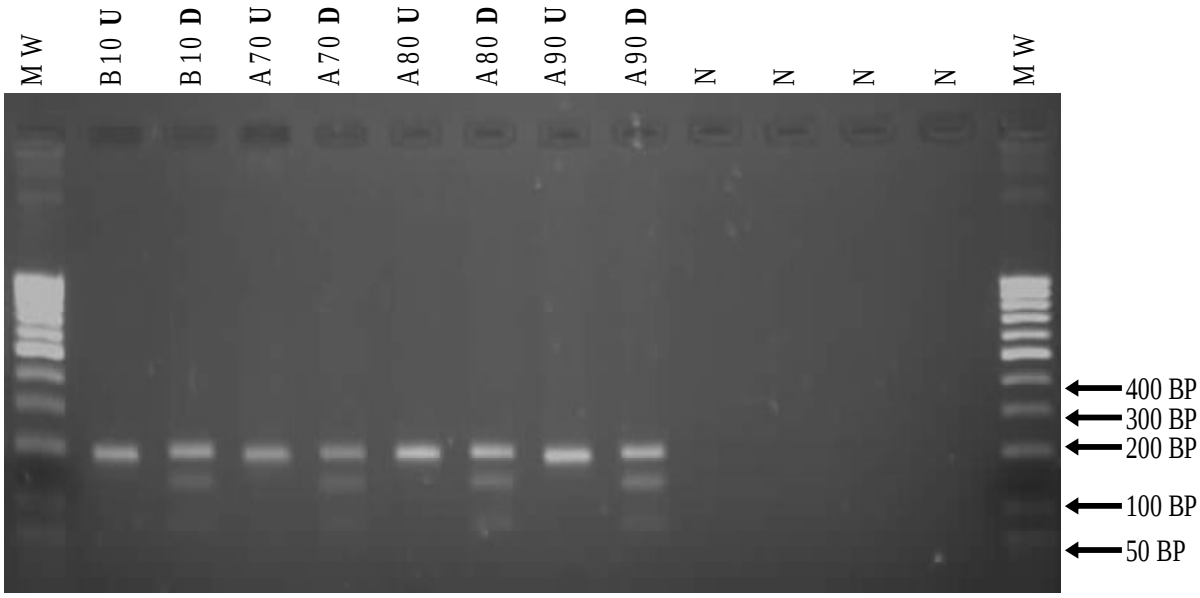
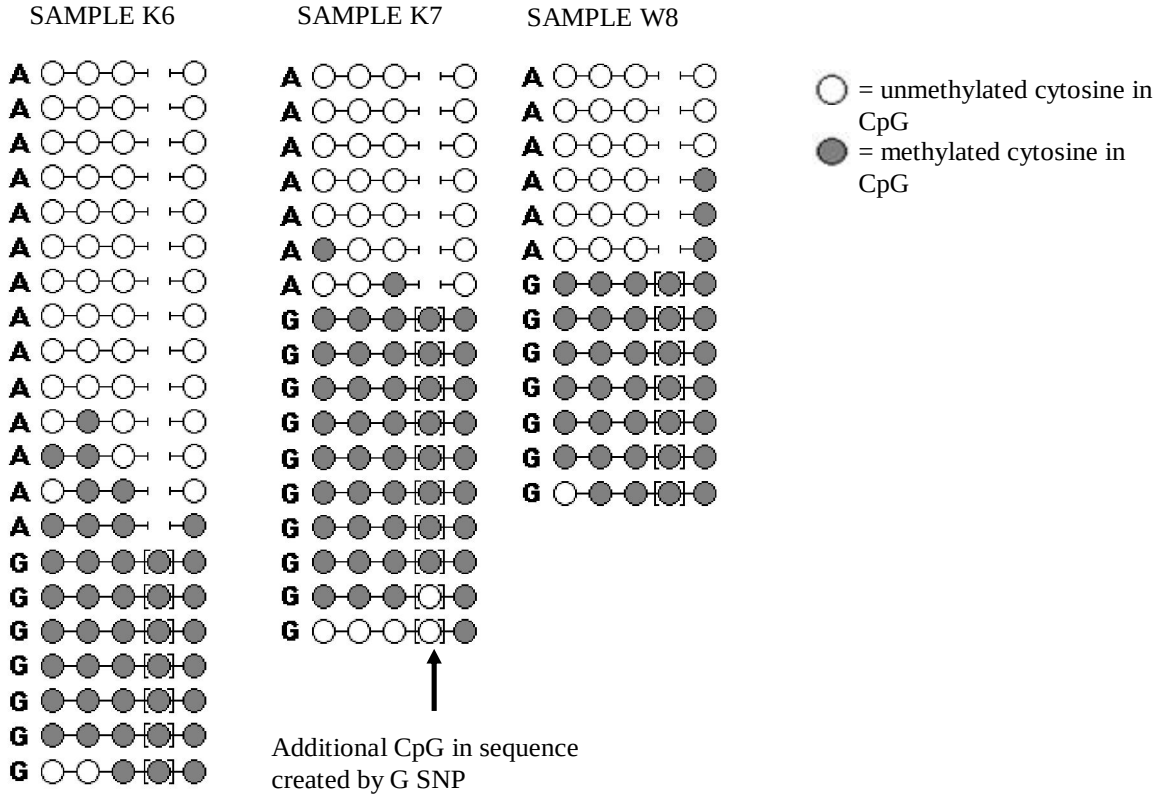


Figure 3.4: Gel electrophoresis image of digested and undigested PCR products of leukocyte and placenta DNA for the *GNAS*1 region
Samples prefixed B10 and A70 are leukocyte DNA samples (BioChain, USA); samples prefixed A80 and A90 are placenta DNA (BioChain, USA); prefixes denote the first three characters of each sample's respective batch number. By coincidence all of the controls were heterozygous for the *GNAS*1 SNP. Lanes B10 **D**, A70 **D**, A80 **D** and A90 **D** all show two bands after digestion with *FnuD* II, when compared to their undigested counterparts in lanes; B10 **U**, A70 **U**, A80 **U** and A90 **U** respectively. The four lanes marked N are PCRs which were run with a non-template control. The lanes marked MW denote the 100bp DNA molecular weight ladder used as a reference for product size (Geneflow, UK).

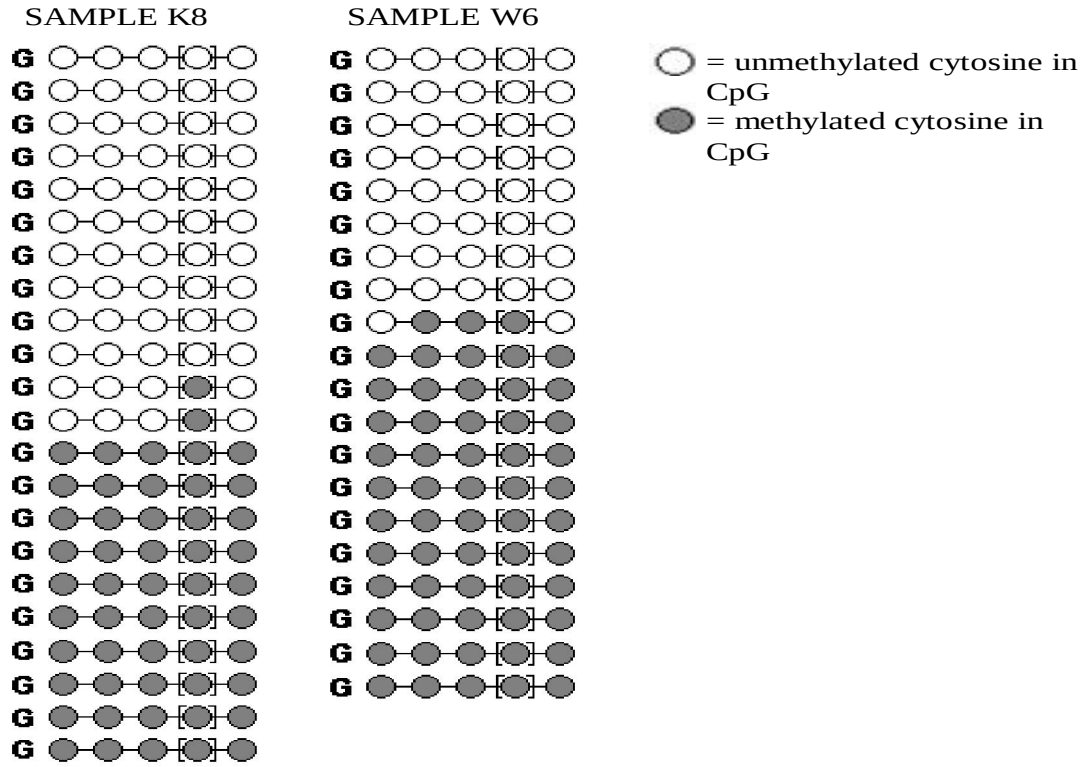
To more accurately assess relative abundances of methylated and unmethylated *GNAS1* sequences, plasma samples were collected from normal healthy individuals and DNA was extracted using the QIAamp DNA blood maxi kit (QIAGEN, U.K). The plasma DNA was then bisulfite converted and the *GNAS*1 region was amplified with bisulfite-specific primers. The bisulfite-specific PCR product was then cleaned on a column (QIAGEN, U.K) and inserted into a vector using the pGEM T-Easy Vector system (Promega, U.S.A). Colonies were selected and the sequencing was processed by molecular biological services (Warwick, U.K), using the ABI prism 3130xl genetic Analyser. The sequences were then analysed with ABI sequence scanner software and aligned using Clustal X v1.81. Overall the ratio of methylated to unmethylated alleles in cfp DNA was 1:1 ($p=0.773$), as assessed by heterozygous samples, the homozygous control samples also had a ~1:1 ratio ($p=0.756$) of

maternal to paternal alleles and acted as a control for any PCR bias which may have arisen due to the presence of the SNP (figure 3.5).

A



B



C

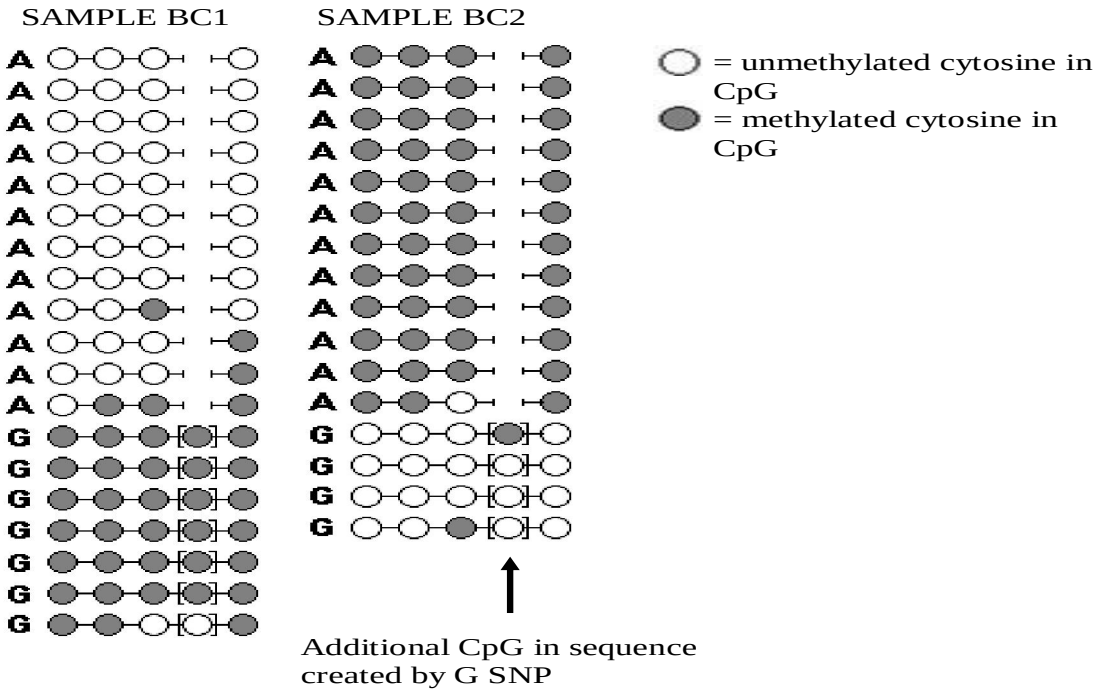


Figure 3.5: Sequencing data from the BS-PCR products of *GNAS1*
In this diagrammatic representation of the *GNAS1* sequence the paternal allele is unmethylated and the maternal allele is methylated. Each CpG in the sequence runs 5' to 3' left to right and is represented as either a clear circle (with unmethylated cytosine) or a filled circle (with methylated cytosine). The transition from A to G allele creates an additional CpG site only present in the G allele. In (A) all of the plasma samples K6, K7 and W8 are from heterozygous donors. The K6 sample has 7 maternal alleles and 14 paternal alleles. The K7 sample has 10 maternal alleles and 7 paternal alleles. The W8 sample has 7 maternal alleles and 6 paternal alleles. In (B) the plasma DNA samples K8 and W6 are homozygous for the G allele. The K8 sample has 10 maternal alleles and 12 paternal alleles. The W6 sample has 12 maternal alleles and 8 paternal alleles. In (C) both of the peripheral blood leukocyte genomic DNA samples are heterozygous. The BC1 sample has 7 maternal alleles and 12 paternal alleles. The BC2 sample has 12 maternal alleles and 4 paternal alleles.

3.4 Allelic ratio controls

As an additional control the relative abundance of a non-imprinted allele was measured, a unique single copy sequence on chromosome 14 was selected which contains an SNP site rs1742500 (NCBI build 36.3). This SNP has a C/T transition and where the C is present a GCGC *Hha* I site is created, whereas in T alleles the site

GTCG is not cleaved by *Hha* I. After the initial testing with restriction a *Hha* I assay to select heterozygotes (data not shown); the plasma DNA samples of several heterozygotes were cloned and sequenced to assess the ratio of C:T alleles and determine the relative abundance of alleles at the non-imprinted locus (figure 3.6). The ratio of C alleles to T alleles was approximately 1:1 showing that there was no specific bias for either the sequences selected.

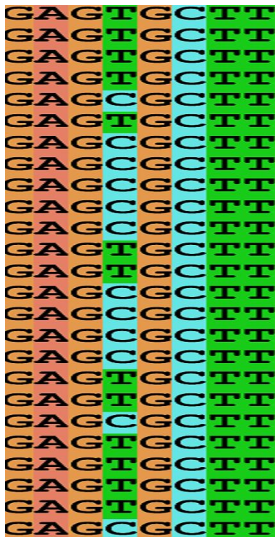


Figure 3.6: Clustal X data, showing sequencing for the Chromosome 14 SNP with short flanking sequence
All of the plasma samples represented here are pooled from donors whom were heterozygous for the rs1742500 SNP; the whole sample set is shown. Of the 25 colonies selected, 12 contain the C allele and 13 contain the T allele. Showing that the ratio of C:T is approximately 1:1.

3.5 Discussion

The imprinted XL locus on *GNAS1* was analysed in the plasma DNA of normal healthy individuals by bisulfite sequencing (figure 3.5). Alleles were distinguished by their methylation profile and as an additional control by their SNP (in heterozygous cases only). In the three plasma samples which were identified as heterozygous (figure 3.5 A) the combined number of maternal (methylated) alleles was 24 and the combined number of paternal (unmethylated) alleles was 27. These data show that the ratio of methylated to unmethylated alleles is approximately equal. The homozygous plasma DNA samples K8 and W6 (figure 3.5 B) also had an almost equal number of methylated to unmethylated alleles, with 22 methylated maternal alleles and 20 unmethylated paternal alleles. These are the same as the heterozygous results and show that there was no PCR bias caused by the presence of the SNP. By including the homozygous plasma DNA samples into the data set, the

number of maternal (methylated) alleles is 46 and the number of paternal (unmethylated) alleles is 47; which is approximately 1:1.

In the adult peripheral-blood leukocyte genomic DNA controls (figure 3.5 C) samples BC1 and BC2 have a combined number of 19 maternal (methylated) alleles and 16 paternal (unmethylated) alleles which is also approximately 1:1. This shows that the relative abundances of methylated to unmethylated DNA sequences in are equal in both the cellular compartment of whole blood and the in the cfp DNA.

As a further control the allelic ratio of another SNP on Chromosome 14 was also assessed on the same set of plasma DNA samples. This genomic locus was assessed as an additional control to ensure that the allelic ratio of a non-imprinted region was also equal and that the results obtained from the analysis of the *GNAS1* region were not by chance (figure 3.6).

These data suggest that there is no specific mechanism for the clearance of CpG unmethylated DNA in the plasma of healthy individuals. The data also suggests that there is no mechanism by which CpG methylated DNA is preferentially retained in the plasma of healthy individuals.

A number of plasma DNA samples were analysed, four of which were found to be heterozygous by methylation-sensitive restriction assay, of the 25 alleles assessed 12 alleles carried the C SNP and 13 alleles had the T SNP (figure 3.6).

These data show for the first time that methylated and unmethylated CpG DNA present in the circulating plasma are equally abundant in normal healthy individuals. This is of great importance in the field non-invasive diagnosis and is crucial for the development of any diagnostic test which relies on the detection of differentially methylated cell-free plasma DNA; such as those being developed in cancer diagnosis and prenatal care. The equal abundance of the DNAs indicates the absence of mechanisms that differentially affect the release or clearance of unmethylated or methylated DNA in plasma. Were differential mechanisms evident, then variation in these mechanisms between individuals, or variation in the immune status or other

factors, could have a drastic consequence for the use of epigenetic DNA biomarkers in plasma.

One interesting use of imprinted genes potentially could be the development of fetal-specific DNA markers for non-invasive prenatal diagnosis. Taking the *GNAS1* locus studied here as an example; by knowing that the maternal allele is methylated in the fetus, it may be possible to detect the switch in methylation patterns if the imprinted allele is combined with an SNP. In cases where the mother is heterozygous for a specific SNP on an imprinted gene and the fetus inherits the opposite SNP from the father whom is homozygous, fetal DNA may be detected by allelic ratio (figure 3.7).

In cases where the heterozygousity of a specific SNP (within an imprinted gene) is approaching 50 % the allelic ratio of fetal DNA could be measured in approximately 6% of pregnancies. This is an unattractive prospect for most diagnostic procedures however; one exciting possibility is the potential use of this technique to determine whether it is the maternal or paternal X chromosome which is inactivated in the placenta or whether X-inactivation is random. If X-inactivation is parent-specific (preferably paternal) this could potentially provide the ability to distinguish the fetal X chromosome from the non-invasive prenatal diagnosis.

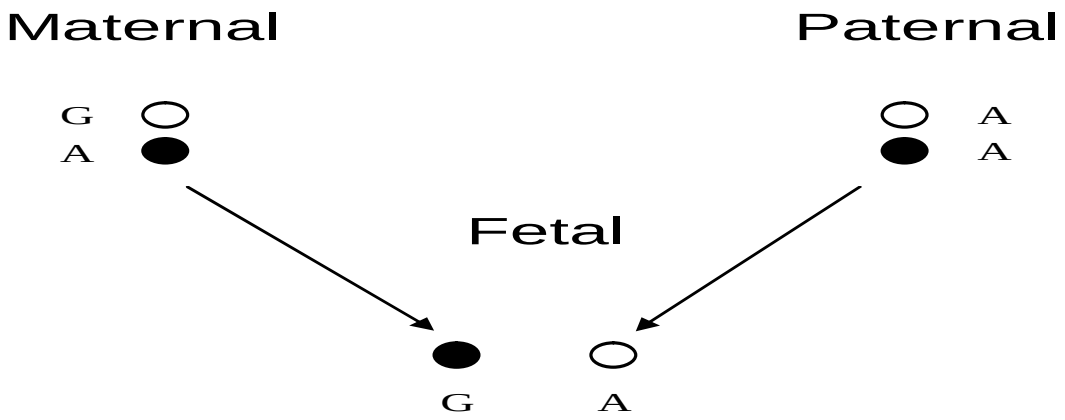


Figure 3.7: Schematic diagram showing the combination of imprinting and SNP analysis for the detection of fetal DNA
In cases where imprinted loci coincide with a SNP, which has a high rate of heterozygousity, fetal DNA which is maternally imprinted can be distinguished completely, but only in the small percentage of cases where the mother is heterozygous for the SNP and the father is homozygous, and the paternal allele inherited is differential to the maternal.

Such a situation could be assessed by the bisulfite sequencing of maternal plasma DNA and by measurement of allelic ratio, as both of the fetal alleles, in the case scenario shown (figure 3.7) are not present in the maternal genome.

Results

Chapter 4

Representation of DNA sequences in cell-free plasma

4.1 Aims

Cell-free DNA in the circulation is derived from a range of cells which are predominantly haematopoietic in origin (Lui *et al*, 2002). It is proposed that apoptosis is the main route by which stable DNA is released by cells into the circulation (Bischoff *et al*, 2005; Sekizawa *et al*, 2000). However the precise mechanism of DNA release has yet to be elucidated and may also include the release of DNA by normal healthy lymphocytes (Anker and Stroun, 2006; Stroun *et al*, 2001). What has been assumed up until this point is that genomic DNA sequences, which are released by cells into the circulation, are equally abundant, at least in the case of single copy genes, so far it has already been demonstrated that *ALU* repeat DNA sequences are more abundant than single copy sequences in plasma and serum (Stroun *et al*, 2001).

In this chapter, by the use of carefully calibrated real-time PCR, the relative abundance of several DNA sequences in plasma is assessed. The measurement of the relative abundance of different DNA sequences is important; if the relative abundance of each sequence is variable between individuals, this may mean that methods such as relative chromosome dosage (RCD) which rely on comparison of DNA abundances are not suitable for the accurate measurement of cell-free plasma DNA.

4.2 Introduction

Cff DNA and cfm DNA have been assessed in many studies and by many methodologies; studies have suggested that cff DNA is placental in origin and cfm DNA is haematopoietic in origin (Alberry *et* al, 2007; Bianchi 2004; Lui *et al.* 2002). Other studies have assessed the increase in cff DNA concentration throughout gestation and its rapid clearance after delivery (Birch *et al*, 2005; Galbiati *et al*, 2005; Lo *et al*, 1998 a and Smid *et al*, 2006). Very few studies have been carried out to assess the relative abundance of different cell-free DNA sequences present in plasma. One investigation into the genomic representation of plasma DNA in pregnant women has been conducted by method of comparative genomic hybridisation (CGH) analysis, coupled with fluorescent *in situ* hybridisation (FISH)

analysis (Chan *et al*, 2005 b). Cell-free plasma DNA was obtained from 10 pregnant women, and amplified by degenerate oligonucleotide primed (DOP)-PCR; the products of DOP-PCR were then labelled with biotin-16-dUTP and digoxigenin by nick translation and stained with either avidin conjugated fluorescein isothiocyanate (FITC) or IgG-conjugated tetramethylrhodamine isothiocyanate (TRITC).

This was followed by counterstaining with an anti-fade solution containing 4,6, diamino-2-phenylindole (DAPI). Digital images of the stained chromosomes were captured and averaged fluorescence ratios along each chromosome were calculated using a digital imaging system (Chan *et al*, 2005 b).

The results of the study were that each sample of plasma DNA from pregnant women contained an approximately even genomic representation of chromosomes. This is a low resolution technique for the approximate measurement of DNA, which averages fluorescence of DNA sequences along the length of a chromosome (Chan *et al*, 2005 b). By contrast the use of real-time PCR to assess the relative abundance of specific cfp DNA sequences is an accurate and direct method of addressing the question of equal representation of cell-free DNA sequences in the circulation.

4.3 Importance of sonicating genomic DNA before calibration of primers

Carefully calibrated real-time PCR has been widely used to determine the amount of fetal (male) or total DNA in plasma during pregnancy (Birch *et al*, 2005). However, the abundance of cell-free plasma DNA sequences has not been accurately assessed relative to other sequences present. Here, 6 sets of primers were calibrated using a dilution series of two preparations of leukocyte DNA (labelled B10 and A70 after the first three characters of their respective batch numbers). Importantly, the leukocyte DNA was sonicated to give a broad size distribution with approximately 80% of the mass between 2.0 kb and 200bp; similar to the size distribution of plasma DNA (Puszyk *et al* 2008b accepted for publication; Chan *et al*, 2004 and Li *et al*, 2004). Not sonicating the DNA before PCR can cause inaccurate quantification of

DNA sequences; as high molecular weight DNA is not representative of cfp DNA and does not amplify so efficiently (figure 4.1; figure 4.2).

It is of note that in most of the published studies on quantification of cfp DNA, sonication has not been employed, leading to a potentially substantial inaccuracy. The extent of the inaccuracy is likely to be sequence-dependent. Thus, the presence of inverted repeat sequences in long fragments would be expected to influence the PCR efficiency and hence apparent quantification (figure 4.1).

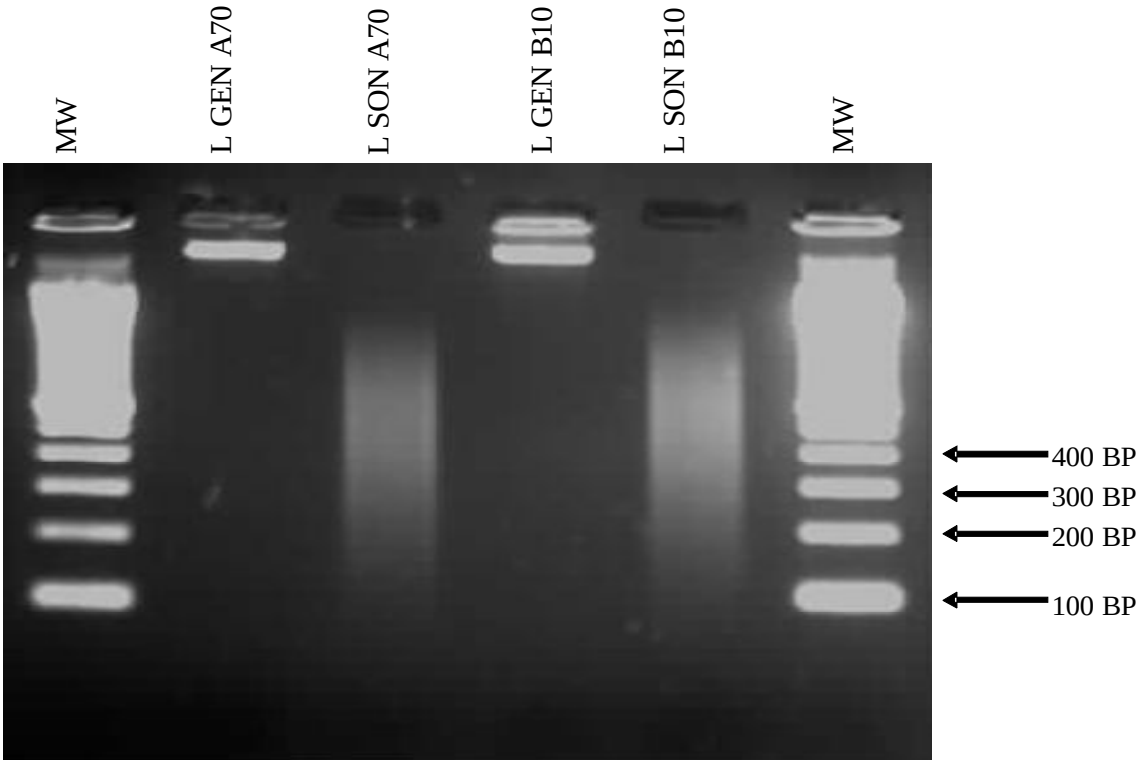


Figure 4.1: Size fractionation data for sonicated and non-sonicated leukocyte DNA

Non-sonicated genomic peripheral blood leukocyte DNA for sample A70 is shown in the lane marked LGEN A70. Non-sonicated genomic leukocyte DNA for sample B10 is shown in the lane marked LGEN B10. Sonicated leukocyte DNA for sample A70 is shown in the lane marked LSON A70. Sonicated leukocyte DNA for sample B10 is shown in the lane marked LSON B10. The size distribution of the sonicated DNA samples is approximately 2000 -100 bp and is representative of the size distribution of DNA in plasma. Lanes marked MW show banding for a 100bp molecular weight DNA ladder (Invitrogen).

4.4 Calibration of individual primer pairs with sonicated and non-sonicated leukocyte DNA in serial dilution

One of the most important things to consider when developing any diagnostic, which relies on the use of cell-free plasma DNA, is whether or not all of the DNA sequences are in equal abundance, and whether or not these abundances vary between individuals. The relative abundances of 6 DNA sequences were assessed in the plasma DNA from healthy individuals. The primer efficiencies were also assessed; and the relative abundance of efficiency-matched primer pairs were found to vary (data not shown).

As well as testing the relative abundances of the specific DNA sequences, the 6 primer pairs were also used as an additional test for the preferential clearance of CpG unmethylated DNA or the retention of CpG methylated DNA in plasma. Three primer pairs were designed to amplify DNA regions known to be methylated; these were *MPH1*, *MPH2* and *PDE9A* (table 4.1). Two primer pairs were designed to amplify DNA regions known to be unmethylated in plasma DNA these were *RASSF1A* and *Beta-Actin*; a third primer pair was designed to contain no CpG sites and therefore be completely unmethylated, this was the CpG-Negative primer set (table 4.1). Using non-sonicated genomic leukocyte DNA to calibrate the 6 primer sets yields a range of data with cycle threshold (Ct) values for real-time PCR reactions ranging from 25.25 Cts to 35.08 Cts (figure 4.2).

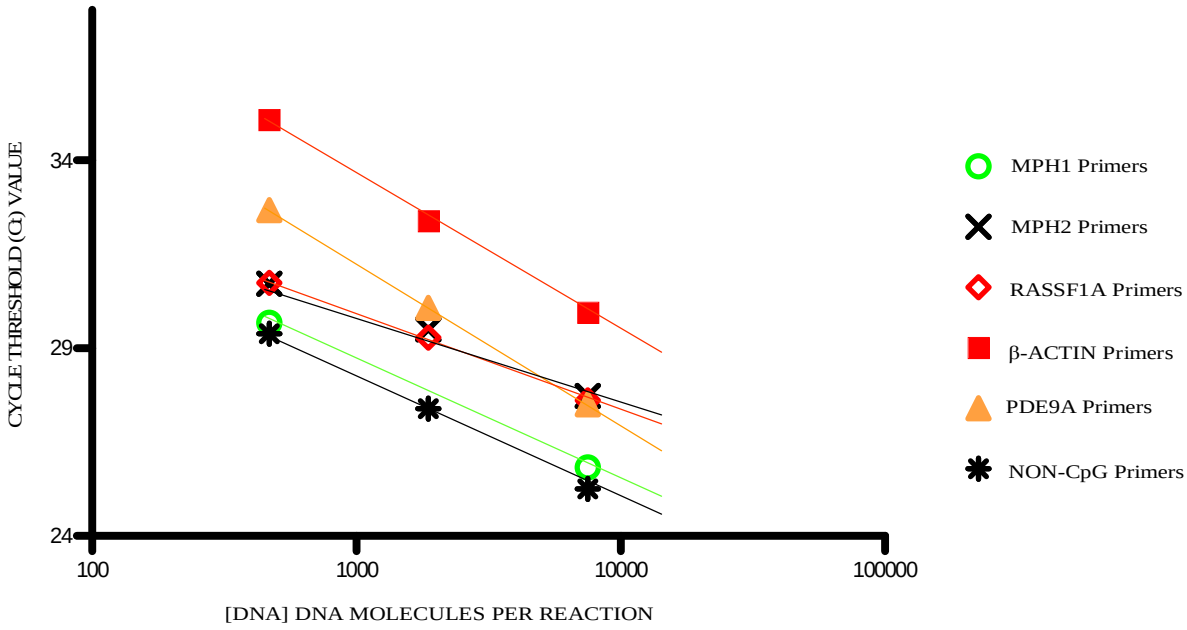


Figure 4.2: Real-time PCR amplification data for primers using a non-sonicated leukocyte DNA template

PCR cycle threshold (Ct) each reaction is plotted against the number of DNA molecules, to give a line of best fit for each primer set. Three sets of primers were designed for DNA regions known to be methylated in plasma: MPH 1 primers (○), MPH 2 primers (**x**) and PDE9A primers (▲). Three sets of primers were designed for regions known to be unmethylated in plasma DNA β-Actin primers (■); RASSF1A primers (◊) and primers which spanned a region not containing any CpG sites, Non-CpG primers (*).

Using sonicated DNA to mimic the size distribution of cell-free plasma DNA changes the dynamic of the Real-time PCR reaction; improving the efficiency of all reactions. The range of cycle threshold (Ct) values for the matched dataset ranging from 24.38 Cts to 32.86 Cts which is lower than the non-sonicated DNA control (figure 4.2; figure 4.3).

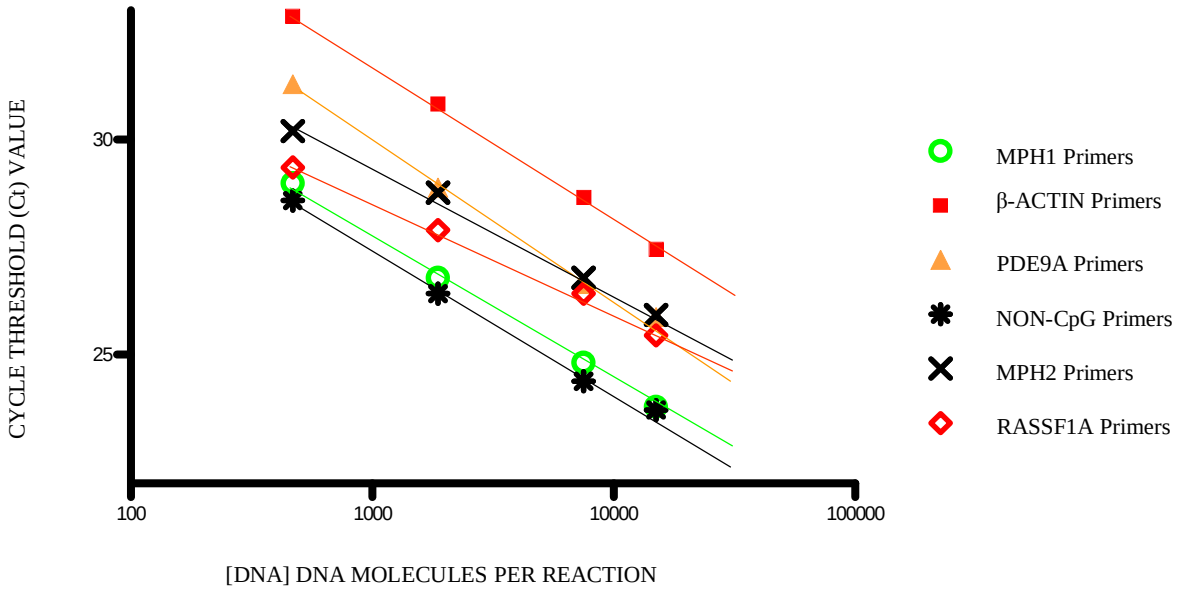


Figure 4.3: Real-time PCR amplification data for primers using a sonicated leukocyte DNA template

PCR cycle threshold (Ct) each reaction is plotted against the number of DNA molecules, to give a line of best fit for each primer set. Three sets of primers were designed for DNA regions known to be methylated in plasma: *MPH* 1 primers (○), MPH 2 primers (**x**) and PDE9A primers (▲). Three sets of primers were designed for regions known to be unmethylated in plasma: β-Actin primers (■); RASSF1A primers (◊) and primers which spanned a region not containing any CpG sites, Non-CpG primers (*). As this DNA is more representative of the size distribution of DNA in plasma the lines of best fit can be used to determine the relative DNA concentration of plasma DNA samples using the Ct values obtained after real-time PCR amplification with a specific primer pair.

Table 4.1: Abundances of DNA copies in cell-free plasma DNA

Primer pairs for the 6 DNA regions listed were used to measure the abundance of cell-free DNA in the plasma of 6 normal, non-pregnant individuals (identified as donors 772-777). The sequences *MPH1* and *MPH2* are regions which, are defined by two primer pairs that lie within the differentially methylated region of the *MASPIN* promoter, which is completely methylated in plasma. *PDE9A* primers define a DNA region which is completely methylated in plasma. The *RASSF1A* primers define a region which is unmethylated in plasma. *CpG-NEGATIVE* primers define a region which lacks the CpG dinucleotide and is therefore unmethylated. The *BETA-ACTIN* region is unmethylated in plasma. Abundances are rounded to nearest value of 10 molecules/ml.

4.5 Primers selected for Real time PCR

		Plasma donor identifier (sex) Plasma DNA copies (molecules/ml)					
DNA region (primers 5'-3', amplicon size, methylation status in plasma)	chromosome	**772 (f)**	**773 (m)**	**774 (m)**	**775 (m)**	**776 (m)**	**777 (m)**
MPH1 (TACTTTTTGTGCCACCAACG , CAGGACCGGAAGGTGAAATA, 95 bp, methylated)	18	930	1430	920	1130	1720	920
MPH2 (TATTTCACCTTCCGGTCCTG , CTTCCAAAAGGCCTCCAAC, 113 bp, methylated)	18	450	590	360	680	2600	330
PDE9A (GGCGTCTGACGTCTCCAG , TGGAGTTGCAGTACTTGCTGA, 94 bp, methylated)	21	310	360	190	380	1270	230
RASSF1A (CTCATTGAGCTGCGGGAGCTGGCA , TGGGGTTGCACGCGGTGCCC, 109 bp, unmethylated)	3	250	180	130	210	1200	210
CpG-NEGATIVE (C21-ORF-128) (TCCCAAAACCCAATAAAAGC, TTTCCCCCTGACAATCTCTC, 83 bp, unmethylated)	21	510	750	550	620	1980	450
BETA-ACTIN (GCGCCGTTCCGAAAGTT , CGGCGGATCGGCAAA, 137 bp, unmethylated)	7	100	100	100	80	200	50

4.6 Discussion

Each of the 6 primer sets were calibrated using a dilution series of two preparations of adult peripheral-blood leukocyte genomic DNA (figure 4.3). The leukocyte DNA was sonicated to give a broad size distribution with approximately 80 % of DNA fragments between 2.0 kb and 200 bp in length, similar to the size distribution of plasma DNA (figure 4.1). Data for both non-sonicated (figure 4.2) and sonicated (figure 4.3) DNA serial dilutions is given, however only data from the sonicated DNA serial dilution was used to calibrate plasma DNA abundances as it more closely mimics the size distribution of plasma DNA. Each of the 6 sequences defined by the primer pairs are unique in the human genome and gave a negligible primer dimer product (data not shown).

The data show that there is no consistent effect from DNA methylation on the relative abundance of DNA sequences in the plasma; as *MPH*2 and *PDE9A* sequences (both methylated sequences) are neither markedly more or less abundant than *RASSF1A* and *CpG-negative* sequences (both unmethylated sequences) (table 4.1).

It is also observed that the pattern of relative abundances of DNA in plasma is not sex dependent, (although further samples would be required to test this observation). The sample from donor 772 was female and the relative abundances from this plasma source do not differ greatly from that of other samples 774 and 777, which were collected from male donors (table 4.1). What is interesting is the range of variation between the relative abundance of different DNA sequences in plasma. The largest difference in sequence abundance is between *MPH1* and *Beta-Actin* sequences, with *MPH1* sequences detected in up-to 18.4 fold higher abundances than *Beta-Actin* sequences, in plasma from donor 777 (table 4.1).

Additionally unexpected results were observed between *MPH1* and *MPH2* sequences. The two primer pairs *MPH1* and *MPH2* prime for two regions on the *MASPIN* gene, and the sequences which they amplify overlap by 32 bp. It is

reasonable to assume that as the sequences overlap they should be relatively equal in abundance. However, the sequences differ by as much as 2.78 fold in plasma DNA with *MPH1* being more abundant than *MPH2*, although not exclusively as is the case with plasma from donor 776 (table 4.1). This may be an indication of some form of non-random cleavage of DNA, which effects relative abundance in a specific way. Note also that donor 776 has the highest overall amounts of cfp DNA. This may indicate that this sample has undergone post-collection cell lysis.

These data show that DNA sequences circulating in plasma of normal healthy individuals are not equally abundant. This is the first time that the question of the relative abundance of specific DNA sequences in plasma has been addressed, challenging the assertion of the previous study, which relied on the averaged measurement of chromosome fluorescence (Chan *et al*, 2005), that all DNA sequences are represented equally in plasma.

This has important implications for non-invasive diagnostic tests. It may not be possible to accurately measure cfp DNA by use of relative chromosome dosage (RCD) unless the relative abundances show a remarkable level of stability. Thus it may be possible to compare sequences using normalised relative chromosome dosage (NRCD). However, inter-individual variation in the relative plasma DNA levels may ultimately mean that assessment by NRCD or RCD is not possible. Currently the most attractive method for the detection of trisomy using cell-free plasma DNA is the allelic ratio (AR) method as this relies on the assessment of alleles of the same sequence and not the comparison of different sequences. Additionally the smaller number of cff DNA molecules required for AR make it a more attractive technique (Table 1.1) (Old and Stallard, accepted for publication).

Results

Chapter 5

Candidate epigenetic biomarker discovery by methylation-sensitive restriction endonuclease analysis of genomic DNA

5.1 Aims

First proposed by Chim and colleagues upon the discovery of the potential epigenetic marker *MASPIN* (Chim *et al*, 2005), the use of peripheral blood leukocyte DNA and term placenta DNA as surrogates for maternal leukocyte DNA and cell-free fetal DNA, has since become the standard technique for the detection of epigenetic biomarkers. Methylation-sensitive restriction endonuclease (MSRE) analysis is one method by which the presence or absence of a methylated cytosine can be assessed. MSRE analysis of adult peripheral blood leukocyte and term placenta can be used to identify candidate epigenetic biomarkers (CEBs) on Chromosome 21 and other autosomes for use in non-invasive prenatal diagnosis. CEBs discovered by this technique require further characterisation on plasma DNA from pregnant women, in order to confirm results using the model system (see chapter 7).

5.2 Introduction

Chim and colleagues discovered the first potential epigenetic marker for cff DNA on the promoter of the *MASPIN* gene on Chromosome 18, using their novel model system of peripheral blood leukocyte and term placenta DNA (Chim *et al, 2005)*. In addition to this a further candidate epigenetic marker of cff DNA has also been identified on the promoter of the *RASSF1A* gene on chromosome 3 (Chan *et al*, 2006) by the application of the same method. Most recently potential epigenetic biomarkers have also been discovered on Chromosome 21 (Chim *et al*, 2008). Using the model system, DNA sequences on chromosomes 1, 2, 12, 18 and 21 were digested with the methylation sensitive restriction endonucleases *Hpa* II, *Hha* I and *FnuD* II; in order to assess the CpG sites within the restriction sites for the presence or absence of methylated cytosine. In total 366 DNA regions were selected by the fulfilment of either one of the three following criteria: 1) Location within or proximal to a gene promoter relative to the transcriptional start site. 2) Location within or proximal to a gene promoter identified as being differential by virtue of expression profile, with data being obtained from a publicly available database *SymAtlas* (Smets *et al*, 2006). 3) Location within a random DNA region, either intergenic or non-intergenic region.

5.3 Identification of differentially methylated regions on Chromosome 21 by methylation-sensitive restriction analysis

MSRE analysis using *Hpa*II restriction digestion, with term placenta and peripheral blood leukocyte genomic DNA, was performed to screen for differentially methylated regions on Chromosome 21. The restriction endonuclease *Hpa*II cuts DNA at unmethylated CCGG target sites, whereas if the site comprises a methylated C residue at the included CpG sequence, the target site remains uncut. DNA regions including one or more methylated (and therefore uncut) target sites are resistant to *Hpa*II and hence can be amplified by polymerase chain reaction (PCR) primer pairs spanning the methylated sites, leading to a DNA band upon agarose gel electrophoresis (figure 1.22; figure 5.1) (see page 72 for primers).

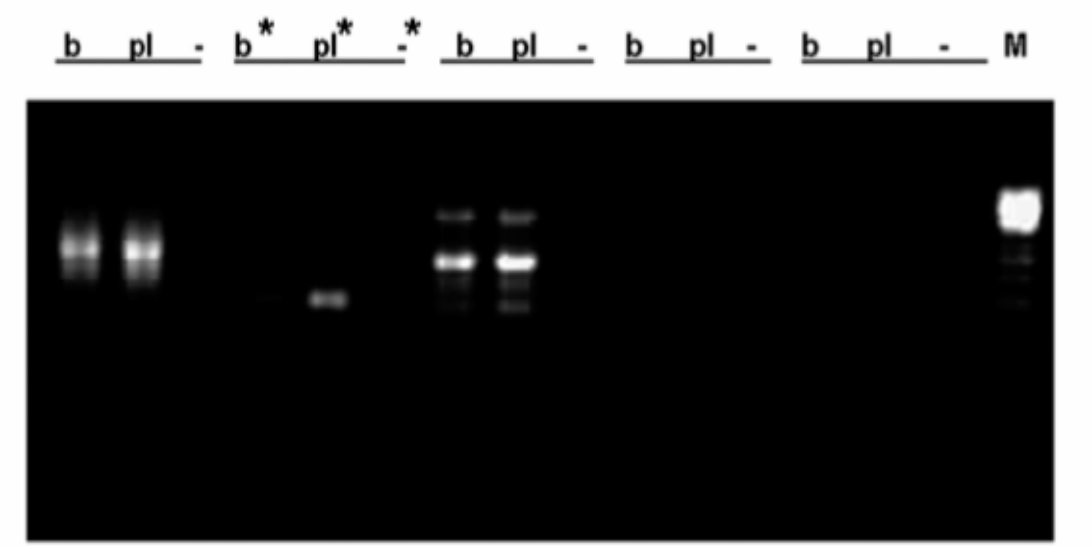


Figure 5.1: Gel electrophoresis image of methylation-sensitive restriction endonuclease analysis on Chromosome 21
*Hpa*II-digested placenta genomic DNA is marked (pl) and *Hpa*II-digested peripheral blood leukocyte DNA is marked (b) non-template PCR controls (–) are indicated. Lanes marked with an asterisk indicate PCR products for the 21–58 primer pair, indicating the absence of a strong band and hence hypomethylation in digested blood genomic DNA, compared with hypermethylation in placenta genomic DNA. Other blood and placenta DNA pairs of lanes indicate non-differential methylation for DNA regions defined by other primers in the screen of Chromosome 21 regions. The presence of bands indicates methylation of *Hpa*II sites in both genomic DNA samples; absence of bands indicates absence of methylation in both genomic DNA samples. M = DNA marker lane. (Taken from Old *et al,* 2007).

5.4 Identification of differentially methylated regions on Chromosome 22 by methylation-sensitive restriction analysis

MSRE analysis of peripheral blood leukocyte and term placenta DNA has also been used to assess sequences on Chromosome 22, for use as control markers on non-trisomic chromosomes. An additional candidate marker was elucidated by following previous work published on the epigenetic profile of Chromosome 22 (Eckhardt *et al*, 2006). Among the regions assessed by MSRE analysis, a candidate epigenetic marker was identified on the *Oncostatin M* (*OSM*) gene on Chromosome 22, which encodes a growth regulator which inhibits the proliferation in a number of tumour cell lines. The region contains 1 *Hpa* II restriction site and 1 *Hha* I restriction site and both contain CpGs with methylated cytosines in placental DNA and hypomethylated in peripheral blood DNA. One region was detected as being differentially methylated in the Chromosome 22 screen (figure 5.2) (see page 71 for primers).

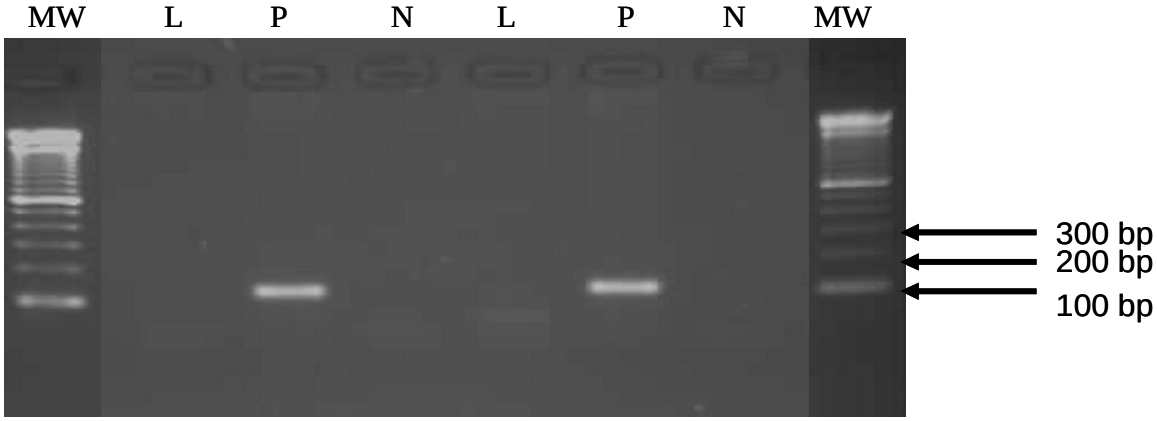


Figure 5.2: Gel electrophoresis image of methylation-sensitive restriction endonuclease analysis of Chromosome 22
Gel electrophoresis image showing PCR products from peripheral blood leukocyte DNA (L) and placenta DNA (P) with non-template controls (N). Each set of DNA products has been digested with either *Hpa* II (first three lanes) or *Hha*I (*) before PCR amplification with primers for the candidate epigenetic biomarker 22-09. These data indicate that placenta DNA is methylated whereas peripheral blood leukocyte (by the presence of a PCR product) and that peripheral blood leukocyte DNA is unmethylated (by the absence of a product band). The 100 bp DNA molecular weight ladder is marked (M).

5.5 Identification of differentially methylated regions on chromosomes 1, 2 and 12 by gene expression profile analysis and methylation-sensitive restriction analysis

Regions were also selected on the basis of their expression profiles, as determined by analysis of expression profiling microarray data, which is publicly available on the SymAtlas website (http://symatlas.gnf.org/SymAtlas/). Some CEBs were detected on chromosomes 1, 2 and 12 (figure 5.3) (see pages 70-72 for primers).

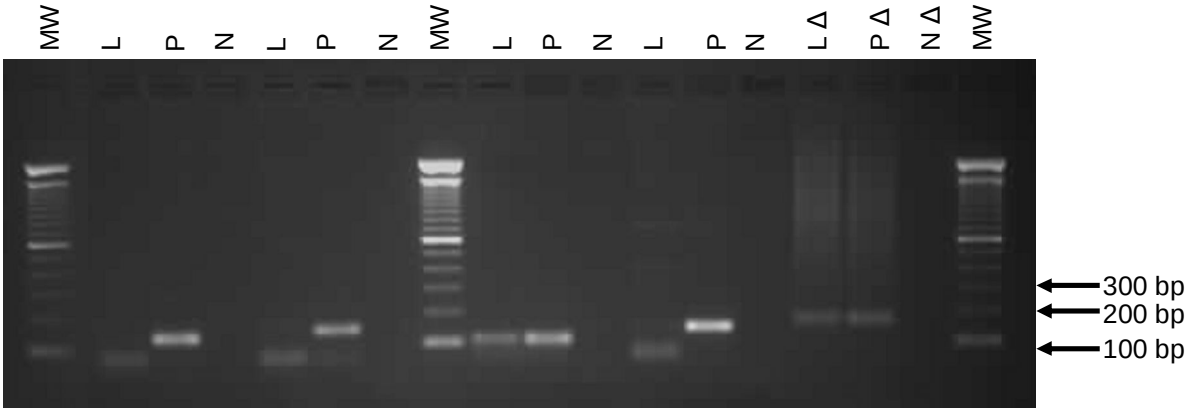


Figure 5.3: Gel electrophoresis image of methylation-sensitive restriction endonuclease analysis of targets identified by expression profiling
Each DNA region selected is represented by three lanes L, P, N. All of the markers from the expression profiling screen shown here are differentially methylated; these lanes show (from left to right) primer pairs 2-28, 1-16, 21-128, 1-50 and 12-42. The DNA region amplified by the last pair of primers 12-42 (denoted by Δ) are not differentially methylated as is shown by the amplification of both peripheral blood leukocyte and placenta DNAs. The third set of primers (21-128) are differentially methylated, however there is a background of methylated DNA in the peripheral blood. MW denotes the 100 bp DNA molecular weight ladder (Invitrogen, UK).

Table 5.1: Panel of candidate epigenetic biomarkers discovered by methylation-sensitive restriction analysis

All of the candidate epigenetic biomarkers detected by methylation-sensitive restriction analysis are shown; with additional information, including chromosome location, laboratory nomenclature, associated Genbank name of the closest gene and methylation status in both placenta DNA and peripheral blood leukocyte DNA.

5.6 Panel of candidate epigenetic biomarkers

Chromosome location	Laboratory designation of marker	Associated Gene (Genbank symbol)	Methylation Status	
			Placenta	Leukocyte
21q22.3	21-58	AIRE	+	-
21q22.3	21-107	CLDN14	+	-
21q22.3	21-128	ERG	+	-
21q22.3	21-210	C21orf1	+	-
1q23.1-q24.1	1-16	CD48	+	-
1q32.1	1-50	FAIM3	+	-
2p14	2-28	ARHGAP25	+	-
12q24	12-42	SELPLG	+	-
22q12	22-09	OSM	+	-

5.7 Discussion

The model system first proposed by Chim and colleagues can be used to successfully identify potential epigenetic biomarkers of cff DNA (Chim *et al*, 2005; Chan *et al*, 2006). A recent development in the field has been the identification of two putative epigenetic biomarkers on Chromosome 21 (Chim *et al*, 2008). By use of the model system it has been possible to identify a panel of candidate epigenetic biomarkers (CEBs) on Chromosome 21 and other autosomes.

Primers were designed for regions of between 300-400 bp on the promoter regions of genes (promoter regions were characterised as the first 2.25 kb upstream of the putative translation initiation site). With DNA regions selected being CpG rich but not within CpG islands. Each region selected contained at least one or more restriction site for the methylation-sensitive restriction enzymes *Hpa* II and *Hha* I.

DNA regions selected were either chosen at random (intergenic spaces or random promoters) or by use of gene expression profiling data from the SymAtlas website (Smets *et al*, 2006) comparing over expression in whole blood and under expression in placenta and *vice versa*. It seems that whichever selection method is used to select regions for analysis, differentially methylated regions can be detected readily, provided that enough regions are tested.

In total 366 regions were assessed on all chromosomes including 152 on Chromosome 21, with only a few regions detected as differentially methylated (approximately 3 %). On the Chromosome 21 analysis the first differentially methylated region detected was the 58th primer pair assessed; this CEB was therefore termed 21-58 (figure 5.1). The DNA region which the 21-58 primers amplify is located on the *AIRE* gene promoter is methylated in placenta and unmethylated in placenta DNA (figure 5.1).

An additional CEB was also detected on Chromosome 22 this CEB was termed 22-09 as it was discovered with the 9th primer pair tested (figure 5.2). CEBs were also

identified from analysis of targets from gene expression profile data, CEBs were detected on chromosomes 1, 2 and 12 (figure 5.3) and on Chromosome 18 (data not shown).

These data show that MSRE analysis can be used to identify CEBs. These data represent the first CEBs discovered on Chromosome 21 (Old *et al*, 2007). Using the model system it is possible to readily detect CEBs which may potentially have a use in non-invasive prenatal diagnosis (table 5.1). Further CEBs are required, as the number of cff DNA molecules in plasma of pregnant women during the first trimester is low, and increasing the number of markers can be taken to have the same effect as taking a larger sample (Old and Stallard, accepted for publication). Also by increasing the number of CEBs the coverage of chromosomes increases and enables the distinction between fetal trisomy and mosaicism or chromosomal translocation. The position of CEBs on each chromosome is important, as in some cases chromosome translocations may occur which duplicate only a part of the chromosome; therefore if for instance CEBs are located on the p arm of a chromosome but a translocation occurs on the q arm of a chromosome, this would not be detected. By having CEBs at a number of loci on any given chromosome, increases the chances of detecting a chromosome translocation.

Results

Chapter 6

Candidate epigenetic biomarker discovery by methylated DNA immunoprecipitation of genomic DNA from chromosomes 21 and 18

6.1 Aims

The data sets presented in this chapter describe how epigenetic biomarkers can be discovered by methylated DNA immunoprecipitation (MeDIP) of DNA. The data sets presented in this chapter were obtained by collaboration with Dr Nigel Carter's group at the Sanger Institute. The data presented in this chapter refer to the discovery and further characterisation of two candidate epigenetic biomarkers, one on Chromosome 21 and one on Chromosome 18 using the MeDIP with whole blood and placenta DNA.

6.2 Introduction

MeDIP is another method by which the methylation of DNA can be assessed. Using the antibody Anti-5-Methylcytosine can provide up to a 90 fold enrichment of methylated DNA in a dose dependent sequence dependent manner (Weber *et* al, 2005; Wilson *et al*, 2006). The methylation profiling of chromosomes 6, 20 and 22 has been partially completed; by bisulfite sequencing of DNA regions (Eckhardt *et al*, 2006). Methylation profiling of chromosomes 18 and 21 using MeDIP has been undertaken at the Sanger institute (Cambridge, UK). Using whole blood and placenta DNA samples were divided equally with one half of the sample precipitated with Anti-5-Methylcytosine and the other not. The samples were then labelled with fluorescent dyes cy3 and cy5 are co-hybridised onto a tiling microarray for either Chromosome 18 or Chromosome 21. Immunoprecipitated DNA was compared with non-immunoprecipitated DNA by hybridisation to a tiling array. Each tiling array consists of 60-70 mer oligonucleotides, distributed along the chromosome with gaps in between them as indicated (figure 6.1; figure 6.3). In order to detect methylated regions for each tissue on either Chromosome 18 or Chromosome 21; the array data for each tissue are then compared to give differentially methylated regions between whole blood DNA and placenta DNA (figure 6.1; figure 6.3). Candidate epigenetic biomarkers were characterised further by MSRE analysis.

Using MeDIP combined with microarray technology enables methylation profiling of entire chromosomes, which greatly improves the efficiency of CEB discovery.

Figure 6.1: Methylation microarray data for whole blood and placenta DNA on Chromosome 18

Shows tiling microarray data for a region on Chromosome 18 ranging from nucleotides 58950400- 58960800 (NCBI build 36.2). Each bar indicates an oligonucleotide on the chromosome and gives a signal after hybridisation of either under-representation or over-representation on a log 2 scale. The difference between immunoprecipitated placenta DNA and untreated placenta DNA is shown in green; the difference between immunoprecipitated whole blood DNA and untreated whole blood DNA is shown in yellow. The subtractive difference between placenta DNA and whole blood DNA is shown in black. The area circled in red and showing two red signals is a region which is hypermethylated in placenta DNA and hypomethylated in whole blood DNA. This signal indicates the region of a candidate epigenetic biomarker on Chromosome 18.

6.3 Differentially methylated regions on Chromosome 18

Position (NCBI36) of differentially methylated region: **58956134-58956604**

Further characterisation of the CEB from the tiling array data by MSRE analysis confirms that the region is differentially methylated between peripheral leukocyte and placenta DNA (figure 6.2).

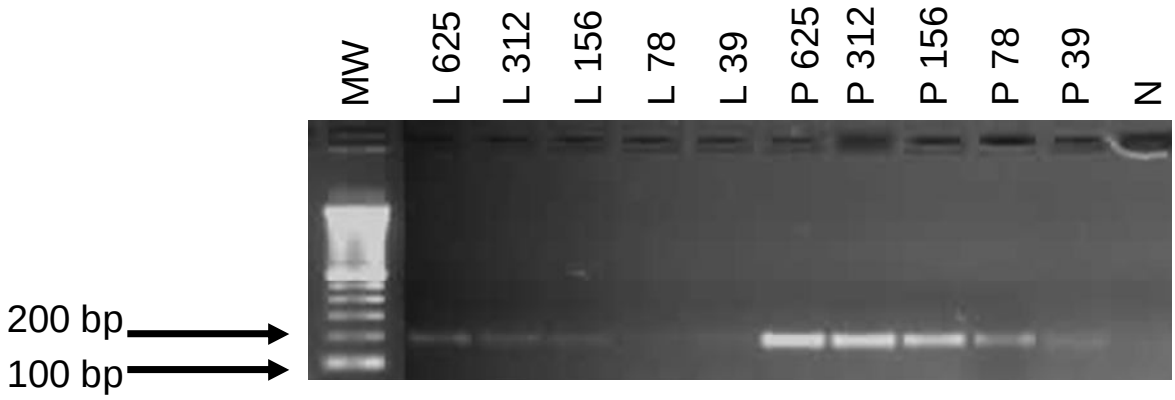
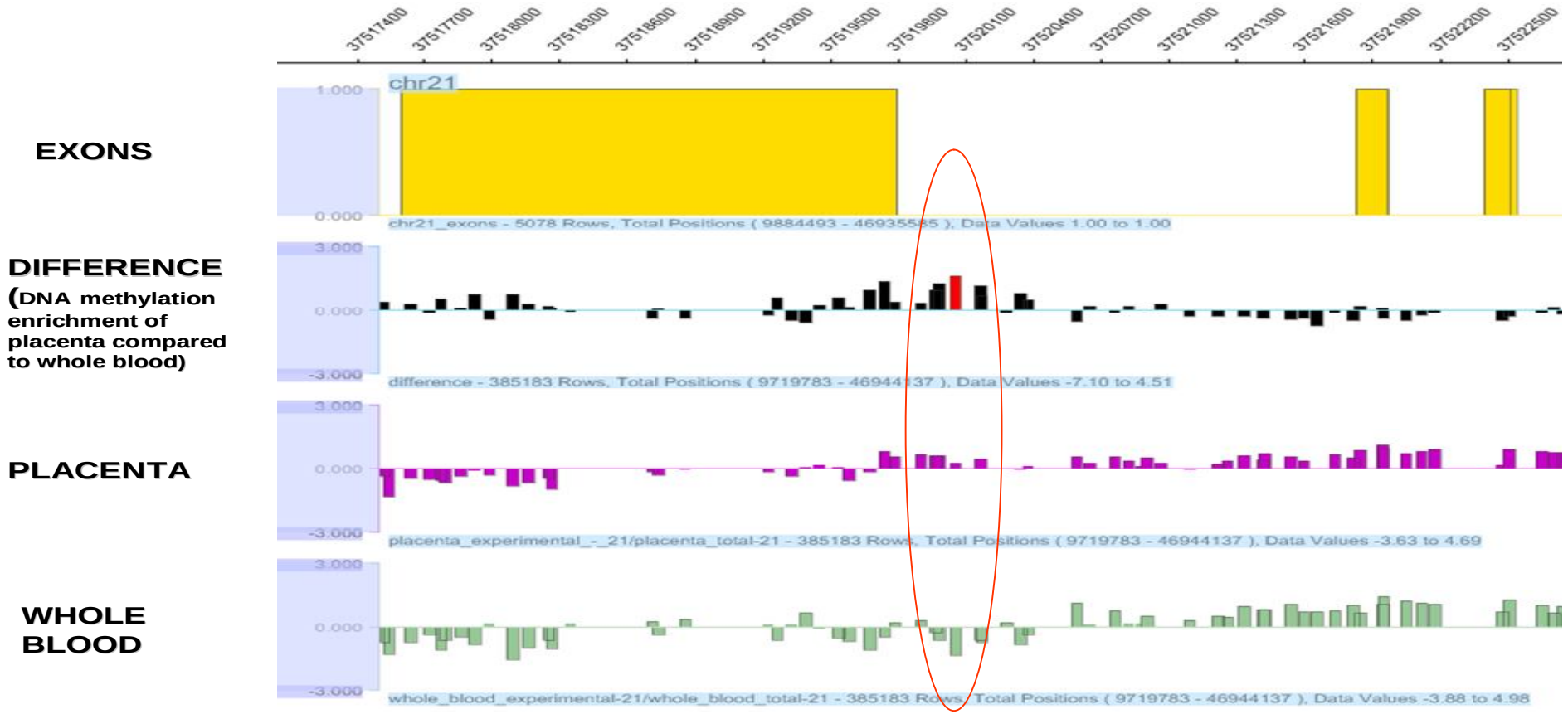


Figure 6.2: Methylation restriction assay data for the LM18-58A primers from the Chromosome 18 tiling array
Primers were designed for the region highlighted by the microarray data, including 200 bp upstream and downstream of the region (for the reason that the DNA was sonicated and fragments between 70-150 bp may have bound to the oligonucleotides which are 60-70 bp in length). Genomic leukocyte (L) and genomic placenta (P) were diluted in serially diluted; the figures given show [DNA] in picograms and double digested with *Hpa* II and *Hha* I. The gel data shows that leukocyte DNA (L625-39) is almost completely digested whereas placenta DNA (P625-39) is undigested. This confirms the differential methylation suggested by the MeDIP array. The lane labelled N is a non -template control and the lane labelled MW contains a 100 bp molecular weight DNA ladder (Invitrogen, UK).

Figure 6.3: Methylation microarray data for whole blood and placenta DNA on Chromosome 21

Shows tiling microarray data for a region on Chromosome 21 ranging from nucleotides 37517400- 37522500 (NCBI build 36.2). Each bar indicates an oligonucleotide on the chromosome and gives a signal after hybridisation of either under-representation or over-representation on a log 2 scale. The difference between immunoprecipitated placenta DNA and untreated placenta DNA is shown in purple; the difference between immunoprecipitated whole blood DNA and untreated whole blood DNA is shown in green. The subtractive difference between placenta DNA and whole blood DNA is shown in black. The area circled in red and showing a red signal is a region which is hypermethylated in placenta DNA and hypomethylated in whole blood DNA. This signal indicates the region of a candidate epigenetic biomarker on Chromosome 21.

6.4 Differentially methylated regions on Chromosome 21



Position (NCBI36) of differentially methylated region: **37519941-37520196**

Further characterisation of the region detected on the Chromosome 21 tiling array by MSRE analysis identifies a region containing two CpG sites which are methylated in placenta DNA and unmethylated in peripheral blood leukocyte DNA (figure 6.4).

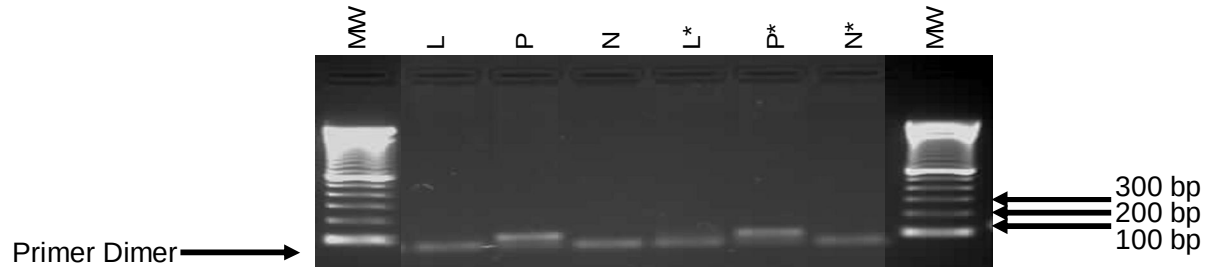


Figure 6.4: Methylation restriction assay data for the 21-37 primers from the Chromosome 21 tiling array
Primers were designed for the region highlighted by the microarray data, including 200 bp upstream and downstream of the region (for the reason that the DNA was sonicated and fragments between 70-150 bp may have bound to the oligonucleotides which are 60-70 bp in length). Genomic leukocyte DNA digested by *Hpa* II is denoted (L) and digested by *Hha* I is denoted (L*) and genomic placenta DNA digested by *Hpa* II is denoted (P) and when digested by *Hha* I is denoted (P*). The gel data shows that leukocyte DNA is completely digested *Hpa* II and almost completely digested by *Hha* I; whereas placenta DNA is undigested. This confirms the differential methylation suggested by the MeDIP array. The lanes labelled (N and N*) are non -template controls and the lanes labelled MW contain a 100 bp molecular weight DNA ladder (Invitrogen, UK).

6.5 Discussion

MeDIP can be used to enrich methylated DNA prior to application to whole chromosome tiling microarray. Selective data received from the Sanger institute (Cambridge, UK) from Chromosome 18 and Chromosome 21 tiling arrays has provided targets for CEBs on these chromosomes.

Tiling array data on Chromosome 18 indicated a number of CEBs, one of which is located between 58956000 bp and 58956800 bp on Chromosome 18 (NCBI 36.2). Placenta and whole blood DNA samples were collected and divided equally, one half of each sample was immunoprecipitated with anti 5-Methylcytosine, and each sample was then labelled with either cy3 or cy5 fluorescent dyes. The immunoprecipitated samples were compared with the untreated samples to establish

methylation maps of chromosomes 18 and 21 for each tissue, a further comparison between the tissue types was then made showing regions where methylated DNA was over represented or under represented (figure 6.1; figure 6.3). Signals on each tiling array represent oligonucleotides from 60-70 bp along the length of each chromosome, and signal strength is measured within a log 2 scale of -3 fold to +3 fold, with positive signals indicating over-representation of methylated DNA and negative signals indicating under-representation of methylated DNA (figure 6.1; figure 6.3).

One signal between 58956000 and 58956800 on Chromosome 18 shows a region which is over-represented in placenta and under-represented in whole blood; this region is indicated by the red histogram signal, indicating a strong subtractive difference between the enrichment of placenta DNA and that of whole blood DNA (figure 6.1). This region was further characterised by design of primers for shorter amplicons between 70-150 bp for use with MSRE analysis. One primer pair termed LM18-58-A amplifies a region of 148 bp containing a single *Hpa* II site and a single *Hha* I site, which are differentially methylated between placenta DNA and peripheral blood leukocyte DNA (figure 6.2).

Tiling array data for Chromosome 21 shows a strong signal between 37519800 bp and 37520100 bp, with methylated DNA indicated as under-represented in whole blood and methylated DNA indicated as over-represented in placenta (figure 6.3). Further characterisation of the region with a primer pair termed 21-37 which amplifies a region of 92 bp containing a single *Hpa* II site and a single *Hha* I site, which are differentially methylated between placenta DNA and peripheral blood leukocyte DNA (figure 6.4).

These data indicate that MeDIP can be used to detect methylated regions across entire chromosomes, and provide target regions for the discovery of CEBs. These regions require further characterisation by higher resolution techniques such as MSRE analysis to define CEBs more accurately.

Results

Chapter 7

Epigenetic profiling of the *AIRE* gene promoter

7.1 Aims

Two of the first candidate epigenetic biomarkers (CEBs) to be identified have been further characterised with in-depth analysis of their methylation profiles. Bisulfite conversion of DNA followed by cloning and sequencing of PCR products (bisulfite sequencing) has been used to compare the methylation profiles of leukocyte and term placenta DNA (BioChain, USA). In addition, methylation-specific primers (MSPs) have been optimised for one of the regions. Data presented in this chapter focuses on the characterisation of 2 CEBs, both of which are located on Chromosome 21, and both of which are located on the promoter region of the *AIRE* gene. The COBRA analysis of the *AIRE* and *MASPIN* products was prepared by Dr Francesco Crea.

7.2 Introduction

Using the model system of leukocyte and term placenta DNA, two CEBs have been characterised on the promoter of the *AIRE* gene. The *AIRE* gene is an autoimmune regulator and the methylation of the gene has been previously studied (Murumagi *et al*, 2003). The region studied by Murumagi and colleagues refers to a CpG island 300 bp upstream of the transcriptional start site and is not overlapped by either of the regions included in this study. Following up on initial data from the methylation sensitive restriction assay (figure 5.1) in addition to the original marker (termed the 58 region another region was detected on the *AIRE* gene promoter by bisulfite conversion of DNA and Bisulfite specific PCR (BS-PCR). This region has been termed the A1 region of the *AIRE* promoter.

Each region has been characterised by a combination of methylation-specific techniques including: bisulfite sequencing, methylation-specific PCR, melt curve analysis of bisulfite-specific PCR products and combined bisulfite and restriction analysis (COBRA).

7.3 Location of the 58 and A1 regions on the *AIRE* promoter on Chromosome 21.

The 58 region is located between -1859 and -1651 nucleotides relative to the putative translational start site and is hypermethylated in placenta DNA and is hypomethylated in peripheral blood leukocyte DNA, it contains 4 CpG sites within the region studied (figure 7.1). The A1 region is located further upstream between -486 and -305 nucleotides relative to the translational start site, and is methylated in the opposite way to the 58 region, with hypomethylated placenta DNA and hypermethylated peripheral blood leukocyte DNA (figure 7.1).

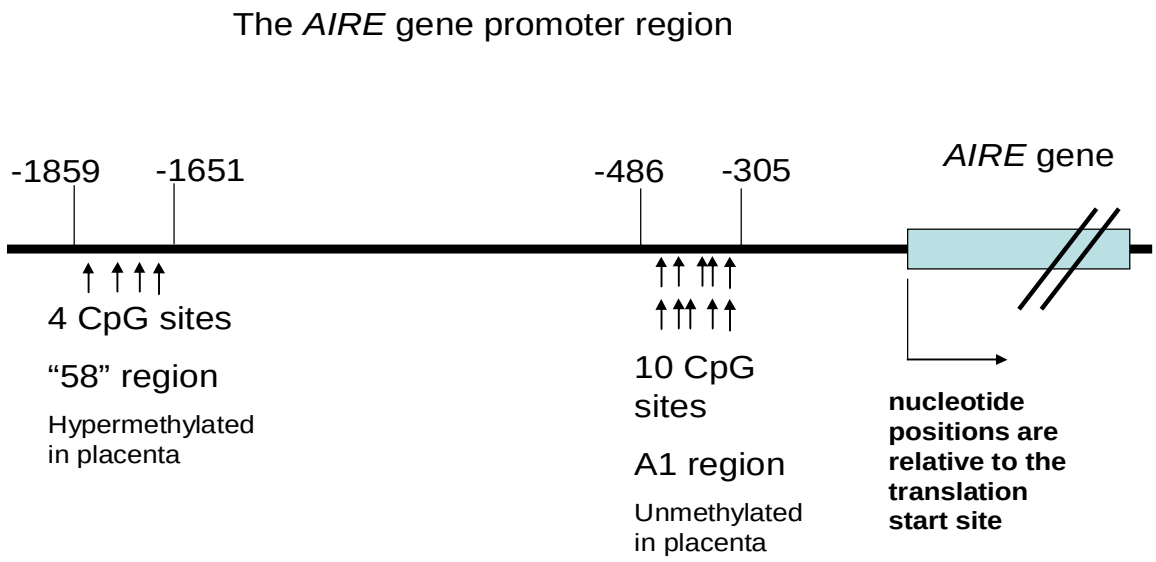


Figure 7.1: Schematic diagram of the *AIRE* gene promoter showing candidate epigenetic biomarkers
The 58 region and A1 region sequences are located within the *AIRE* gene promoter region. The nucleotide numbers that define the regions are indicated, relative to the transcriptional start site of the *AIRE* gene. Vertical arrows indicate the presence of CpG sites.

7.4 Epigenetic profile of the 58 region

Cloning and sequencing of BS-PCR products for the 58 region of the *AIRE* gene promoter was carried out in order to assess the methylation status of DNA in peripheral blood leukocyte and term placenta tissues (table 7.1). The data show that

leukocyte DNA is largely unmethylated, whereas placenta DNA is heavily methylated (table 7.1).

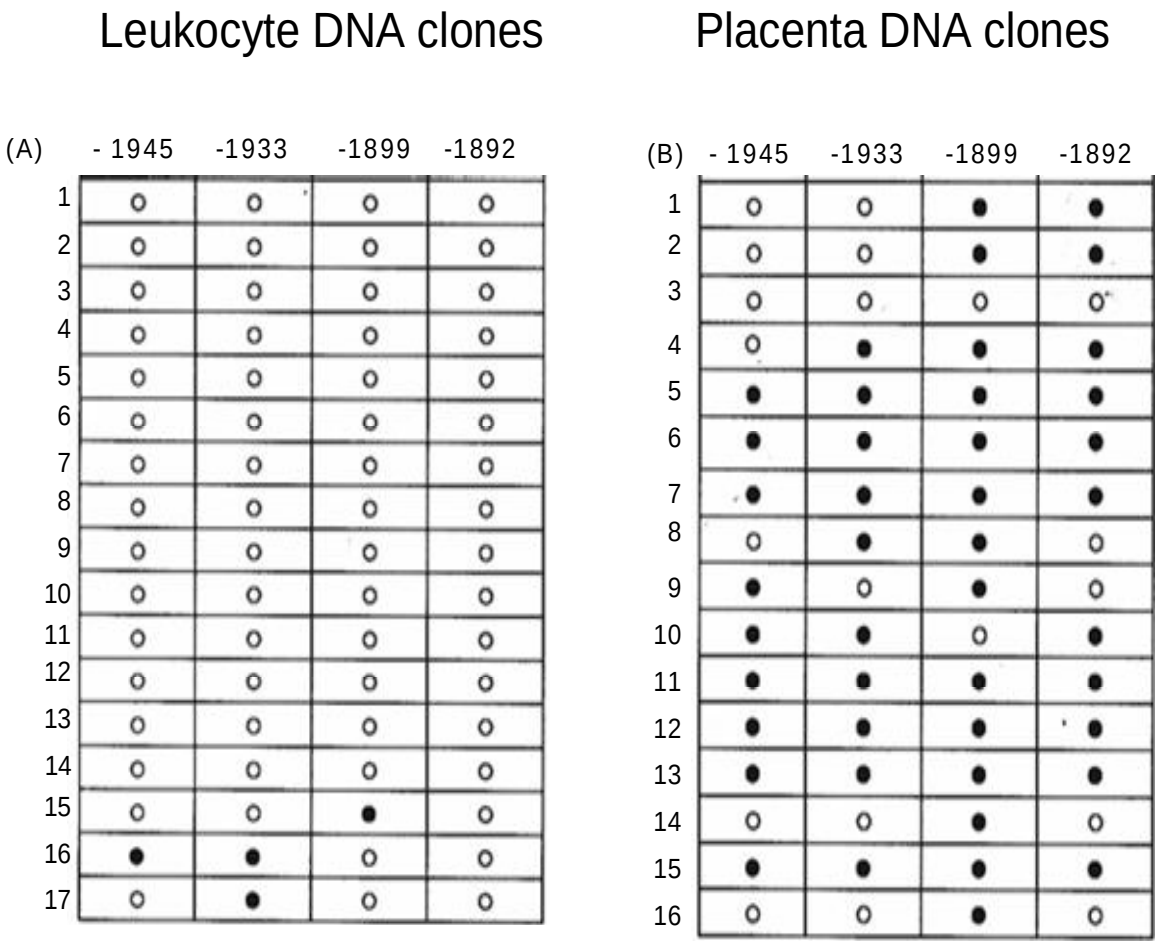


Table 7.1: Methylation profile of the 58 region of the *AIRE* promoter
Shows methylation of the four CpG sites in the 58 region sequence from 5' to 3' ends. Cloning and sequencing of BS-PCR amplicons, showing methylated (filled circles) and unmethylated CpG sites (open circles) at four CpG sites within the 58 region. The relevant cytosine bases are located at nucleotides –1945, –1933, –1899 and –1892 relative to the translational start site of the *AIRE* gene (Murumagi *et al*, 2003). Most clones show substantial methylation in term placenta DNA (B), while nearly all are unmethylated in peripheral blood leukocyte DNA (A).

7.5 Optimisation of methylation specific primers for the 58 region of the *AIRE* gene promoter

The data from the bisulfite sequencing of the 58 region show that the 58 region is a good candidate epigenetic biomarker. On the basis of this data methylation-specific primers (MSPs) were designed using the web-base program MethPrimer (Li and Dahiya, 2002). MSPs contain CpG sites, which after bisulfite conversion of DNA can be used to amplify methylated DNA but not unmethylated DNA, due to the retention of cytosines in CpGs in methylated DNA sequences and the loss of cytosines in CpGs in unmethylated DNA sequences. The first pair of 58 MSPs was able to detect low amounts of methylated placental DNA, in a background of unmethylated peripheral blood leukocyte DNA, however this assay required nested PCR to efficiently detect placental DNA (figure 7.2) (see page 70 for primers).

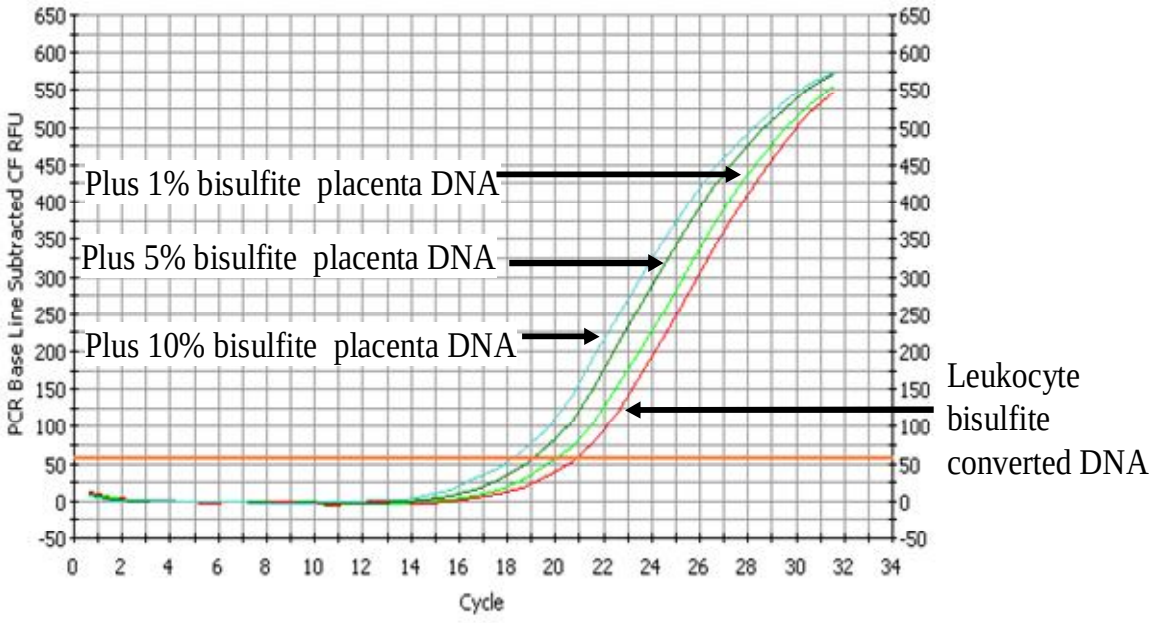


Figure 7.2: PCR quantification data from nested-PCR with placenta DNA spiked in a background of leukocyte DNA
During the initial stages of methylation-specific primer (MSP) optimisation for the 58 region, primers were not highly differential in single-amplification PCR format. However after nested PCR, it was observed that as little as 1% placenta DNA could be detected in a background of leukocyte DNA. Leukocyte bisulfite converted DNA had a Ct value of 20.9; leukocyte bisulfite converted DNA spiked with 1% of placental bisulfite converted DNA had a Ct value of 20.0 and when spiked with 5% placenta bisulfite converted DNA a Ct value of 19.1 and with 10% placenta DNA 18.2 Ct.

The original set of 58 region MSPs demonstrated that relatively low amounts of methylated placental DNA could be detected in a background of unmethylated peripheral blood leukocyte DNA however the primers were inefficient (figure 7.2). New MSPs were designed using the MethPrimer programme (Li and Dahiya, 2002). The new primers were designed to be more efficient and able to preferentially amplify methylated placental DNA in a single PCR reaction (figure 7.3) (see page 72 for primers).

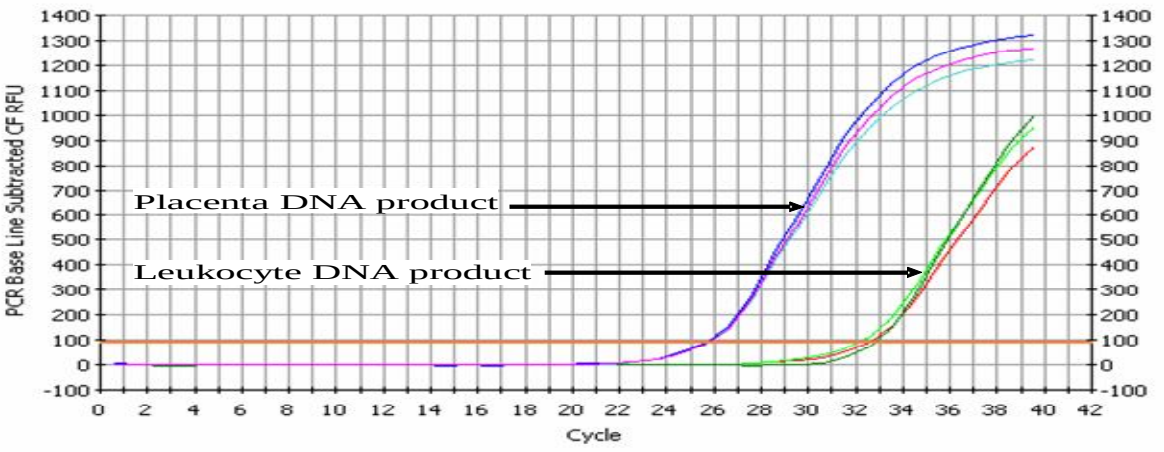


Figure 7.3: Real-time PCR quantification data for the optimised methylation-specific primers for the 58 region
Optimised MSPs for the 58 region of the *AIRE* promoter on Chromosome 21 were used for real-time PCR amplification of bisulphite-converted DNA. The mean cycle threshold (Ct) for bisulphite-converted placenta DNA is 25.7 Cts and for bisulphite-converted blood DNA it is 32.4 Cts. Thus placental DNA is amplified at a much earlier Ct as indicated by the ΔCt of 6.7 cycles. RT-PCR analysis was in triplicate. By extrapolation, this degree of specificity is expected to allow identification of 1% cell-free fetal DNA of placental origin in maternal plasma.

7.6 Characterisation of the A1 region of the *AIRE* gene promoter

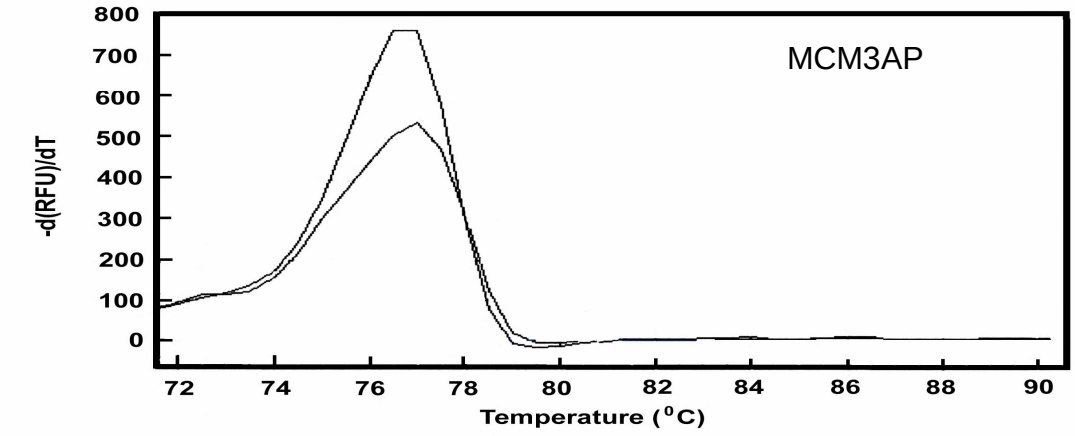
In addition to the 58 region of the *AIRE* gene promoter, further methylation analysis was also carried out on the A1 region of the *AIRE* promoter. After BS-PCR the BIORAD iCycler can be programmed to perform a temperature gradient, by which the PCR products are melted and data are recorded regarding the melting temperature (Tm) of each product. The temperature gradient can be used to detect differences between the melting temperatures of BS-PCR products derived from the amplification of peripheral blood leukocyte and term placenta DNAs. After bisulfite

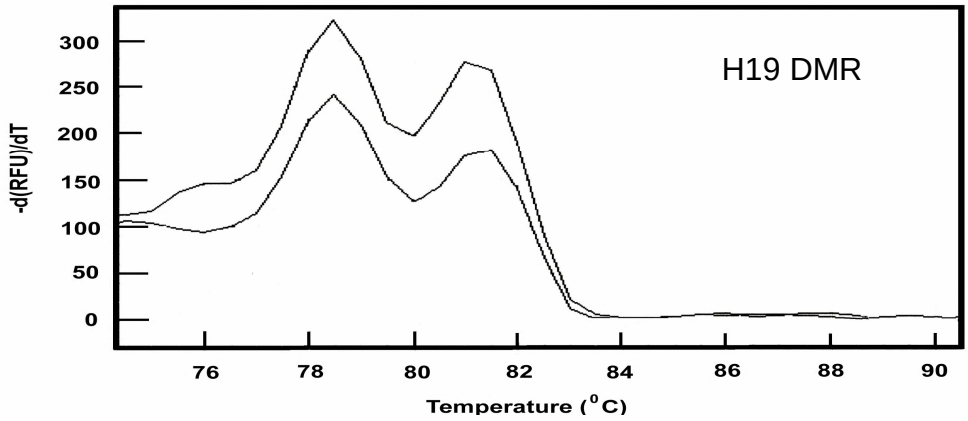
conversion of DNA unmethylated cytosines are converted to uracil, and during PCR and are incorporated into the PCR product as thymine by Taq DNA polymerase; the overall effect is to cause a C → T transition in unmethylated DNA, whereas cytosines within CpG sites in methylated DNA sequences are protected. One advantage of this technique is that CpGs are measured in the context of the whole amplicon; therefore there must be a significant cumulative difference in the methylation of the CpGs, for a different Tm to be detected.

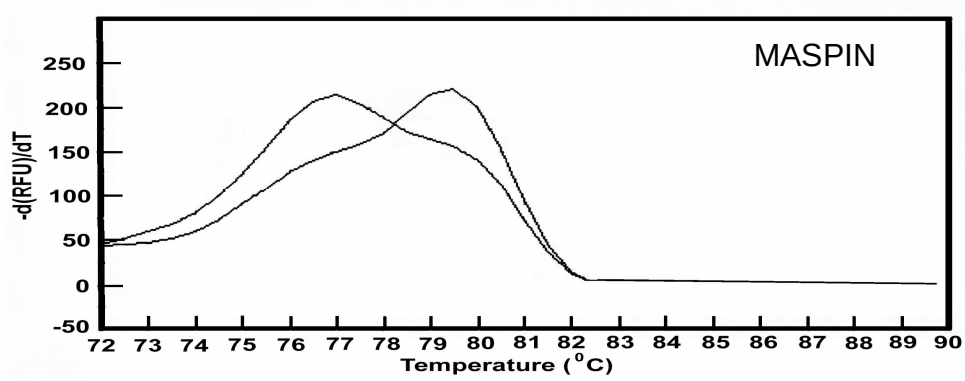
The melting profile of the A1 region was assessed, along with the *MASPIN* putative epigenetic marker, the imprinted locus *H19 DMR* and a non-differentially methylated region *MCM3AP* (figure 7.4). The *MASPIN* promoter, was used as a positive control, and has previously been shown to include a 204 bp region that is largely unmethylated in term placenta DNA, but methylated in leukocyte DNA. As expected, the melting profile of the leukocyte amplicons peak at about 2.5°C higher than placenta DNA amplicons (figure 7.4 C). The *H19* locus on human chromosome 11 was also analysed. The melting profiles of this region are bimodal curves, with the difference in melting profile peaks being relatively large, about 3.0°C (figure 7.4 B). The bimodal curves are similar for both the leukocyte and placenta DNAs (figure 7.4 B). The *MCM3AP* promoter region displays melting profiles with single peaks from leukocyte and placenta DNA, with no clear shift in the melting profiles (figure 7.4 A). For the A1 region, leukocyte DNA is hypermethylated in comparison with placenta DNA, as indicated by the shift of 1.5°C in the melting profile peaks of the 182 bp amplicons (figure 7.4 D). The A1 region of the *AIRE* gene promoter region was therefore identified as a potential epigenetic biomarker for NIPD of trisomy 21.

The differences in methylation status of the A1 region of the *AIRE* gene promoter were analysed by cloning and sequencing the 182 bp amplicons from bisulfite-converted leukocyte and placenta DNA (table 7.2). Of particular interest is the presence of clones (2/29) from placenta DNA that are completely unmethylated at all 10 sites of the *AIRE* promoter region, with no example (0/29) of a completely unmethylated clone in leukocyte DNA. Considering sites 1, 3, 5, and 7 together, 17/29 clones from placenta are unmethylated at all of these four sites, whereas there is no clone (0/29) from leukocyte DNA that is unmethylated at all four sites (table 7.2 A; 7.2 B).

Figure 7.4: Melt curve analysis of BS-PCR products
Shows the melting profiles of BS-PCR products from four primer pairs: *MCM3AP*; *H19 DMR*; *MASPIN* and the 21-A1 region. Melting profiles for a 122 bp upstream region of the *MCM3AP* gene located at Chromosome 21q22.3. The upper curve represents leukocyte DNA, the lower placenta DNA. Melting profiles for a 225 bp region of the imprinted *DMR* of the *H19* locus positioned at chromosome 11p15.5. The upper curve represents leukocyte DNA, the lower placenta DNA. Melting profiles for a 204 bp region of the *MASPIN* promoter located at Chromosome 18q21.3. Note that the curve for leukocyte DNA displays a peak shifted to a temperature about 2.5 °C higher than that for placenta DNA. Melting profiles for the 181 bp A1 region upstream of the *AIRE* gene at Chromosome 21q22.3. Note that the curve for leukocyte DNA displays a peak shifted to a temperature about 1.5 °C higher than that for placenta DNA.


MCM3AP
-d(RFU)/dT
800
700
600
500
400
300
200
100
0
72
74
76
78
80
82
84
86
88
90
Temperature (°C)


H19 DMR
-d(RFU)/dT
300
250
200
150
100
50
0
76
78
80
82
84
86
88
90
Temperature (°C)


MASPIN
-d(RFU)/dT
250
200
150
100
50
0
-50
72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
Temperature (°C)

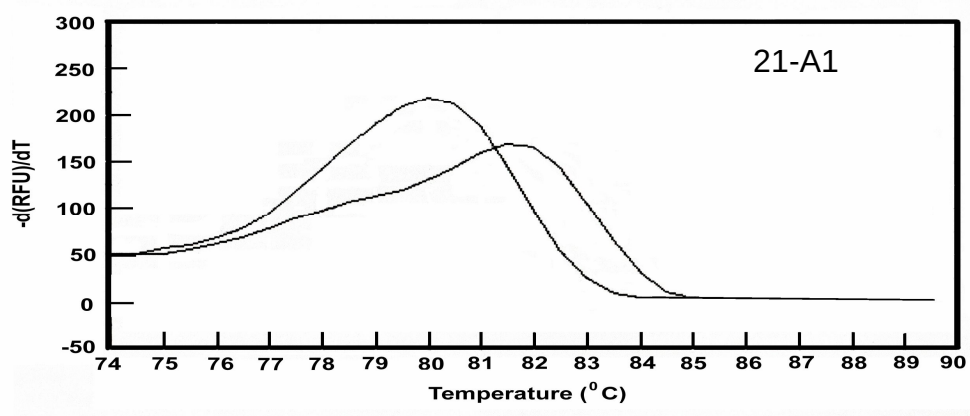

21-A1
-d(RFU)/dT
300
250
200
150
100
50
0
-50
74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90
Temperature (°C)

Table 7.2: Methylation profile of bisulfite-converted DNAs from the A1 region of the *AIRE* promoter

Shows the specific methylation profile of the A1 region in leukocyte and placenta tissues. The position of each CpG for each of the 10 CpG sites is given relative to the translational start site. The rows depict the methylation status of each clone. Filled and unfilled circles represent methylated and unmethylated CpG sites, respectively. (A) Shows the methylation profile for the A1 region in leukocyte DNA, which is hypermethylated, with ~77% of CpGs methylated. (B) Shows the methylation profile of the A1 region in the placenta DNA, which is hypomethylated with ~74% of CpGs unmethylated. The placenta DNA shows a general reduction in methylation patterns in comparison with leukocyte DNA. The four CpG sites: - 461, - 448, - 418 and - 402 form 4 *Taq* I restriction sites when methylated. All of the leukocyte sequences (29/29) have at least one methylated CpG at 1 of the 4 sites. Whereas (17/29) sequences in placenta DNA are unmethylated at all 4 of the CpG sites.

	- 461	- 451	- 448	- 433	- 418	- 405	- 402	- 380	-356	- 347
01	●	●	●	●	●	●	●	●	●	●
02	●	●	●	●	●	●	o	o	o	o
03	●	●	●	●	●	●	o	●	●	●
04	●	●	●	●	o	●	o	●	●	●
05	o	●	●	●	o	o	o	o	o	o
06	●	●	●	o	●	●	●	●	●	●
07	●	●	●	●	●	●	●	●	●	●
08	●	●	●	●	●	●	●	●	●	●
09	●	o	●	●	●	o	●	●	●	●
10	●	●	●	●	o	●	●	●	●	●
11	o	●	o	●	o	●	●	●	●	o
12	o	●	●	●	●	●	o	●	●	●
13	●	●	●	●	●	o	●	●	●	●
14	●	●	●	●	●	o	●	●	●	●
15	●	●	o	●	●	o	o	o	o	o
16	●	●	●	●	●	●	o	●	●	●
17	●	●	●	●	●	o	o	●	●	●
18	●	●	●	●	●	o	●	●	●	●
19	●	●	o	o	o	●	o	●	o	o
20	●	●	●	●	●	●	●	●	●	●
21	●	●	o	o	●	●	●	●	●	●
22	●	●	●	●	o	o	o	●	o	o
23	●	●	●	●	●	o	●	●	●	●
24	o	o	o	●	o	●	●	●	●	●
25	●	●	●	●	o	o	o	●	o	●
26	●	●	●	●	●	o	●	●	●	●
27	o	●	o	●	●	o	●	●	●	o
28	●	●	●	●	●	●	●	●	●	●
29	●	●	●	●	●	o	o	●	o	o

Table 7.2 A: Methylation profile of bisulfite converted leukocyte DNA for the A1 region of the *AIRE* promoter

	- 461	- 451	- 448	- 433	- 418	- 405	- 402	- 380	- 356	- 347
01	o	●	●	o	o	o	o	o	o	o
02	●	o	o	o	o	o	o	o	o	o
03	●	●	●	●	●	●	●	●	●	●
04	o	●	o	o	o	o	o	o	o	o
05	●	●	●	●	●	●	●	●	●	●
06	o	o	o	o	o	●	o	o	o	o
07	o	●	o	o	o	o	o	o	o	o
08	o	●	o	o	●	o	o	o	o	o
09	●	●	o	o	o	o	o	o	o	o
10	o	o	o	o	o	o	o	o	o	o
11	●	●	●	●	o	o	o	o	o	o
12	o	●	o	o	o	o	o	o	o	o
13	o	●	●	o	o	o	o	o	o	o
14	o	●	o	●	o	o	o	o	o	o
15	o	●	o	o	o	o	o	o	o	o
16	o	o	o	●	o	o	o	o	o	o
17	o	●	o	o	o	o	o	o	o	o
18	o	●	●	o	o	o	o	o	o	o
19	●	●	●	●	o	●	●	●	●	●
20	o	●	o	●	o	o	o	o	o	o
21	o	●	o	●	o	o	o	o	o	o
22	o	●	o	o	o	o	o	o	o	o
23	o	●	o	o	o	o	o	o	o	o
24	o	●	o	o	o	●	o	o	o	o
25	o	●	●	o	●	●	●	●	●	●
26	o	●	o	●	o	●	o	o	o	o
27	o	●	o	o	o	o	o	o	o	o
28	o	●	●	●	o	●	o	●	o	o
29	o	o	o	o	o	o	o	o	o	o

Table 7.2 B: Methylation profile of bisulfite converted placenta DNA for the A1 region of the *AIRE* promoter

Combined bisulfite conversion and restriction assay (COBRA) was used to examine the methylation status of the A1 region. When methylated, the four CpG sites at positions 1, 3, 5 and 7 of the A1 region make up four *Taq I* restriction sites in bisulfite converted DNA. In unmethylated DNA these CpG sites do not give rise to *Taq I* restriction sites. To carry out a high resolution and quantitative COBRA of the A1 region PCR was performed on bisulfite-converted DNA, the products were digested with *Taq I*, and analysed by capillary electrophoresis, on a BioRad Experion microfluidic electrophoresis apparatus. The *MASPIN* gene was used as a control, and was highly discriminatory between placenta and leukocyte DNA (figure 7.6 A). The A1 region was also highly discriminatory, with an undigested product present in the placental DNA product and no undigested product in the leukocyte DNA product (figure 7.6 B). These data show that the A1 region is a good CEB for the detection of fetal trisomy 21.

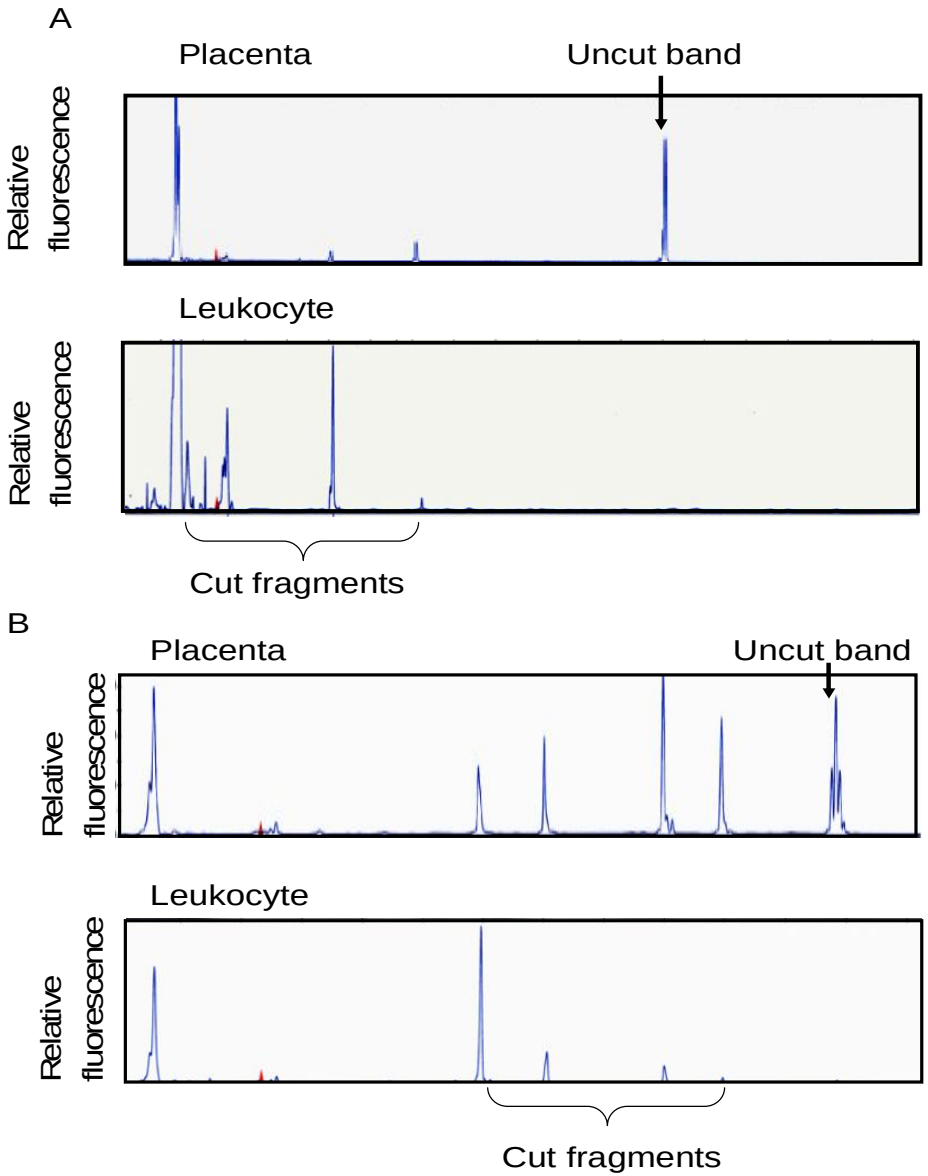


Figure 7.5: High resolution COBRA using capillary electrophoresis

Amplicons of the *MASPIN* and A1 region were derived from placenta and leukocyte bisulfite converted DNA. This was followed by end-labelling using a 5'-FAM labelled reverse primer. The products were digested with *Taq I* restriction endonuclease, and analysed by capillary electrophoresis using fluorescence detection of end-labelled fragments. (A) The fluorescence trace of the *MASPIN* amplicons demonstrates very clearly the difference between the placenta and leukocyte DNA with a large uncut peak for the placenta but no uncut signal for the leukocyte amplicons. (B) The A1 amplicons show a similar differential pattern with a large uncut peak for placenta and a very small signal for the leukocyte amplicons; although there are methylated fragments detected this is expected as the placenta is bathed in maternal blood, this data is concordant with that of the sequencing. The large peaks of low molecular weight product at the far left of each run are caused by primer dimerisation.

7.7 Discussion

Two of the earliest CEBs identified on Chromosome 21 have been extensively characterised by MSRE analysis and bisulfite sequencing and melt curve analysis. The two CEBs are both located on the *AIRE* gene promoter and have opposing methylation profiles (figure 7.1).

Bisulfite sequencing of peripheral blood leukocyte and term placenta DNA shows that the 58 region is hypomethylated on leukocyte DNA and hypermethylated in placenta DNA. With 5.9 % of CpGs methylated in leukocyte DNA and 68.8 % of CpGs methylated in placenta DNA (table 7.1).

Methylation-specific primers were optimised for the 58 region of the *AIRE* promoter. The initial primers required nested PCR to detect differences in methylation of placenta and leukocyte DNA, although 1% of placenta DNA could be detected in a background of 99% leukocyte DNA, with increases of placenta DNA detectable in a dosage dependent manner (figure 7.2). Optimisation of methylation-specific primers for the 58 region was successful with the primers amplifying placenta DNA 6.7 cycle thresholds before leukocyte DNA, meaning that the 58 MSPs are able to detect 1.03 % of placental DNA in a background of leukocyte DNA (figure 7.3).

These data indicate that the 58 region is a strong CEB and could be used in non-invasive diagnostic test, for the detection of cff DNA in a large background of maternal DNA. As in all tests there is a consistent difference between the methylated placenta DNA and the unmethylated leukocyte DNA, which shows that it is possible to detect the low amount of ~1 % of placenta DNA in a background of leukocyte DNA; this level of detection is required as cff DNA can constitute as little as 1% of DNA in the maternal circulation.

The methylation profile for the two tissues in the A1 region is opposite to that of the 58 region; with peripheral blood leukocyte DNA being methylated and term placenta DNA being unmethylated in the A1 region. When identifying CEBs the methylation

of the leukocyte DNA is not preferable, as it means additional techniques are required to assess the epigenetics of the biomarker.
Instead of direct assessment by MSRE analysis alternative techniques are required, which alter unmethylated DNA sequences (bisulfite conversion) and allow analysis by temperature gradient (melt curve analysis) or enable the digestion of the methylated fraction (COBRA).

Alongside several controls the melt products of bisulfite converted DNA for the A1 region of the *AIRE* promoter were analysed (figure 7.4). The data show that the leukocyte DNA BS-PCR product has a higher melting temperature than the placenta DNA BS-PCR product. The placenta DNA product melts at 80 ^{0}C and the leukocyte DNA product 81.5 ^{0}C. This cumulative difference between the 10 CpG sites of the A1 region sequence indicates that leukocyte DNA is hypermethylated and placenta DNA is hypomethylated.

Further characterisation of the A1 region by bisulfite sequencing shows the methylation profile of the region with the methylation of every CpG in the sequence. The data show that leukocyte DNA is hypermethylated with approximately 77% of all CpGs being methylated and that placenta DNA is hypomethylated with 74% of all CpGs being unmethylated (table 7.2).

Additional testing by combined bisulfite restriction analysis (COBRA) also confirmed the consistent differences in methylated between leukocyte and placenta DNA on the A1 region. After bisulfite conversion the BS-PCR products are digested with *Taq* I the methylated CpGs are digested and the unmethylated cytosines transformed to Thymine, removing the *Taq* I restriction site (figure 7.5). *MASPIN* is used as a control as it is methylated in the same way as the A1 region (figure 7.5 A). The data from the A1 COBRA products shows that there is a large amount of undigested-intact product in the placenta DNA; however the leukocyte DNA product is completely digested (figure 7.5 B). This is because the digestion of bisulfite converted DNA relies on the retention of cytosines in the Taq I sites in the A1 sequence. Coincidently there are four CpG sites which lie within Taq I restriction sites in the A1 region; these CpGs are located at -461, -448, -418 and -402. 0% of

clones being unmethylated at all four sites in leukocyte DNA and 58% of clones are completely unmethylated at all four sites in placenta DNA (table 7.2).

The data from the bisulfite sequencing shows that the four CpG site which make the *Taq* I restriction sites are completely unmethylated in 58% of the sequences assessed, this means that the remaining sequences have methylation in at least one of these sites (table 7.2 B). This finding is supported by COBRA data, showing some fragments are cleaved in the placenta DNA (figure 7.5 B), which means that only a proportion of the placenta DNA can be assessed.

Taken together these data strongly indicate that the A1 region is a good CEB and that with further development, this marker could potentially be used as part of a non-invasive diagnostic test for the detection of fetal trisomy. Further testing with plasma DNA from pregnant women is required to validate both CEBs, to determine if the placenta DNA methylation signature is detectable in the plasma DNA of pregnant women and not present in the plasma DNA of non-pregnant women.

Results

Chapter 8

Assessment of candidate epigenetic biomarkers with plasma DNA from pregnant women

8.1 Aims

Since the discovery of cell free fetal DNA (cff DNA) in the maternal circulation (Lo *et al*, 1997) unique biomarkers have been sought after to provide, a method of identifying cff DNA which is independent of fetal sex. So far several potential epigenetic markers have been reported (Chim *et al*, 2005; Chan *et al*, 2006 and Chim *et al*, 2008). Using the model system of leukocyte versus placenta DNA a novel panel of candidate epigenetic biomarkers (CEBs) has been developed, data in this chapter describes the further characterisation of these CEBs using plasma DNA from pregnant individuals. Having discovered several CEBs on a number of chromosomes (see chapters 5, 6, and 7) further characterisation and validation of the fetal specificity of these potential markers is required. This chapter presents data on the characterisation of CEBs using plasma DNA samples collected from pregnant women and non-pregnant women.

8.2 Introduction

In order to assess each CEB, primers for each region were required to be redesigned in order amplify the smaller DNA fragments present in plasma; using the web-based program Primer3input (Rozen and Skaletsky, 2000). This is due to the fact that cff DNA has a smaller size distribution of DNA molecules compared to that of cfm DNA and genomic placenta DNA. Various studies have assessed the properties of cff DNA in the maternal circulation and all agree that cff DNA has a smaller size distribution of DNA molecules than cfm DNA (Chan *et al*, 2004; Zimmerman *et al*, 2004; Chan *et al*, 2005 and Koide *et al*, 2005). It has been established the majority of cff DNA is shorter than 300 bp in length (Li et *al*, 2004); and suggested that most cff DNA has undergone nucleosomal cleavage and is in fact shorter than 193 bp in length (Chan *et al*, 2004). One study has shown that by using a range of primers to define amplicons of different sizes (63 bp – 524 bp) on the SRY gene; that the amount of cff DNA detected is in relation to the size of the amplicon. It was shown that in comparison to a 63 bp amplicon for the SRY gene that each longer amplicon detected fewer copies of the gene in matched samples. An amplicon of 107 bp only detected 53.1 % of the sequences detected by the 63 bp amplicon; this amount decreased to 0.0% when the largest amplicon of 524 bp was used (Koide *et al*, 2005).

The CEBs were redesigned to give amplicons of between 70-150 bp in order to maximise their potential to amplify cff DNA. Suitable markers from the leukocyte and placenta experiments were then assessed with plasma DNA collected from pregnant women prior to delivery.

8.3 Redesign of candidate epigenetic biomarkers for use on plasma DNA

Primers for each of the CEBs were designed using the web-based programme Primer3input (Rozen and Skaletsky, 2000) to give amplicons of between 70 and 150 bp, these primers were then tested on leukocyte and placenta DNA in order to ensure that they still amplify the correct regions efficiently (figure 8.1). The most auspicious CEBs shown are: 2-28, 1-50, 21-107, 18-CR and 21-58 (figure 8.1). These CEBs are methylated in placenta DNA and unmethylated in peripheral blood leukocyte DNA; other candidates were discarded after further testing.

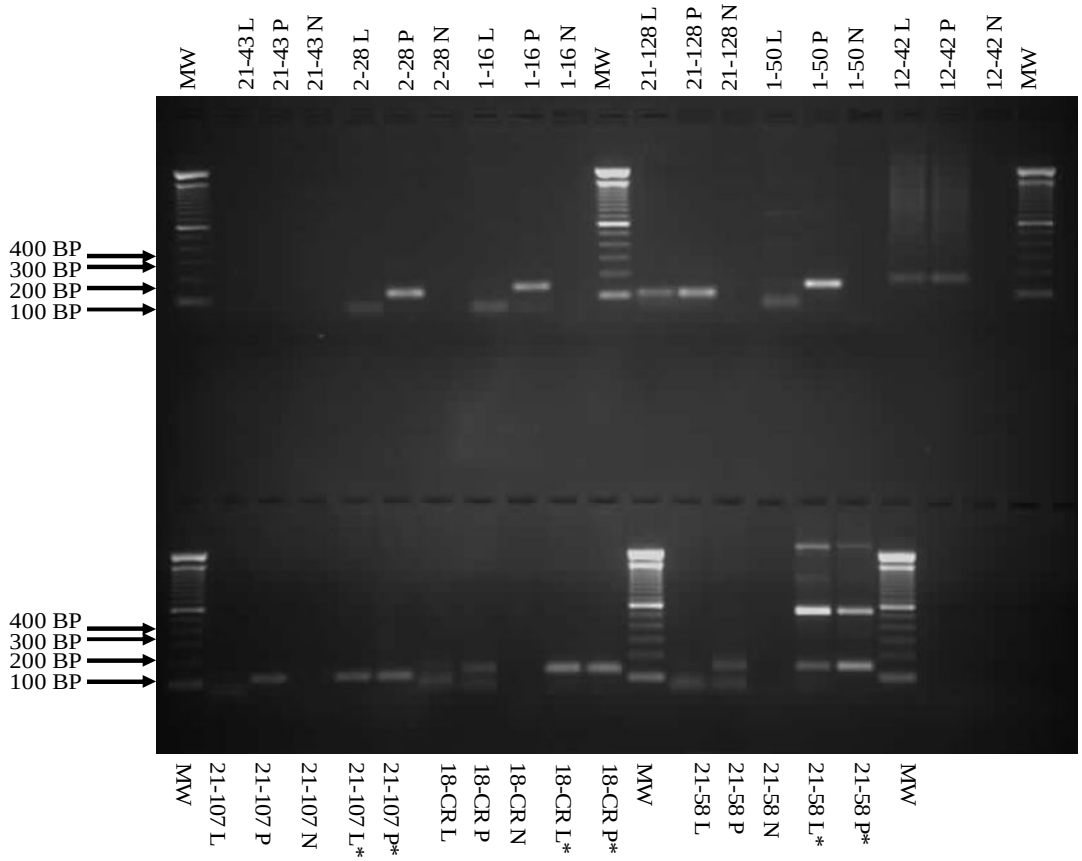


Figure 8.1: Optimisation of short amplicon primers for candidate epigenetic biomarkers

Each set of primers is denoted by the following nomenclature, the first number denotes the chromosome on which the primers are designed and the second numbers and/or letters denote the screen or battery of tests which those primers came from. Each set of PCR products is digested with *Hpa* II or *Hha* I (*). The products of each specific primer pair are labelled (L) for leukocyte DNA, (P) for placenta DNA and (N) for non-template control. Primers 21-43 are either non-differential or have not amplified (lanes 21-43 L and 21-43 P), this may meant that either both leukocyte and placenta DNA are hypomethylated of the primers have failed to amplify in both cases. Primers 12-42 are non-differential, this is shown for the 12-42 primers as both leukocyte and placenta DNA products have amplified a PCR product (lanes 12-42 L and 12-42 P) meaning that they are both hypermethylated. The primer sets 2-28, 1-16, 1-50, 21-107, 18-CR, and 21-58 have amplified regions which are differentially methylated shown in lanes (2-28 L and 2-28 P; 1-16 L and 1-16 P; 1-50 L and1-50 P; 21-107 L and 21-107 P; 18-CR L and18-CR P; 21-58 L and 21-58 P). The three primer sets 21-107, 18-CR and 21-58 also contain *Hha* I sites and were assessed at these sites also shown in lanes marked *. Only 21-58 was partially differentially for a *Hha* I site shown (21-58 L*). MW denotes the 100 bp DNA ladder (Invitrogen, UK).

8.4 Assessment of candidate epigenetic biomarkers on plasma DNA

Further development of CEBs was then carried out using plasma DNA extracted from 3 pregnant women and 3 non-pregnant women. First the total cell-free plasma DNA was assessed using primers for beta globin (figure 8.2). This was followed by the assessment of CEBs 21-107, 22-09 and LM18-58A, which were the most promising candidate epigenetic markers, data forLM18-58A is shown (figure 8.4). Although the CEB 21-128 was not shown to be completely differential in peripheral blood leukocyte and placenta DNA (figure 8.1) it was also assessed to provide an additional control (figure 8.3).

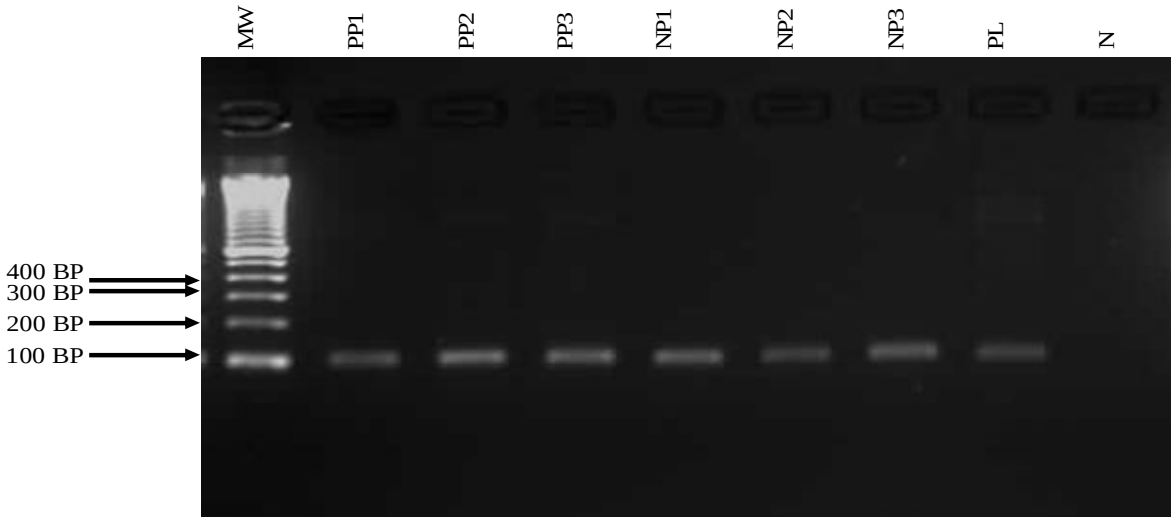


Figure 8.2: Assessment of total cell-free plasma DNA by PCR with *beta globin* primers
Lanes marked PP1-PP3 show PCR products for beta globin from plasma DNAs from pregnant women in their third trimester. Lanes marked NP1-NP3 show PCR products for beta globin from plasma DNAs from non-pregnant women. The lane marked PL is a placental DNA positive control and the lane marked N is a non-template negative control. MW denotes the 100bp DNA ladder (Invitrogen, UK).

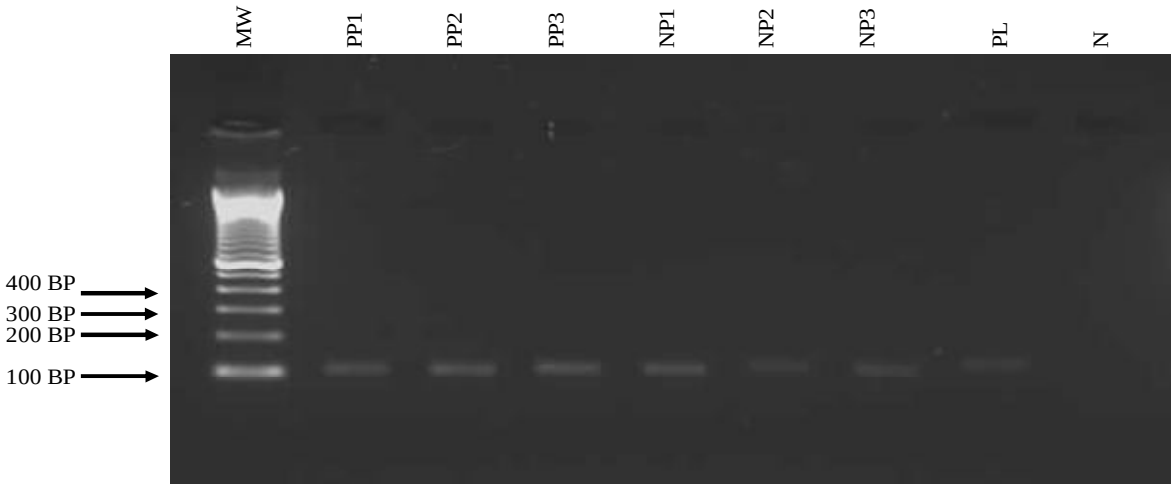


Figure 8.3: Assessment of 21-128 primers on pregnant and non-pregnant plasma DNA

Lanes marked PP1-PP3 show PCR products for 21-128 primers after digestion with *Hpa* II and *Hha* I; on plasma DNAs from pregnant women in their third trimester. Lanes marked NP1-NP3 show PCR products for 21-128 primers after digestion with *Hpa* II and *Hha* I; from plasma DNAs from non-pregnant women. The lane marked PL is a placental DNA positive control and the lane marked N is a non-template negative control. MW denotes the 100bp DNA ladder (Invitrogen, UK). There is a product in all of the samples therefore the 21-128 marker is non-specific for fetal DNA.

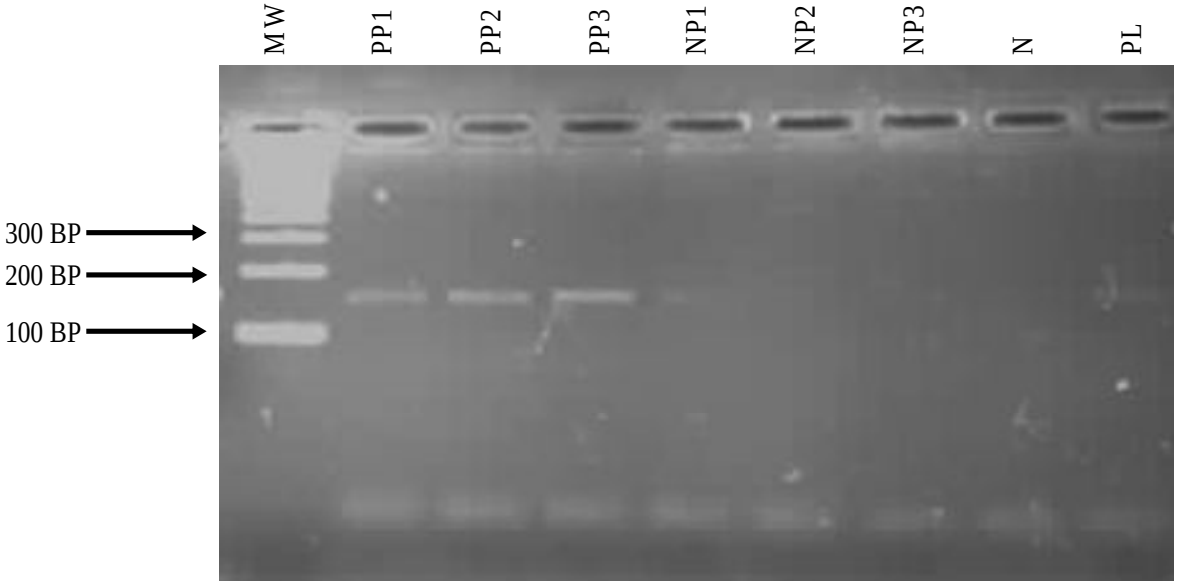


Figure 8.4: Assessment of LM18-58A primers on pregnant and non-pregnant plasma DNA
Lanes marked PP1-PP3 show PCR products for LM18-58A primers after digestion with *Hpa* II and *Hha* I; on plasma DNAs from pregnant women in their third trimester. Lanes marked NP1-NP3 show PCR products for LM18-58A primers after digestion with *Hpa* II and *Hha* I; from plasma DNAs from non-pregnant women. The lane marked PL is a placental DNA positive control and the lane marked N is a non-template negative control. MW denotes the 100bp DNA ladder (Invitrogen, UK). LM18-58A is shown to be fetal specific as there is product band in all three pregnant samples PP1-PP3 and no band in any of three non-pregnant samples.

8.5 Additional testing of the most promising candidate epigenetic biomarkers on plasma DNA

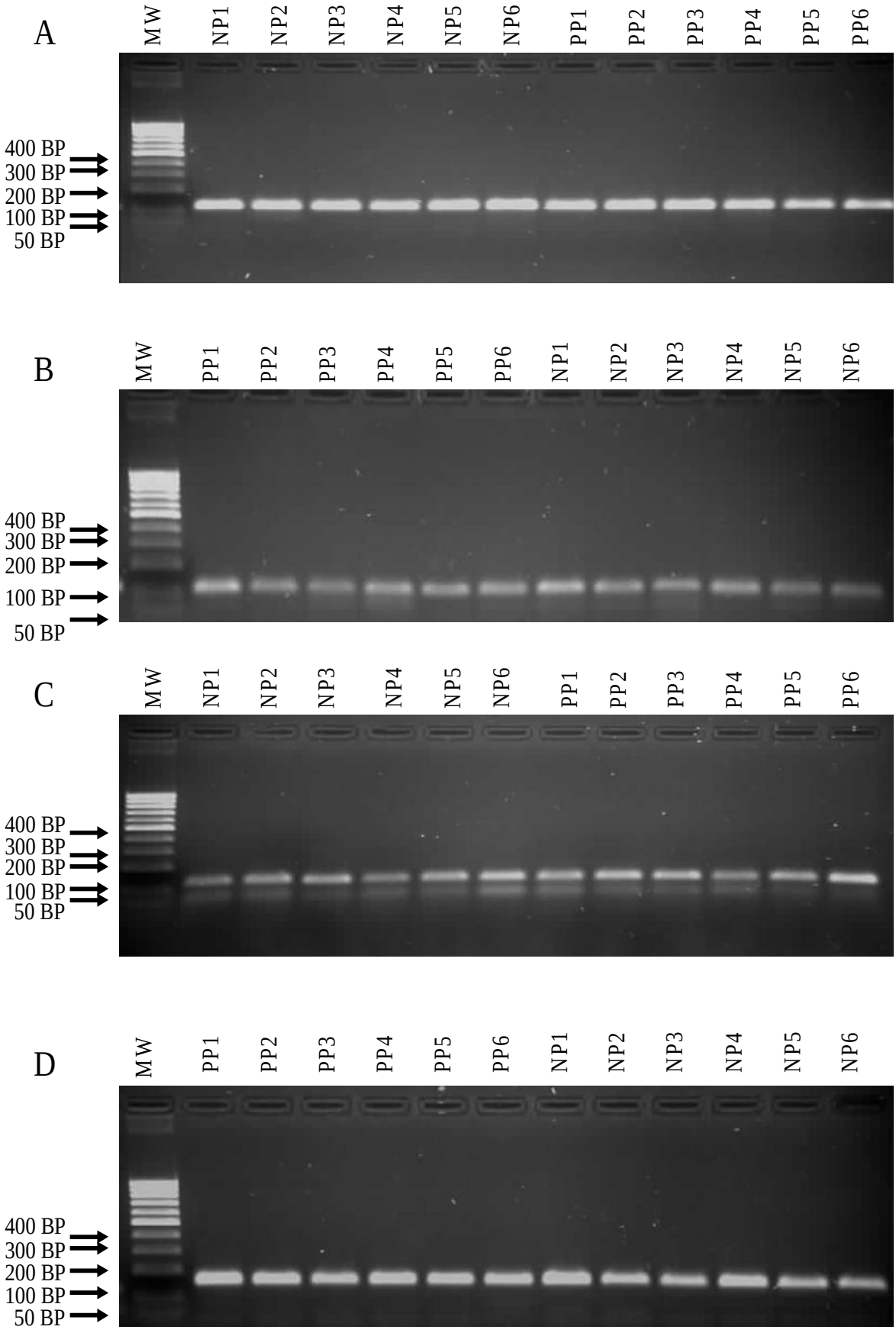
CEBs 21-107, 22-09 and LM18-58A all seemed to be specific for fetal DNA in the maternal circulation, showing distinctive products when used to amplify DNA extracted from the plasma of pregnant women, compared with a distinct lack of amplification when used to amplify DNA extracted from non-pregnant women (figures 8.4) (data for 21-107 and 22-09 is not shown due to low resolution of reproduced images). This indicated that these putative markers were suitable for further testing on plasma DNA samples and could potentially be used in non-invasive prenatal diagnosis of fetal aneuploidies. To that end additional plasma samples were assessed using these CEBs (figure 8.5). In the first instance the total amount of cell free plasma DNA was assessed by the use of *BETA GLOBIN* to

amplify DNA extracted from the plasma of six samples collected from pregnant women and six samples collected from non-pregnant women (figure 8.5 A).

However, of the three CEBs tested, none have been found to be specific for cell-free fetal DNA. All three markers 21-107 (figure 8.5 B), 22-09 (figure 8.5 C) and LM18-58A (figure 8.5 D) each amplified DNA in all of the pregnant and non-pregnant samples tested.

Figure 8.5: Further testing of candidate epigenetic biomarkers with pregnant plasma DNA versus non-pregnant plasma DNA

The total cell-free plasma DNA is assessed using beta globin primers in (A); the total amount of in pregnant samples (PP1-PP6) is approximately equal to that of non-pregnant plasma DNA samples (NP1-NP6). The PCR products for 21-107 (B) are not specific to cff DNA; the PCR product is present in pregnant plasma DNA samples (PP1-PP6) is also present in approximately the same concentration as in the non-pregnant samples (NP1-NP-6). The PCR products for 22-09 (C) are not specific to cff DNA; the product present in pregnant plasma DNA samples (PP1-PP6) is also present in approximately the same concentration as in the non-pregnant samples (NP1-NP-6). The PCR products for LM18-58A (D) are not specific to cff DNA; the product present in pregnant plasma DNA samples (PP1-PP6) is also present in approximately the same concentration as in the non-pregnant samples (NP1-NP-6). MW denotes the 100 bp DNA ladder (Geneflow, UK).


A
MW
NP1
NP2
NP3
NP4
NP5
NP6
PP1
PP2
PP3
PP4
PP5
PP6
400 BP
300 BP
200 BP
100 BP
50 BP
B
MW
PP1
PP2
PP3
PP4
PP5
PP6
NP1
NP2
NP3
NP4
NP5
NP6
400 BP
300 BP
200 BP
100 BP
50 BP
C
MW
NP1
NP2
NP3
NP4
NP5
NP6
PP1
PP2
PP3
PP4
PP5
PP6
400 BP
300 BP
200 BP
100 BP
50 BP
D
MW
PP1
PP2
PP3
PP4
PP5
PP6
NP1
NP2
NP3
NP4
NP5
NP6
400 BP
300 BP
200 BP
100 BP
50 BP

8.6 Further characterisation of the A1 candidate epigenetic biomarker on the *AIRE* gene promoter; with pregnant plasma DNA.

As one of the strongest CEBs the A1 region of the *AIRE* promoter was also assessed; by Real-Time PCR amplification of DNA extracted from plasma samples obtained from pregnant women 24 hours prior to elective delivery and 24 hours after delivery. The PCRs were carefully calibrated with peripheral blood leukocyte serial dilutions. All pregnancies were later verified to be male births. The A1 CEB was compared to the SRY cff DNA biomarker, in order to assess it clearance after delivery. The data show that SRY sequences are cleared rapidly from the maternal plasma after delivery (figure 8.6 A), whereas the unmethylated "fetal" fraction of the A1 DNA sequences are not cleared as rapidly from the maternal circulation, and are still detectable after delivery (figure 8.6 B).

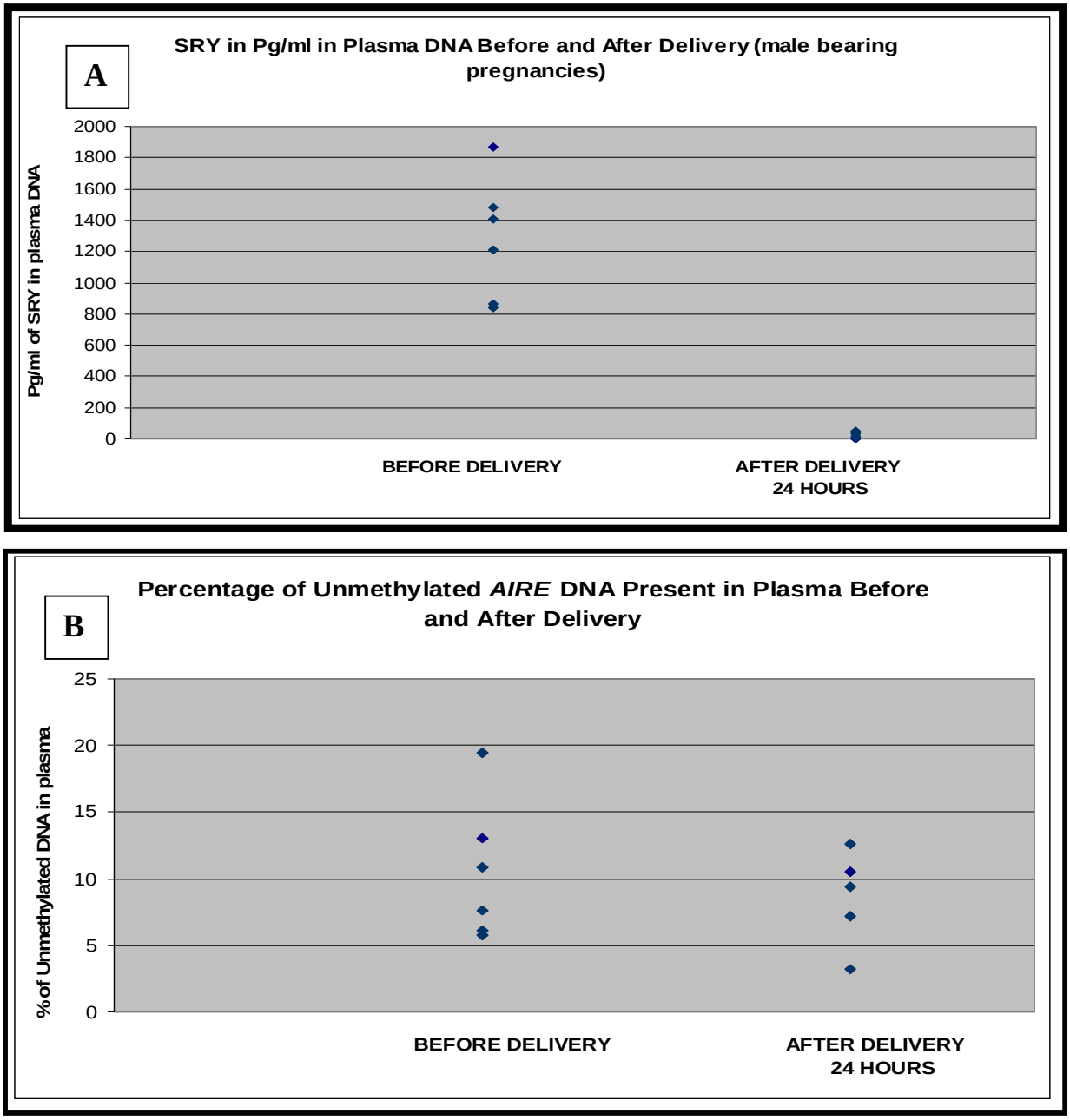


Figure 8.6: Detection of DNA sequences before and after delivery
(A) Shows the SRY sequence is detected in the plasma DNA of pregnant women before delivery, the absolute quantification for the 6 samples assessed ranges from 1867 pg/ml to 840 pg/ml. 24 hours after delivery SRY DNA sequences are not detected in any of the samples tested. (B) Shows the unmethylated A1 cff DNA sequences detected as a fraction of total plasma DNA, before delivery and 24 hours post delivery; there is a trend towards clearance of A1 sequences after delivery although it is not as pronounced as the clearance of SRY sequences.

8.7 Discussion

In order to assess all of the CEBs, primers were redesigned to give shorter amplicons between 70-150 bp in order to ensure maximum efficiency when amplifying the fragmented plasma DNA. The 21-43 primers did not amplify a PCR product and the 12-42 primers were non-specific (figure 8.1). Primers for CEBs: 2-28, 1-16, 1-50, 21-107, 18-CR, 21-58 were clearly differentially methylated between leukocyte and placenta DNA, whereas the 21-128 CEB was only partially differential (figure 8.1). CEBs 21-107, 18-CR, 21-58, 2-28, 1-16 and 1-50 were amenable to further testing on plasma DNA from pregnant women collected in the third trimester.

Several CEBs were assessed using plasma DNA, extracted from whole blood obtained from pregnant women just prior to delivery, with a control group of plasma DNA collected from the whole blood of non-pregnant women. *BETA GLOBIN* was used to assess the total amount of cell-free DNA in each plasma sample (figure 8.2). The levels of total DNA were approximately equal between all samples, meaning that any differences detected would not be due to lower DNA concentrations in any of the samples.

The CEB 21-107 was detected in pregnant plasma DNA and was not detected in the non-pregnant control plasma DNA (data not shown). The presence of a product band indicated that there was a methylated sequence in the plasma of pregnant women which is not present in non-pregnant women. 21-107 appeared to be a promising CEB which required further validation with additional plasma samples.

The CEB 21-128 was also tested with plasma DNA from pregnant women, this marker had been shown to be partially differential when using peripheral blood and placenta DNA (figure 8.1) and it was assumed that this would also be the case in plasma DNA, however 21-128 amplified DNA in both pregnant and non-pregnant DNA equally (figure 8.3), indicating that this CEB is not suitable for further characterisation, for use as a marker in non-invasive prenatal diagnosis.

The CEB 22-09 was also assessed on plasma DNA from pregnant and non-pregnant women (data not shown), this CEB also showed a product in 3 pregnant samples but

not in the 3 non-pregnant samples. These data showed that the 22-09 sequence is methylated in pregnant plasma DNA but not in non-pregnant DNA, which could mean that 22-09 is a suitable for further testing on plasma DNA samples.

Analysis of the LM18-58A CEB showed a product present in the 3 pregnant samples and no product in the 3 non-pregnant samples, except for a very faint band in one sample (figure 8.4). This meant that the sequence for LM18-58A is methylated in pregnant plasma DNA but unmethylated in non-pregnant plasma DNA, and is also suitable for further testing on plasma DNA samples.

In the first round of testing with plasma DNA, the three CEBs tested appeared to work well and were selected for further testing, these were 21-107, 22-09 and LM18-58A; other markers tested on the same plasma samples did not work so well, these were 21-128 (figure 8.3); 21-58 and 21-43 (data not shown). These data indicated that 21-107, 22-09 and LM18-58A were markers for differentially methylated DNA and were apparently detecting cff DNA which is methylated in the pregnant plasma, as well as other CEBs which did not appear to function.

As the number of samples tested was very small (n=3) further testing on plasma DNA was undertaken with a group of 6 pregnant samples and 6 non-pregnant samples of plasma DNA (figure 8.5). The total DNA for each plasma DNA sample was assessed using beta globin (figure 8.5 A). The CEBs 21-107, 22-09 and LM18-58A were assessed using the same plasma DNA samples (figure 8.5 B, C and D respectively). Each of the CEBs amplified a product in both the pregnant and non-pregnant DNA samples. This means that the markers are not consistent and cannot be used for as part of test for non-invasive prenatal diagnosis as they cannot accurately distinguish epigenetic differences between pregnant and non-pregnant plasma DNA samples.

One reason why these CEBs may have worked in the first experiment may be due to inter-individual variation of methylation patterns; it could be that by chance the 3 pregnancies selected were methylated in fetal fraction for all 3 markers but not in the additional 6 samples tested. This is highly unlikely but a possibility none the less. Sample contamination is ruled out, as digestions were carried out using a digestion

in-tube protocol before PCR, therefore no PCR products could have contaminated any prior reactions as PCR products are unmethylated and would be cleaved by either *Hpa* II or *Hha* I.

One CEB has shown some promise, the A1 region of the *AIRE* gene promoter (21-A1). 21-A1 was tested on plasma DNA collected from pregnant women prior to delivery by elective surgery; and on plasma DNA collected from women 24 hours after delivery, as a control SRY was also assessed (figure 8.6). All pregnancies were later confirmed as male-bearing pregnancies. In all samples assessed the SRY sequence was present before delivery and rapidly cleared 24 hours after delivery (figure 8.6 A). 21-A1 sequences were detected before delivery and show a trend towards clearance 24 hours after delivery (figure 8.6 B).

21-A1 may be a partial marker for fetal-specific DNA, and could potentially form part of a larger diagnostic test if more markers were discovered. Even if the markers discovered are not completely specific for cff DNA if there is a large enough panel they could form a diagnostic test for fetal aneuploidy as the power of any diagnostic test increases with the number of markers used.

The difficulty in obtaining epigenetic markers for cff DNA using the model system is an important issue that should be addressed, and any future discovery of cff DNA markers should be done by analysis of first trimester plasma DNA from pregnant women, by combination of either MeDIP or methylation-specific restriction endonuclease analysis, with whole genome amplification and tiling microarray analysis. This would lead to a more accurate and rapid detection of epigenetic biomarkers with a lower rate of marker failure upon validation.

Discussion

Chapter 9

General discussion

9.1 Discussion

Since the discovery of cell-free fetal DNA in the circulation of pregnant women (Lo *et al*, 1997) it has been a major goal to develop fetal-specific biomarkers for the detection of trisomy 21. The main aim of this research project has been to discover epigenetic biomarkers for the detection of fetal DNA; with a view to developing non-invasive prenatal diagnosis of fetal aneuploidy, primarily for the detection of trisomy 21 but also for the detection of trisomy of chromosomes 18 (Edward syndrome) and 13 (Patau syndrome).

The criteria for fetal DNA-specific epigenetic biomarkers are: they must be differentially methylated in order to distinguish cell-free fetal DNA (cff DNA) from cell-free maternal DNA (cfm DNA); they must be independent of fetal sex; they must be detectable during pregnancy and cleared rapidly after delivery and they must be detectable during the first trimester of pregnancy.

Over the last few years potential epigenetic DNA biomarkers have been discovered on chromosomes 18, 3 and 21 these markers: *MASPIN*, *RASSF1A* and *PDE9A* have been detected in the plasma DNA of pregnant women during the third trimester (Chim *et al*, 2005; Chan *et al*, 2006; Chim *et al*, 2008). The discovery of these markers shows that it is possible to detect cff DNA molecules in the plasma of pregnant women without relying on the detection of paternally inherited sequences.

However, these epigenetic biomarkers do not yet fulfil the criteria for use as biomarkers for non-invasive prenatal diagnosis (NIPD), as they have not been shown to accurately detect cff DNA in the plasma of pregnant women, during either the first or second trimesters.

In addition to our discovery of candidate epigenetic biomarkers (CEBs) for fetal aneuploidy, some basic scientific questions are addressed here with regards to the presence and properties of cell-free plasma DNA (cfp DNA).

The knowledge gained from studying the properties of cfp DNA is not only applicable to the field of NIPD; but also to the field of cancer diagnosis and

prognosis; as hypermethylation of tumour suppressor genes and in some cases the hypomethylation of oncogenes has been linked to the progression or recurrence of many forms of cancer including: hepatocellular carcinoma and oesophageal adenocarcinoma (Franco *et al*, 2008; Kawakami *et al*, 2000; Tischoff and Tannapfel, 2008).

One of the properties of cfp DNA which was previously undetermined is the relative abundance of methylated and unmethylated CpG DNA from single copy sequences in plasma. Prior literature on the presence of CpG methylation in the circulation, states that unmethylated CpG DNA may act as a pathogen-associated molecular pattern (PAMP), and elicit an immunogenic response (Cornelie *et al*, 2004; Januchowski *et al*, 2004). The implication of this may be that CpG unmethylated DNA may interact with immune responsive cells in the blood, which may lead to preferential clearance of unmethylated DNA from the circulation.

Another publication suggests that the compact structure of heterochromatin may confer protection to CpG methylated DNA sequences and prevent the action of DNAases present in the blood (Herman, 2004), this may suggest that CpG methylated DNA is not cleared as rapidly as unmethylated DNA. Taking these ideas together it is reasonable to hypothesise that methylated DNA sequences are over-represented in plasma by comparison to unmethylated DNA sequences.

In order to assess the abundance of methylated and unmethylated DNA in plasma a novel system was developed for the quantification of both types of DNA in the circulation. We have developed our own protocol for bisulfite conversion of plasma DNA to enable successful bisulfite sequencing of plasma DNA without over-conversion of methylated cytosines (see chapter 2). Using the XL region of the *GNAS1* imprinted locus it is possible to amplify both the methylated and unmethylated alleles, assessing both DNA types in an unbiased way. Using *GNAS1* as a measure of both methylated and unmethylated DNA it has been possible to determine that CpG methylated and CpG unmethylated DNA sequences are present in equal abundance, circulating in the plasma of healthy individuals.

This conclusion is supported by the data obtained by the amplification of CpG methylated and CpG unmethylated DNA sequences from plasma DNA samples (table 4.1) these data show that CpG methylated DNA sequences and CpG unmethylated sequences do not differ strikingly, and are neither retained nor cleared in a preferential way.

The implication of this data is that it is possible to develop epigenetic biomarkers using cfp DNA, without bias caused by an excess or absence of either CpG methylated or CpG unmethylated DNA. These experiments should be repeated with the use of plasma DNA obtained from pregnant women, primarily from the first trimester and with the use of more imprinted loci, in order to determine if there are any caveats specific to pregnancy, which may alter the ratio of methylated to unmethylated DNA.

An additional discovery emerging from this data is that imprinted loci which coincide with an SNP could be used as fetal-specific DNA biomarkers (figure 3.7). In addition genetic combination of a polymorphic locus and DNA methylation could be used to determine whether or not X chromosome inactivation in the placenta is parent-of-origin-specific in humans, as it is in the mouse (Okamoto *et al*, 2005) where the paternal X chromosome is inactivated in the trophoblasts of female fetuses.

One important property of cfp DNA is the relative abundance of each sequence. Because cfp DNA is released into the circulation largely from cells undergoing apoptosis (Babochkina *et al*, 2005; Gavrieli *et al*, 1992, Kolialexi *et al*, 2004; Sekizawa *et al*, 2004), it may be possible that the DNA released is fragmented or damaged in such a way as to alter the abundance of specific sequences. Previous work has shown that the relative abundance of repeat *ALU* sequences is greater than that of single gene sequences (Stroun *et al*, 2001). However, here we address the separate issue of the relative abundances of single gene sequences compared with each other. This is important as the development of a non-invasive prenatal diagnostic test may require the use of such techniques as relative chromosome dosage (RCD), which compares the relative abundance of DNA sequences on different chromosomes for the detection of trisomy.

The largest difference in sequence abundance detected was between *MPH1* and *Beta-Actin* sequences, showing an average difference of 12.3 fold; even accounting for small discrepancies which may arise from pipetting errors or well-reading, which may occur during PCR, this level of variation in sequence abundance may mean that RCD is not a suitable technique for the quantification of cfp DNA; unless such differences are maintained between individuals. By assessing individual sequences using real-time PCR, a closer analysis of cfp DNA has been achieved, building on knowledge obtained in the original study (Chan *et al*, 2005 b), increasing the understanding of the kinetics of cfp DNA.

This experiment should also be repeated using plasma DNA collected from pregnant women during the first trimester, in order to establish the situation with regards to sequence abundance during pregnancy. Also the relative abundance of cff DNA may also be affected by the routine acquisition of aneuploidy of cytotrophoblasts which invade the maternal circulation and are thought to contribute the majority of cff DNA to the circulation (Fisher *et al*, 1989; Librach *et al*, 1991; Weier *et al*, 2005). However, the relative abundance of fetal DNA sequences can only be assessed in this way once fetal-specific DNA markers have been developed.

In agreement with the data presented by Old and Stallard (Old and Stallard, accepted for publication) the conclusions made here also support the idea that cfp DNA should be assessed by method of allelic ratio. As the data show that the relative abundance of single copy gene sequences are not equal in cfp DNA.

The model system of peripheral blood leukocyte DNA versus term placenta DNA can be used to successfully identify candidate epigenetic biomarkers (CEBs) (Old *et al*, 2007). Out of a total of 366 DNA regions assessed ~3% were found to be differentially methylated between peripheral-blood leukocyte and term placenta DNA. When searching for differentially methylated regions it does not appear to be the case that concerted methylation is exclusive to promoter regions, but may also be detected within intergenic and exogenic regions also.

Using methylation-sensitive restriction endonuclease (MSRE) analysis a panel of CEBs has been identified (some presented Old *et al*, 2007) on chromosomes 21, 18,

12, 2 and 1 (see chapter 5). In addition to MSRE analysis, methylated DNA immunoprecipitation (MeDIP) and combined bisulfite restriction analysis (COBRA) can also be used to identify CEBs (see chapters 6 and 7). Tiling microarrays can be designed to give coverage of entire chromosomes, with the exception of the sequences which make up the gaps between oligonucleotides and regions for which it is not possible to synthesise oligonucleotides such as telomere repeats and CpG rich regions (Weber *et al*, 2005; Wilson *et al*, 2006). Combined with either MeDIP or bisulfite sequencing, tiling microarrays can be used to generate chromosomal methylation maps, which can be used to identify differentially methylated regions.

So far the development of epigenetic biomarkers has led to the development of one epigenetic biomarker which is partially specific for cff DNA; this is the 21-A1 epigenetic biomarker, which shows a trend towards clearance from maternal plasma after delivery. All other CEBs which have been successfully tested on peripheral blood leukocyte and term placenta DNA failed to be validated on pregnant samples; i.e. the placental methylation signature was detected in both pregnant and non-pregnant samples.

This data calls into question the validity of using the model system of peripheral blood leukocyte DNA versus term placenta DNA for the identification of epigenetic biomarkers for the detection of fetal aneuploidy. The epigenetic biomarkers *MASPIN*, *RASSF1A* and *PDE9A* have yet to be successfully developed into a non-invasive diagnostic test as yet, so far it has been demonstrated that the placental DNA methylation signatures of these markers are only detectable in the third trimester (Chim *et al*, 2005; Chan *et al*, 2006; Chim *et al*, 2008). It would appear that term placenta DNA is a poor surrogate for the placental DNA circulating in plasma during the first and second trimesters, and that peripheral blood-leukocyte DNA is a poor substitute for maternal plasma DNA. The most logical way of discovering epigenetic biomarkers for cff DNA, is to develop markers using first trimester plasma DNA from pregnant women.

In the first instance first trimester plasma DNA should be used to confirm the properties of cfp DNA, such as the ratio of methylated to unmethylated DNA and the relative abundance of DNA sequences. Once the general properties of cfp DNA

during pregnancy have been ascertained, the next step is to proceed with the discovery of epigenetic biomarkers.

Pooled samples of first trimester plasma DNA could be either be treated by bisulfite conversion or MeDIP, and applied to tiling microarrays for chromosomes 21, 18 and 13, which have the highest incidence of trisomies. By using a whole chromosome tiling array, the efficiency of biomarker discovery is increased.

If no reliable epigenetic biomarkers were discovered by this method, this experiment would finally end the search for fetal-specific DNA epigenetic biomarkers, for the use in non-invasive prenatal diagnosis of fetal aneuploidy, meaning that efforts should be redirected into either placental RNA biomarker discovery which so far is the most promising approach, or detection of fetal/pregnancy specific proteins in the maternal circulation, in order to improve current screening methods.

However, if a suitable panel of reliable epigenetic biomarkers were discovered, they could be used to form a non-invasive prenatal diagnostic test, which would allow the safe detection of fetal aneuploidy and prevent the loss of pregnancies which are caused by the current invasive techniques. Either way this experiment could provide vital information for the future of non-invasive prenatal diagnosis of fetal aneuploidy.

It is possible that RNA biomarkers such as *PLAC4* may provide suitable coverage of the population for trisomy detection, provided that a sufficient number of heterozygous SNPs can be found to provide informative markers across the population (Chan *et al*, 2006). To date *PLAC4* has been the most successful RNA fetal biomarker discovered, which may mean that other markers may not be as successful.

9.2 Conclusions

Methylated and unmethylated CpG DNA sequences are present in equal abundances, in the plasma of healthy individuals. This discovery shows that there are no specific mechanisms in the blood, which may preferentially clear either methylated or unmethylated CpG DNA sequences. This means that analysis of methylated cfp

DNA is not biased by an excess of either DNA type and could be exploited for the development of epigenetic biomarkers, for use with either NIPD or cancer diagnosis.

Cfp DNA molecules which are derived from single-copy gene sequences are not present in equal abundance in the plasma of normal healthy individuals; this is an important factor to take into account when developing any diagnostic test.

The model system developed by Chim and colleagues (Chim *et al*, 2005) is flawed; and has not been validated by any of the candidate epigenetic biomarkers discovered in this project; any future epigenetic biomarker discovery for the purpose of NIPD should be developed by a simple and direct assessment of first trimester plasma DNA.

It is expected that the issue of rapid and accurate NIPD of fetal trisomy 21 will be solved in the coming years. RNA testing has been reported by Sequenom (see chapter1). However, it is likely that NIPD of trisomy 21 will ultimately be carried out with DNA biomarkers, as the direct assessment of DNA is the most likely source of fetal genetic material to give a direct quantitation of chromosome number.

http://www.bcgdownsyndrome.org/about.cfm (accessed regularly throughout 06/08)

Agren, J., Thiemermann, C., Foster, S. J., Wang, J. E. and Aasen, A. O. (2006). Cytokine responses to CpG DNA in human leukocytes. *Scand.J.Immunol.* **64,** 61-68.

Ahmad-Nejad, P., Hacker, H., Rutz, M., Bauer, S., Vabulas, R. M. and Wagner, H. (2002). Bacterial CpG-DNA and lipopolysaccharides activate Toll-like receptors at distinct cellular compartments. *Eur.J.Immunol.* **32,** 1958-1968.

Alberry, M., Maddocks, D., Jones, M., Abdel Hadi, M., Abdel-Fattah, S., Avent, N. and Soothill, P. W. (2007). Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. *Prenat.Diagn.* **27,** 415-418.

Amicucci, P., Gennarelli, M., Novelli, G. and Dallapiccola, B. (2000). Prenatal diagnosis of myotonic dystrophy using fetal DNA obtained from maternal plasma. *Clin.Chem.* **46,** 301-302.

Angell, R. (1997). First-meiotic-division nondisjunction in human oocytes. *Am.J.Hum.Genet.* **61,** 23-32.

Angert, R. M., Leshane, E. S., Yarnell, R. W., Johnson, K. L. and Bianchi, D. W. (2004). Cell-free fetal DNA in the cerebrospinal fluid of women during the peripartum period. *Am.J.Obstet.Gynecol.* **190,** 1087-1090.

Anker, P. and Stroun, M. (2006). Immunological aspects of circulating DNA. *Ann.N.Y.Acad.Sci.* **1075,** 34-39.

Ariga, H., Ohto, H., Busch, M. P., Imamura, S., Watson, R., Reed, W. and Lee, T. H. (2001). Kinetics of fetal cellular and cell-free DNA in the maternal circulation during and after pregnancy: implications for noninvasive prenatal diagnosis. *Transfusion.* **41,** 1524-1530.

Artlett, C. M., Smith, J. B. and Jimenez, S. A. (1998). Identification of fetal DNA and cells in skin lesions from women with systemic sclerosis. *N.Engl.J.Med.* **338,** 1186-1191.

Avgidou, K., Papageorghiou, A., Bindra, R., Spencer, K. and Nicolaides, K. H. (2005). Prospective first-trimester screening for trisomy 21 in 30,564 pregnancies. *Am.J.Obstet.Gynecol.* **192,** 1761-1767.

Avner, P. and Heard, E. (2001). X-chromosome inactivation: counting, choice and initiation. *Nat.Rev.Genet.* **2,** 59-67.

Babochkina, T., Mergenthaler, S., Dinges, T. M., Holzgreve, W. and Hahn, S. (2005). Direct detection of fetal cells in maternal blood: a reappraisal using a combination of two different Y chromosome-specific FISH probes and a single X chromosome-specific probe. *Arch.Gynecol.Obstet.* **273,** 166-169.

Baker, D. J., Jeganathan, K. B., Cameron, J. D., Thompson, M., Juneja, S., Kopecka, A., Kumar, R., Jenkins, R. B., de Groen, P. C., Roche, P. and van Deursen, J. M. (2004). BubR1 insufficiency causes early onset of aging-associated phenotypes and infertility in mice. *Nat.Genet.* **36,** 744-749.

Bayrak-Toydemir, P., Pergament, E. and Fiddler, M. (2003). Applying a test system for discriminating fetal from maternal cells. *Prenat.Diagn.* **23,** 619-624.

Bestor, T. H. (2000). The DNA methyltransferases of mammals. *Hum.Mol.Genet.* **9,** 2395-2402.

Bianchi, D. W. (2004). Circulating fetal DNA: its origin and diagnostic potential-a review. *Placenta.* **25 Suppl A,** S93-S101.

Bianchi, D. W., Flint, A. F., Pizzimenti, M. F., Knoll, J. H. and Latt, S. A. (1990). Isolation of fetal DNA from nucleated erythrocytes in maternal blood. *Proc.Natl.Acad.Sci.U.S.A.* **87,** 3279-3283.

Bianchi, D. W., LeShane, E. S. and Cowan, J. M. (2001). Large amounts of cell-free fetal DNA are present in amniotic fluid. *Clin.Chem.* **47,** 1867-1869.

Bianchi, D. W., Simpson, J. L., Jackson, L. G., Elias, S., Holzgreve, W., Evans, M. I., Dukes, K. A., Sullivan, L. M., Klinger, K. W., Bischoff, F. Z., Hahn, S., Johnson, K. L., Lewis, D., Wapner, R. J. and de la Cruz, F. (2002). Fetal gender and aneuploidy detection using fetal cells in maternal blood: analysis of NIFTY I data. National Institute of Child Health and Development Fetal Cell Isolation Study. *Prenat.Diagn.* **22,** 609-615.

Bianchi, D. W., Zickwolf, G. K., Weil, G. J., Sylvester, S. and DeMaria, M. A. (1996). Male fetal progenitor cells persist in maternal blood for as long as 27 years postpartum. *Proc.Natl.Acad.Sci.U.S.A.* **93,** 705-708.

Bianchi, D. W., Zickwolf, G. K., Yih, M. C., Flint, A. F., Geifman, O. H., Erikson, M. S. and Williams, J. M. (1993). Erythroid-specific antibodies enhance detection of fetal nucleated erythrocytes in maternal blood. *Prenat.Diagn.* **13,** 293-300.

Birch, L., English, C. A., O'Donoghue, K., Barigye, O., Fisk, N. M. and Keer, J. T. (2005). Accurate and robust quantification of circulating fetal and total DNA in maternal plasma from 5 to 41 weeks of gestation. *Clin.Chem.* **51,** 312-320.

Bird, A. (2007). Perceptions of epigenetics. *Nature.* **447,** 396-398.

Bischoff, F. Z., Lewis, D. E. and Simpson, J. L. (2005). Cell-free fetal DNA in maternal blood: kinetics, source and structure. *Hum.Reprod.Update.* **11,** 59-67.

Bischoff, F. Z., Marquez-Do, D. A., Martinez, D. I., Dang, D., Horne, C., Lewis, D. and Simpson, J. L. (2003). Intact fetal cell isolation from maternal blood: improved isolation using a simple whole blood progenitor cell enrichment approach (RosetteSep). *Clin.Genet.* **63,** 483-489.

Boehm, F. H., Salyer, S. L., Dev, V. G. and Reed, G. W. (1993). Chorionic villus sampling: quality control--a continuous improvement model. *Am.J.Obstet.Gynecol.* **168,** 1766-75; discussion 1775-7.

Borrell, A., Casals, E., Fortuny, A., Farre, M. T., Gonce, A., Sanchez, A., Soler, A., Cararach, V. and Vanrell, J. A. (2004). First-trimester screening for trisomy 21 combining biochemistry and ultrasound at individually optimal gestational ages. An interventional study. *Prenat.Diagn.* **24,** 541-545.

Botezatu, I., Serdyuk, O., Potapova, G., Shelepov, V., Alechina, R., Molyaka, Y., Ananev, V., Bazin, I., Garin, A., Narimanov, M., Knysh, V., Melkonyan, H., Umansky, S. and Lichtenstein, A. (2000). Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. *Clin.Chem.* **46,** 1078-1084.

Brambati, B. and Tului, L. (2005). Chorionic villus sampling and amniocentesis. *Curr.Opin.Obstet.Gynecol.* **17,** 197-201.

Byard, R. W. (2007). Forensic issues in Down syndrome fatalities. *J.Forensic.Leg.Med.* **14,** 475-481.

Cameron, E. E., Bachman, K. E., Myohanen, S., Herman, J. G. and Baylin, S. B. (1999). Synergy of demethylation and histone deacetylase inhibition in the re-expression of genes silenced in cancer. *Nat.Genet.* **21,** 103-107.

Carpenter, A. T. (1994). Chiasma function. *Cell.* **77,** 957-962.

Cassidy, S. B., Lai, L. W., Erickson, R. P., Magnuson, L., Thomas, E., Gendron, R. and Herrmann, J. (1992). Trisomy 15 with loss of the paternal 15 as a cause of Prader-Willi syndrome due to maternal disomy. *Am.J.Hum.Genet.* **51,** 701-708.

Chan, K. C., Ding, C., Gerovassili, A., Yeung, S. W., Chiu, R. W., Leung, T. N., Lau, T. K., Chim, S. S., Chung, G. T., Nicolaides, K. H. and Lo, Y. M. (2006). Hypermethylated RASSF1A in maternal plasma: A universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. *Clin.Chem.* **52,** 2211-2218.

Chan, K. C., Hui, A. B., Wong, N., Lau, T. K., Leung, T. N., Lo, K. W. and Lo, Y. M. (2005). Investigation of the genomic representation of plasma DNA in pregnant women by comparative genomic hybridization analysis: a feasibility study. *Clin.Chem.* **51,** 2398-2401.

Chan, K. C., Yeung, S. W., Lui, W. B., Rainer, T. H. and Lo, Y. M. (2005). Effects of preanalytical factors on the molecular size of cell-free DNA in blood. *Clin.Chem.* **51,** 781-784.

Chan, K. C., Zhang, J., Hui, A. B., Wong, N., Lau, T. K., Leung, T. N., Lo, K. W., Huang, D. W. and Lo, Y. M. (2004). Size distributions of maternal and fetal DNA in maternal plasma. *Clin.Chem.* **50,** 88-92.

Chen, H. P., Wang, T. R., Xu, J. P., Xu, X. Y., Dangol, S. D. and He, G. F. (2004). Fetal origin of single nucleated erythroblasts and free DNA in the peripheral blood of pregnant women. *Int.J.Gynaecol.Obstet.* **85,** 1-5.

Chen, Z. X., Mann, J. R., Hsieh, C. L., Riggs, A. D. and Chedin, F. (2005). Physical and functional interactions between the human DNMT3L protein and members of the de novo methyltransferase family. *J.Cell.Biochem.* **95,** 902-917.

Chim, S. S., Jin, S., Lee, T. Y., Lun, F. M., Lee, W. S., Chan, L. Y., Jin, Y., Yang, N., Tong, Y. K., Leung, T. Y., Lau, T. K., Ding, C., Chiu, R. W. and Lo, Y. M. (2008). Systematic search for placental DNA-methylation markers on chromosome 21: toward a maternal plasma-based epigenetic test for fetal trisomy 21. *Clin.Chem.* **54,** 500-511.

Chim, S. S., Tong, Y. K., Chiu, R. W., Lau, T. K., Leung, T. N., Chan, L. Y., Oudejans, C. B., Ding, C. and Lo, Y. M. (2005). Detection of the placental epigenetic signature of the maspin gene in maternal plasma. *Proc.Natl.Acad.Sci.U.S.A.* **102,** 14753-14758.

Chinnapapagari, S. K., Holzgreve, W., Lapaire, O., Zimmermann, B. and Hahn, S. (2005). Treatment of maternal blood samples with formaldehyde does not alter the proportion of circulatory fetal nucleic acids (DNA and mRNA) in maternal plasma. *Clin.Chem.* **51,** 652-655.

Chiu, R. W., Lau, T. K., Cheung, P. T., Gong, Z. Q., Leung, T. N. and Lo, Y. M. (2002). Noninvasive prenatal exclusion of congenital adrenal hyperplasia by maternal plasma analysis: a feasibility study. *Clin.Chem.* **48,** 778-780.

Chiu, R. W., Lau, T. K., Leung, T. N., Chow, K. C., Chui, D. H. and Lo, Y. M. (2002). Prenatal exclusion of beta thalassaemia major by examination of maternal plasma. *Lancet.* **360,** 998-1000.

Chiu, R. W., Lui, W. B., El-Sheikhah, A., Chan, A. T., Lau, T. K., Nicolaides, K. H. and Lo, Y. M. (2005). Comparison of protocols for extracting circulating DNA and RNA from maternal plasma. *Clin.Chem.* **51,** 2209-2210.

Chung, G. T., Chiu, R. W., Chan, K. C., Lau, T. K., Leung, T. N. and Lo, Y. M. (2005). Lack of dramatic enrichment of fetal DNA in maternal plasma by formaldehyde treatment. *Clin.Chem.* **51,** 655-658.

Cicero, S., Rembouskos, G., Vandecruys, H., Hogg, M. and Nicolaides, K. H. (2004). Likelihood ratio for trisomy 21 in fetuses with absent nasal bone at the 11-14-week scan. *Ultrasound Obstet.Gynecol.* **23,** 218-223.

Clerc, P. and Avner, P. (1998). Role of the region 3' to Xist exon 6 in the counting process of X-chromosome inactivation. *Nat.Genet.* **19,** 249-253.

Cornelie, S., Poulain-Godefroy, O., Lund, C., Vendeville, C., Ban, E., Capron, M. and Riveau, G. (2004). Methylated CpG-containing plasmid activates the immune system. *Scand.J.Immunol.* **59,** 143-151.

Costa, J. M., Gautier, E. and Benachi, A. (2004). Genetic analysis of the fetus using maternal blood. *Gynecol.Obstet.Fertil.* **32,** 646-650.

Cuckle, H. S. and van Lith, J. M. (1999). Appropriate biochemical parameters in first-trimester screening for Down syndrome. *Prenat.Diagn.* **19,** 505-512.

Dahl, C. and Guldberg, P. (2003). DNA methylation analysis techniques. *Biogerontology.* **4,** 233-250.

Dhallan, R., Au, W. C., Mattagajasingh, S., Emche, S., Bayliss, P., Damewood, M., Cronin, M., Chou, V. and Mohr, M. (2004). Methods to increase the percentage of free fetal DNA recovered from the maternal circulation. *JAMA.* **291,** 1114-1119.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A., Haefliger, C., Horton, R., Howe, K., Jackson, D. K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K. and Beck, S. (2006). DNA methylation profiling of human chromosomes 6, 20 and 22. *Nat.Genet.* **38,** 1378-1385.

Eddleman, K. A., Malone, F. D., Sullivan, L., Dukes, K., Berkowitz, R. L., Kharbutli, Y., Porter, T. F., Luthy, D. A., Comstock, C. H., Saade, G. R., Klugman, S., Dugoff, L., Craigo, S. D., Timor-Tritsch, I. E., Carr, S. R., Wolfe, H. M. and D'Alton, M. E. (2006). Pregnancy loss rates after midtrimester amniocentesis. *Obstet.Gynecol.* **108,** 1067-1072.

Eden, S. and Cedar, H. (1994). Role of DNA methylation in the regulation of transcription. *Curr.Opin.Genet.Dev.* **4,** 255-259.

Eggermann, T., Meyer, E., Schonherr, N., Flick, F., Chauvistre, H., Mavany, M. and Wollmann, H. A. (2007). Mutation analysis of GNAS1 and overlapping transcripts in Silver-Russell syndrome patients. *Mol.Genet.Metab.* **90,** 224-226.

Elias, S., Lewis, D. E., Bischoff, F. Z. and Simpson, J. L. (1996). Isolation and genetic analysis of fetal nucleated red blood cells from maternal blood: the Baylor College of Medicine experience. *Early Hum.Dev.* **47 Suppl,** S85-8.

Filkins, K. and Koos, B. J. (2005). Ultrasound and fetal diagnosis. *Curr.Opin.Obstet.Gynecol.* **17,** 185-195.

Finning, K. M., Martin, P. G., Soothill, P. W. and Avent, N. D. (2002). Prediction of fetal D status from maternal plasma: introduction of a new noninvasive fetal RHD genotyping service. *Transfusion.* **42,** 1079-1085.

Firth, H. V., Boyd, P. A., Chamberlain, P., MacKenzie, I. Z., Lindenbaum, R. H. and Huson, S. M. (1991). Severe limb abnormalities after chorion villus sampling at 56-66 days' gestation. *Lancet.* **337,** 762-763.

Fisher, S. J., Cui, T. Y., Zhang, L., Hartman, L., Grahl, K., Zhang, G. Y., Tarpey, J. and Damsky, C. H. (1989). Adhesive and degradative properties of human placental cytotrophoblast cells in vitro. *J.Cell Biol.* **109,** 891-902.

Franco, R., Schoneveld, O., Georgakilas, A. G. and Panayiotidis, M. I. (2008). Oxidative stress, DNA methylation and carcinogenesis. *Cancer Lett.* **266,** 6-11.

Freeman, S. B., Allen, E. G., Oxford-Wright, C. L., Tinker, S. W., Druschel, C., Hobbs, C. A., O'Leary, L. A., Romitti, P. A., Royle, M. H., Torfs, C. P. and Sherman, S. L. (2007). The National Down Syndrome Project: design and implementation. *Public Health Rep.* **122,** 62-72.

Frid, C., Drott, P., Lundell, B., Rasmussen, F. and Anneren, G. (1999). Mortality in Down's syndrome in relation to congenital malformations. *J.Intellect.Disabil.Res.* **43 (Pt 3),** 234-241.

Frommer, M., McDonald, L. E., Millar, D. S., Collis, C. M., Watt, F., Grigg, G. W., Molloy, P. L. and Paul, C. L. (1992). A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc.Natl.Acad.Sci.U.S.A.* **89,** 1827-1831.

Galbiati, S., Smid, M., Gambini, D., Ferrari, A., Restagno, G., Viora, E., Campogrande, M., Bastonero, S., Pagliano, M., Calza, S., Ferrari, M. and Cremonesi, L. (2005). Fetal DNA detection in maternal plasma throughout gestation. *Hum.Genet.* **117,** 243-248.

Gavrieli, Y., Sherman, Y. and Ben-Sasson, S. A. (1992). Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. *J.Cell Biol.* **119,** 493-501.

Giacona, M. B., Ruben, G. C., Iczkowski, K. A., Roos, T. B., Porter, D. M. and Sorenson, G. D. (1998). Cell-free DNA in human blood plasma: length measurements in patients with pancreatic cancer and healthy controls. *Pancreas.* **17,** 89-97.

Glass, R. H., Aggeler, J., Spindle, A., Pedersen, R. A. and Werb, Z. (1983). Degradation of extracellular matrix by mouse trophoblast outgrowths: a model for implantation. *J.Cell Biol.* **96,** 1108-1116.

Grunau, C., Clark, S. J. and Rosenthal, A. (2001). Bisulfite genomic sequencing: systematic investigation of critical experimental parameters. *Nucleic Acids Res.* **29,** E65-5.

Guetta, E., Gordon, D., Simchen, M. J., Goldman, B. and Barkai, G. (2003). Hematopoietic progenitor cells as targets for non-invasive prenatal diagnosis: detection of fetal CD34+ cells and assessment of post-delivery persistence in the maternal circulation. *Blood Cells Mol.Dis.* **30,** 13-21.

Guibert, J., Benachi, A., Grebille, A. G., Ernault, P., Zorn, J. R. and Costa, J. M. (2003). Kinetics of SRY gene appearance in maternal serum: detection by real time PCR in early pregnancy after assisted reproductive technique. *Hum.Reprod.* **18,** 1733-1736.

Guldberg, P., Worm, J. and Gronbaek, K. (2002). Profiling DNA methylation by melting analysis. *Methods.* **27,** 121-127.

Gussin, H. A., Bischoff, F. Z., Hoffman, R. and Elias, S. (2002). Endothelial precursor cells in the peripheral blood of pregnant women. *J.Soc.Gynecol.Investig.* **9,** 357-361.

Hasle, H., Clemmensen, I. H. and Mikkelsen, M. (2000). Risks of leukaemia and solid tumours in individuals with Down's syndrome. *Lancet.* **355,** 165-169.

Hasselmann, D. O., Rappl, G., Tilgen, W. and Reinhold, U. (2001). Extracellular tyrosinase mRNA within apoptotic bodies is protected from degradation in human serum. *Clin.Chem.* **47,** 1488-1489.

Hassold, T., Hall, H. and Hunt, P. (2007). The origin of human aneuploidy: where we have been, where we are going. *Hum.Mol.Genet.* **16 Spec No. 2,** R203-8.

Hassold, T. and Hunt, P. (2001). To err (meiotically) is human: the genesis of human aneuploidy. *Nat.Rev.Genet.* **2,** 280-291.

Hemmi, H., Takeuchi, O., Kawai, T., Kaisho, T., Sato, S., Sanjo, H., Matsumoto, M., Hoshino, K., Wagner, H., Takeda, K. and Akira, S. (2000). A Toll-like receptor recognizes bacterial DNA. *Nature.* **408,** 740-745.

Herman, J. G. (2004). Circulating methylated DNA. *Ann.N.Y.Acad.Sci.* **1022,** 33-39.

Hermann, A., Schmitt, S. and Jeltsch, A. (2003). The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J.Biol.Chem.* **278,** 31717-31721.

Hitzler, J. K. (2007). Acute megakaryoblastic leukemia in Down syndrome. *Pediatr.Blood Cancer.* **49,** 1066-1069.

Hook, E. B. (1995). Paternal age and Down syndrome in British Columbia. *Epidemiology*. **6,** 640-641.

Hook, E. B. and Cross, P. K. (1983). Spontaneous abortion and subsequent Down syndrome livebirth. *Hum.Genet.* **64,** 267-270.

Howell, C. Y., Bestor, T. H., Ding, F., Latham, K. E., Mertineit, C., Trasler, J. M. and Chaillet, J. R. (2001). Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell.* **104,** 829-838.

Hunt, P., LeMaire, R., Embury, P., Sheean, L. and Mroz, K. (1995). Analysis of chromosome behavior in intact mammalian oocytes: monitoring the segregation of a univalent chromosome during female meiosis. *Hum.Mol.Genet.* **4,** 2007-2012.

Ingargiola, I., Vaerman, J. L., Debieve, F., Palgen, G., Verellen-Dumoulin, C. and Hubinont, C. (2003). Free fetal DNA concentration in maternal plasma during normal labour at term. *Prenat.Diagn*. **23,** 1077-1082.

Irniger, S. (2006). Preventing fatal destruction: inhibitors of the anaphase-promoting complex in meiosis. *Cell.Cycle.* **5,** 405-415.

Jahr, S., Hentze, H., Englisch, S., Hardt, D., Fackelmayer, F. O., Hesch, R. D. and Knippers, R. (2001). DNA fragments in the blood plasma of cancer patients: quantitations and evidence for their origin from apoptotic and necrotic cells. *Cancer Res.* **61,** 1659-1665.

Januchowski, R., Prokop, J. and Jagodzinski, P. P. (2004). Role of epigenetic DNA alterations in the pathogenesis of systemic lupus erythematosus. *J.Appl.Genet.* **45,** 237-248.

Jyothy, A., Kumar, K. S., Mallikarjuna, G. N., Babu Rao, V., Uma Devi, B., Sujatha, M. and Reddy, P. P. (2001). Parental age and the origin of extra chromosome 21 in Down syndrome. *J.Hum.Genet.* **46,** 347-350.

Kawakami, K., Brabender, J., Lord, R. V., Groshen, S., Greenwald, B. D., Krasna, M. J., Yin, J., Fleisher, A. S., Abraham, J. M., Beer, D. G., Sidransky, D., Huss, H. T., Demeester, T. R., Eads, C., Laird, P. W., Ilson, D. H., Kelsen, D. P., Harpole, D.,

Moore, M. B., Danenberg, K. D., Danenberg, P. V. and Meltzer, S. J. (2000). Hypermethylated APC DNA in plasma and prognosis of patients with esophageal adenocarcinoma. *J.Natl.Cancer Inst.* **92,** 1805-1811.

Kazazian, H. H.,Jr and Moran, J. V. (1998). The impact of L1 retrotransposons on the human genome. *Nat.Genet.* **19,** 19-24.

Kierszenbaum, A. L. (2002). Genomic imprinting and epigenetic reprogramming: unearthing the garden of forking paths. *Mol.Reprod.Dev.* **63,** 269-272.

Klose, R. J. and Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem.Sci.* **31,** 89-97.

Koide, K., Sekizawa, A., Iwasaki, M., Matsuoka, R., Honma, S., Farina, A., Saito, H. and Okai, T. (2005). Fragmentation of cell-free fetal DNA in plasma and urine of pregnant women. *Prenat.Diagn.* **25,** 604-607.

Kolialexi, A., Tsangaris, G. T., Antsaklis, A., Kitsiou-Tzeli, S. and Mavrou, A. (2004). Use of annexin V for the identification of fetal cells in maternal circulation. *In Vivo*. **18,** 629-632.

Kolvraa, S., Christensen, B., Lykke-Hansen, L. and Philip, J. (2005). The fetal erythroblast is not the optimal target for non-invasive prenatal diagnosis: preliminary results. *J.Histochem.Cytochem.* **53,** 331-336.

Kopreski, M. S., Benko, F. A., Kwak, L. W. and Gocke, C. D. (1999). Detection of tumor messenger RNA in the serum of patients with malignant melanoma. *Clin.Cancer Res.* **5,** 1961-1965.

Krabchi, K., Gadji, M., Samassekou, O., Gregoire, M. C., Forest, J. C. and Drouin, R. (2006). Quantification of fetal nucleated cells in maternal blood of pregnant women with a male trisomy 21 fetus using molecular cytogenetic techniques. *Prenat.Diagn.* **26,** 28-34.

Krabchi, K., Gros-Louis, F., Yan, J., Bronsard, M., Masse, J., Forest, J. C. and Drouin, R. (2001). Quantification of all fetal nucleated cells in maternal blood between the 18th and 22nd weeks of pregnancy using molecular cytogenetic techniques. *Clin.Genet.* **60,** 145-150.

Lapaire, O., Holzgreve, W., Oosterwijk, J. C., Brinkhaus, R. and Bianchi, D. W. (2007). Georg Schmorl on trophoblasts in the maternal circulation. *Placenta.* **28,** 1-5.

Ledbetter, D. H., Martin, A. O., Verlinsky, Y., Pergament, E., Jackson, L., Yang-Feng, T., Schonberg, S. A., Gilbert, F., Zachary, J. M. and Barr, M. (1990). Cytogenetic results of chorionic villus sampling: high success rate and diagnostic accuracy in the United States collaborative study. *Am.J.Obstet.Gynecol.* **162,** 495-501.

Ledbetter, D. H., Zachary, J. M., Simpson, J. L., Golbus, M. S., Pergament, E., Jackson, L., Mahoney, M. J., Desnick, R. J., Schulman, J. and Copeland, K. L. (1992). Cytogenetic results from the U.S. Collaborative Study on CVS. *Prenat.Diagn.* **12,** 317-345.

Lee, J., Inoue, K., Ono, R., Ogonuki, N., Kohda, T., Kaneko-Ishino, T., Ogura, A. and Ishino, F. (2002). Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development.* **129,** 1807-1817.

Li, E. (2002). Chromatin modification and epigenetic reprogramming in mammalian development. *Nat.Rev.Genet.* **3,** 662-673.

Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell.* **69,** 915-926.

Li, L. C. and Dahiya, R. (2002). MethPrimer: designing primers for methylation PCRs. *Bioinformatics.* **18,** 1427-1431.

Li, Y., Holzgreve, W., Page-Christiaens, G. C., Gille, J. J. and Hahn, S. (2004). Improved prenatal detection of a fetal point mutation for achondroplasia by the use of size-fractionated circulatory DNA in maternal plasma--case report. *Prenat.Diagn.* **24,** 896-898.

Li, Y., Zimmermann, B., Rusterholz, C., Kang, A., Holzgreve, W. and Hahn, S. (2004). Size separation of circulatory DNA in maternal plasma permits ready detection of fetal DNA polymorphisms. *Clin.Chem.* **50,** 1002-1011.

Librach, C. L., Werb, Z., Fitzgerald, M. L., Chiu, K., Corwin, N. M., Esteves, R. A., Grobelny, D., Galardy, R., Damsky, C. H. and Fisher, S. J. (1991). 92-kD type IV collagenase mediates invasion of human cytotrophoblasts. *J.Cell Biol.* **113,** 437-449.

Lo, K. W., Lo, Y. M., Leung, S. F., Tsang, Y. S., Chan, L. Y., Johnson, P. J., Hjelm, N. M., Lee, J. C. and Huang, D. P. (1999). Analysis of cell-free Epstein-Barr virus associated RNA in the plasma of patients with nasopharyngeal carcinoma. *Clin.Chem.* **45,** 1292-1294.

Lo, Y. M. (2005). Recent advances in fetal nucleic acids in maternal plasma. *J.Histochem.Cytochem.* **53,** 293-296.

Lo, Y. M., Corbetta, N., Chamberlain, P. F., Rai, V., Sargent, I. L., Redman, C. W. and Wainscoat, J. S. (1997). Presence of fetal DNA in maternal plasma and serum. *Lancet.* **350,** 485-487.

Lo, Y. M., Hjelm, N. M., Fidler, C., Sargent, I. L., Murphy, M. F., Chamberlain, P. F., Poon, P. M., Redman, C. W. and Wainscoat, J. S. (1998). Prenatal diagnosis of fetal RhD status by molecular analysis of maternal plasma. *N.Engl.J.Med.* **339,** 1734-1738.

Lo, Y. M., Tein, M. S., Lau, T. K., Haines, C. J., Leung, T. N., Poon, P. M., Wainscoat, J. S., Johnson, P. J., Chang, A. M. and Hjelm, N. M. (1998). Quantitative analysis of fetal DNA in maternal plasma and serum: implications for noninvasive prenatal diagnosis. *Am.J.Hum.Genet.* **62,** 768-775.

Lo, Y. M., Tsui, N. B., Chiu, R. W., Lau, T. K., Leung, T. N., Heung, M. M., Gerovassili, A., Jin, Y., Nicolaides, K. H., Cantor, C. R. and Ding, C. (2007). Plasma placental RNA allelic ratio permits noninvasive prenatal chromosomal aneuploidy detection. *Nat.Med.* **13,** 218-223.

Lo, Y. M., Zhang, J., Leung, T. N., Lau, T. K., Chang, A. M. and Hjelm, N. M. (1999). Rapid clearance of fetal DNA from maternal plasma. *Am.J.Hum.Genet.* **64,** 218-224.

Lui, Y. Y., Chik, K. W., Chiu, R. W., Ho, C. Y., Lam, C. W. and Lo, Y. M. (2002). Predominant hematopoietic origin of cell-free DNA in plasma and serum after sex-mismatched bone marrow transplantation. *Clin.Chem.* **48,** 421-427.

Manotaya, S., Elias, S., Lewis, D. E., Simpson, J. L. and Bischoff, F. Z. (2002). Evaluation of a culture system for enrichment of CD34+ hematopoietic progenitor cells present in maternal blood. *Fetal.Diagn.Ther.* **17,** 90-96.

Maymon, R., Levinsohn-Tavor, O., Cuckle, H., Tovbin, Y., Dreazen, E., Wiener, Y. and Herman, A. (2005). Second trimester ultrasound prenasal thickness combined with nasal bone length: a new method of Down syndrome screening. *Prenat.Diagn.* **25,** 906-911.

Miranda, T. B. and Jones, P. A. (2007). DNA methylation: the nuts and bolts of repression. *J.Cell.Physiol.* **213,** 384-390.

Montagna, M., Santacatterina, M., Torri, A., Menin, C., Zullato, D., Chieco-Bianchi, L. and D'Andrea, E. (1999). Identification of a 3 kb Alu-mediated BRCA1 gene rearrangement in two breast/ovarian cancer families. *Oncogene.* **18,** 4160-4165.

Morelli, M. A. and Cohen, P. E. (2005). Not all germ cells are created equal: aspects of sexual dimorphism in mammalian meiosis. *Reproduction.* **130,** 761-781.

Mujezinovic, F. and Alfirevic, Z. (2007). Procedure-related complications of amniocentesis and chorionic villous sampling: a systematic review. *Obstet.Gynecol.* **110,** 687-694.

Mulcahy, H. E., Croke, D. T. and Farthing, M. J. (1996). Cancer and mutant DNA in blood plasma. *Lancet.* **348,** 628.

Muller, F., Rebiffe, M., Taillandier, A., Oury, J. F. and Mornet, E. (2000). Parental origin of the extra chromosome in prenatally diagnosed fetal trisomy 21. *Hum.Genet.* **106,** 340-344.

Murumagi, A., Vahamurto, P. and Peterson, P. (2003). Characterization of regulatory elements and methylation pattern of the autoimmune regulator (AIRE) promoter. *J.Biol.Chem.* **278,** 19784-19790.

Nan, X., Campoy, F. J. and Bird, A. (1997). MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell.* **88,** 471-481.

Ng, E. K., Leung, T. N., Tsui, N. B., Lau, T. K., Panesar, N. S., Chiu, R. W. and Lo, Y. M. (2003). The concentration of circulating corticotropin-releasing hormone mRNA in maternal plasma is increased in preeclampsia. *Clin.Chem.* **49,** 727-731.

Ng, E. K., Tsui, N. B., Lam, N. Y., Chiu, R. W., Yu, S. C., Wong, S. C., Lo, E. S., Rainer, T. H., Johnson, P. J. and Lo, Y. M. (2002). Presence of filterable and

nonfilterable mRNA in the plasma of cancer patients and healthy individuals. *Clin.Chem.* **48,** 1212-1217.

Ng, E. K., Tsui, N. B., Lau, T. K., Leung, T. N., Chiu, R. W., Panesar, N. S., Lit, L. C., Chan, K. W. and Lo, Y. M. (2003). mRNA of placental origin is readily detectable in maternal plasma. *Proc.Natl.Acad.Sci.U.S.A.* **100,** 4748-4753.

Nicolaides, K. H. (2005). First-trimester screening for chromosomal abnormalities. *Semin.Perinatol.* **29,** 190-194.

Nicolaides, K. H. (2004). Nuchal translucency and other first-trimester sonographic markers of chromosomal abnormalities. *Am.J.Obstet.Gynecol.* **191,** 45-67.

Nicolaides, K. H., Chervenak, F. A., McCullough, L. B., Avgidou, K. and Papageorghiou, A. (2005). Evidence-based obstetric ethics and informed decision-making by pregnant women about invasive diagnosis after first-trimester assessment of risk for trisomy 21. *Am.J.Obstet.Gynecol.* **193,** 322-326.

Nicolaides, K. H., Spencer, K., Avgidou, K., Faiola, S. and Falcon, O. (2005). Multicenter study of first-trimester screening for trisomy 21 in 75 821 pregnancies: results and estimation of the potential impact of individual risk-orientated two-stage first-trimester screening. *Ultrasound Obstet.Gynecol.* **25,** 221-226.

Niemimaa, M., Suonpaa, M., Perheentupa, A., Seppala, M., Heinonen, S., Laitinen, P., Ruokonen, A. and Ryynanen, M. (2001). Evaluation of first trimester maternal serum and ultrasound screening for Down's syndrome in Eastern and Northern Finland. *Eur.J.Hum.Genet.* **9,** 404-408.

Ohashi, Y., Miharu, N., Honda, H., Samura, O. and Ohama, K. (2002). Correlation of fetal DNA and human chorionic gonadotropin concentrations in second-trimester maternal serum. *Clin.Chem.* **48,** 386-388.

Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteche, C. M., Avner, P. and Heard, E. (2005). Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature.* **438,** 369-373.

Okano, M., Bell, D. W., Haber, D. A. and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell.* **99,** 247-257.

Oliver, T. R., Feingold, E., Yu, K., Cheung, V., Tinker, S., Yadav-Shah, M., Masse, N. and Sherman, S. L. (2008). New insights into human nondisjunction of chromosome 21 in oocytes. *PLoS Genet.* **4,** e1000033.

Poon, L. L., Leung, T. N., Lau, T. K. and Lo, Y. M. (2000). Presence of fetal RNA in maternal plasma. *Clin.Chem.* **46,** 1832-1834.

Ptashne, M. (2007). On the use of the word 'epigenetic'. *Curr.Biol.* **17,** R233-6.

Puszyk, W. M., Crea, F. and Old, R. W. (2008). Noninvasive prenatal diagnosis of aneuploidy using cell-free nucleic acids in maternal blood: promises and unanswered questions. *Prenat.Diagn.* **28,** 1-6.

Reddi, K. K. and Holland, J. F. (1976). Elevated serum ribonuclease in patients with pancreatic cancer. *Proc.Natl.Acad.Sci.U.S.A.* **73,** 2308-2310.

Rijnders, R. J., van der Schoot, C. E., Bossers, B., de Vroede, M. A. and Christiaens, G. C. (2001). Fetal sex determination from maternal plasma in pregnancies at risk for congenital adrenal hyperplasia. *Obstet.Gynecol.* **98,** 374-378.

Robertson, K. D. and Wolffe, A. P. (2000). DNA methylation in health and disease. *Nat.Rev.Genet.* **1,** 11-19.

Rodriguez De Alba, M., Palomino, P., Jurado, A., Sanz, R., Ibanez, M. A., Fernandez-Moya, J. M., Ayuso, C., Diaz-Recasens, J., Lahoz, C. and Ramos, C. (1999). Prenatal diagnosis on fetal cells obtained from maternal peripheral blood: report of 66 cases. *Prenat.Diagn.* **19,** 934-940.

Rozen, S. and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. *Methods Mol.Biol.* **132,** 365-386.

Santacroce, R., Vecchione, G., Tomaiyolo, M., Sessa, F., Sarno, M., Colaizzo, D., Grandone, E. and Margaglione, M. (2006). Identification of fetal gender in maternal blood is a helpful tool in the prenatal diagnosis of haemophilia. *Haemophilia.* **12,** 417-422.

Scholl, T., Stein, Z. and Hansen, H. (1982). Leukemia and other cancers, anomalies and infections as causes of death in Down's syndrome in the United States during 1976. *Dev.Med.Child Neurol.* **24,** 817-829.

Sekizawa, A., Purwosunu, Y., Matsuoka, R., Koide, K., Okazaki, S., Farina, A., Saito, H. and Okai, T. (2007). Recent advances in non-invasive prenatal DNA diagnosis through analysis of maternal blood. *J.Obstet.Gynaecol.Res.* **33,** 747-764.

Sekizawa, A., Samura, O., Zhen, D. K., Falco, V., Farina, A. and Bianchi, D. W. (2000). Apoptosis in fetal nucleated erythrocytes circulating in maternal blood. *Prenat.Diagn.* **20,** 886-889.

Sekizawa, A., Sugito, Y., Iwasaki, M., Watanabe, A., Jimbo, M., Hoshi, S., Saito, H. and Okai, T. (2001). Cell-free fetal DNA is increased in plasma of women with hyperemesis gravidarum. *Clin.Chem.* **47,** 2164-2165.

Sherman, S. L., Takaesu, N., Freeman, S. B., Grantham, M., Phillips, C., Blackston, R. D., Jacobs, P. A., Cockwell, A. E., Freeman, V. and Uchida, I. (1991). Trisomy 21: association between reduced recombination and nondisjunction. *Am.J.Hum.Genet.* **49,** 608-620.

Shiraishi, M., Oates, A. J. and Sekiya, T. (2002). An overview of the analysis of DNA methylation in mammalian genomes. *Biol.Chem.* **383,** 893-906.

Sitar, G., Brambati, B., Baldi, M., Montanari, L., Vincitorio, M., Tului, L., Forabosco, A. and Ascari, E. (2005). The use of non-physiological conditions to isolate fetal cells from maternal blood. *Exp.Cell Res.* **302,** 153-161.

Smets, E. M., Visser, A., Go, A. T., van Vugt, J. M. and Oudejans, C. B. (2006). Novel biomarkers in preeclampsia. *Clin.Chim.Acta.* **364,** 22-32.

Smid, M., Galbiati, S., Lojacono, A., Valsecchi, L., Platto, C., Cavoretto, P., Calza, S., Ferrari, A., Ferrari, M. and Cremonesi, L. (2006). Correlation of fetal DNA levels in maternal plasma with Doppler status in pathological pregnancies. *Prenat.Diagn.* **26,** 785-790.

Smid, M., Galbiati, S., Vassallo, A., Gambini, D., Ferrari, A., Viora, E., Pagliano, M., Restagno, G., Ferrari, M. and Cremonesi, L. (2003). No evidence of fetal DNA persistence in maternal plasma after pregnancy. *Hum.Genet.* **112,** 617-618.

Snijders, R. J., Sundberg, K., Holzgreve, W., Henry, G. and Nicolaides, K. H. (1999). Maternal age- and gestation-specific risk for trisomy 21. *Ultrasound Obstet.Gynecol.* **13,** 167-170.

Soares, S. R., Templado, C., Blanco, J., Egozcue, J. and Vidal, F. (2001). Numerical chromosome abnormalities in the spermatozoa of the fathers of children with trisomy 21 of paternal origin: generalised tendency to meiotic non-disjunction. *Hum.Genet.* **108,** 134-139.

Spencer, K., Cicero, S., Atzei, A., Otigbah, C. and Nicolaides, K. H. (2005). The influence of maternal insulin-dependent diabetes on fetal nuchal translucency thickness and first-trimester maternal serum biochemical markers of aneuploidy. *Prenat.Diagn.* **25,** 927-929.

Steuerwald, N., Cohen, J., Herrera, R. J., Sandalinas, M. and Brenner, C. A. (2001). Association between spindle assembly checkpoint expression and maternal age in human oocytes. *Mol.Hum.Reprod.* **7,** 49-55.

Stroun, M., Anker, P., Lyautey, J., Lederrey, C. and Maurice, P. A. (1987). Isolation and characterization of DNA from the plasma of cancer patients. *Eur.J.Cancer Clin.Oncol.* **23,** 707-712.

Stroun, M., Lyautey, J., Lederrey, C., Mulcahy, H. E. and Anker, P. (2001). Alu repeat sequences are present in increased proportions compared to a unique gene in plasma/serum DNA: evidence for a preferential release from viable cells? *Ann.N.Y.Acad.Sci.* **945,** 258-264.

Tischoff, I. and Tannapfe, A. (2008). DNA methylation in hepatocellular carcinoma. *World J.Gastroenterol.* **14,** 1741-1748.

Tong, Y. K., Ding, C., Chiu, R. W., Gerovassili, A., Chim, S. S., Leung, T. Y., Leung, T. N., Lau, T. K., Nicolaides, K. H. and Lo, Y. M. (2006). Noninvasive prenatal detection of fetal trisomy 18 by epigenetic allelic ratio analysis in maternal plasma: Theoretical and empirical considerations. *Clin.Chem.* **52,** 2194-2202.

Tremblay, K. D., Saam, J. R., Ingram, R. S., Tilghman, S. M. and Bartolomei, M. S. (1995). A paternal-specific methylation imprint marks the alleles of the mouse H19 gene. *Nat.Genet.* **9,** 407-413.

Tsui, D. W., Chan, K. C., Chim, S. S., Chan, L. W., Leung, T. Y., Lau, T. K., Lo, Y. M. and Chiu, R. W. (2007). Quantitative aberrations of hypermethylated RASSF1A gene sequences in maternal plasma in pre-eclampsia. *Prenat.Diagn.* **27,** 1212-1218.

Tsui, N. B., Chim, S. S., Chiu, R. W., Lau, T. K., Ng, E. K., Leung, T. N., Tong, Y. K., Chan, K. C. and Lo, Y. M. (2004). Systematic micro-array based identification of placental mRNA in maternal plasma: towards non-invasive prenatal gene expression profiling. *J.Med.Genet.* **41,** 461-467.

Tsui, N. B., Ng, E. K. and Lo, Y. M. (2002). Stability of endogenous and added RNA in blood specimens, serum, and plasma. *Clin.Chem.* **48,** 1647-1653.

Van der Schoot, C. E., Soussan, A. A., Koelewijn, J., Bonsel, G., Paget-Christiaens, L. G. and de Haas, M. (2006). Non-invasive antenatal RHD typing. *Transfus.Clin.Biol.* **13,** 53-57.

Vodicka, R., Vrtel, R., Dusek, L., Singh, A. R., Krizova, K., Svacinova, V., Horinova, V., Dostal, J., Oborna, I., Brezinova, J., Sobek, A. and Santavy, J. (2007). TSPY gene copy number as a potential new risk factor for male infertility. *Reprod.Biomed.Online.* **14,** 579-587.

Wachi, T. and Kitagawa, M. (2004). Studies on preliminary concentration methods for recovery of fetal nucleated red blood cells in maternal blood. *Congenit Anom (Kyoto).* **44,** 196-203.

Wakuda, K. and Yoshida, Y. (1992). DNA ploidy and proliferative characteristics of human trophoblasts. *Acta Obstet.Gynecol.Scand.* **71,** 12-16.

Wald, N. J., Cuckle, H. S., Densem, J. W., Nanchahal, K., Royston, P., Chard, T., Haddow, J. E., Knight, G. J., Palomaki, G. E. and Canick, J. A. (1988). Maternal serum screening for Down's syndrome in early pregnancy. *BMJ.* **297,** 883-887.

Wapner, R. J. (2005). Invasive prenatal diagnostic techniques. *Semin.Perinatol.* **29,** 401-404.

Weber, M., Davies, J. J., Wittig, D., Oakeley, E. J., Haase, M., Lam, W. L. and Schubeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat.Genet.* **37,** 853-862.

Weier, J. F., Weier, H. U., Jung, C. J., Gormley, M., Zhou, Y., Chu, L. W., Genbacev, O., Wright, A. A. and Fisher, S. J. (2005). Human cytotrophoblasts acquire aneuploidies as they differentiate to an invasive phenotype. *Dev.Biol.* **279,** 420-432.

Williams, J.,3rd, Medearis, A. L., Chu, W. H., Kovacs, G. D. and Kaback, M. M. (1987). Maternal cell contamination in cultured chorionic villi: comparison of chromosome Q-polymorphisms derived from villi, fetal skin, and maternal lymphocytes. *Prenat.Diagn.* **7,** 315-322.

Wolstenholme, J. (1996). Confined placental mosaicism for trisomies 2, 3, 7, 8, 9, 16, and 22: their incidence, likely origins, and mechanisms for cell lineage compartmentalization. *Prenat.Diagn.* **16,** 511-524.

Wong, B. C., Chiu, R. W., Tsui, N. B., Chan, K. C., Chan, L. W., Lau, T. K., Leung, T. N. and Lo, Y. M. (2005). Circulating placental RNA in maternal plasma is associated with a preponderance of 5' mRNA fragments: implications for noninvasive prenatal diagnosis and monitoring. *Clin.Chem.* **51,** 1786-1795.

Wright, D., Bradbury, I., Cuckle, H., Gardosi, J., Tonks, A., Standing, S. and Benn, P. (2006). Three-stage contingent screening for Down syndrome. *Prenat.Diagn.* **26,** 528-534.

Wu, J., Issa, J. P., Herman, J., Bassett, D. E.,Jr, Nelkin, B. D. and Baylin, S. B. (1993). Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc.Natl.Acad.Sci.U.S.A.* **90,** 8891-8895.

Xiong, Z. and Laird, P. W. (1997). COBRA: a sensitive and quantitative DNA methylation assay. *Nucleic Acids Res.* **25,** 2532-2534.

Yamada, Y., Watanabe, H., Miura, F., Soejima, H., Uchiyama, M., Iwasaka, T., Mukai, T., Sakaki, Y. and Ito, T. (2004). A comprehensive analysis of allelic methylation status of CpG islands on human chromosome 21q. *Genome Res.* **14,** 247-266.

Yoon, P. W., Freeman, S. B., Sherman, S. L., Taft, L. F., Gu, Y., Pettay, D., Flanders, W. D., Khoury, M. J. and Hassold, T. J. (1996). Advanced maternal age and the risk of Down syndrome characterized by the meiotic stage of chromosomal error: a population-based study. *Am.J.Hum.Genet.* **58,** 628-633.

Yoshikawa, T., Imazu, S., Gao, J. Q., Hayashi, K., Tsuda, Y., Okada, N., Tsutsumi, Y., Akashi, M., Mayumi, T. and Nakagawa, S. (2006). Non-methylated CpG motif packaged into fusogenic liposomes enhance antigen-specific immunity in mice. *Biol.Pharm.Bull.* **29,** 105-109.

Zipursky, A., Poon, A. and Doyle, J. (1992). Leukemia in Down syndrome: a review. *Pediatr.Hematol.Oncol.* **9,** 139-149.

Zwaan, M. C., Reinhardt, D., Hitzler, J. and Vyas, P. (2008). Acute leukemias in children with Down syndrome. *Pediatr.Clin.North Am.* **55,** 53-70, x.