

UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
INSTITUTE OF MOLECULAR AND CELL BIOLOGY

Hindrekk Teder

Evaluation of the possibility to detect fetal chromosome
trisomies based on a defined set of single nucleotide
polymorphisms for non-invasive prenatal testing

Master's Thesis (30 ECTS)

Gene Technology

Supervisors: Kaarel Krjutškov, PhD

Prof. Ants Kurg, PhD

TARTU 2016

Evaluation of the possibility to detect fetal chromosome trisomies based on a defined set of single nucleotide polymorphisms for non-invasive prenatal testing

Abstract

Non-invasive prenatal testing (NIPT) of fetal aneuploidy using cell-free fetal (cffDNA) from mother's blood sample has shown to be an accurate and reliable screening tool. The current NIPT protocols are based on targeted or whole genome sequencing, which demand resource-intensive bioinformatical capacity. The complexity of current NIPT technology is trustworthy but the comprehensive adaption of the application is still limited due to the high cost.

Mother- and fetus-specific genotypes, according to the nature of cell-free DNA (cfDNA) during pregnancy, were simulated and used in further analysis. Simulations and theoretical calculations demonstrate the characteristic patterns of allelic ratios in case of normal number of chromosomes or trisomy where extra chromosome is inherited from mother or father. Here described analytical approach managed to identify fetal trisomy by comparing the allelic ratios of the risk chromosome with the expected allelic ratios using the *t*-test and hidden Markov model (HMM) analysis.

An average, 3/4 of all highly polymorphic single nucleotide polymorphisms (SNPs) can be used in analysis based on comparison of the allelic ratios. As a result, at least 300 highly polymorphic SNPs over risk and reference chromosomes are needed to detect fetal trisomy using *t*-test alone. In addition, the HMM analysis can independently detect fetal trisomy and have the ability to distinguish the parental origin of trisomy.

Based on the simulated data, the existence and the origin of fetal trisomy is theoretically detectable using a novel and highly quantitative SNP-based approach that is under development by our research group. However, further testing has to be carried out with the real data to confirm the theory.

Keywords

prenatal screening, non-invasive prenatal testing, single nucleotide polymorphism, hidden Markov model, cell-free DNA, cell-free fetal DNA, allelic ratio, aneuploidy, trisomy, Down syndrome

CERCS

B110 Bioinformatics, medical informatics, biomathematics, biometrics

Mitte-invasiivse prenataalse skriiningu taustauuring – kas loote kromosoomi trisoomiaid on võimalik määrata ühenukleotiidsete polümorfismide valimiga?

Lühikokkuvõte

Loote mitte-invasiivne prenataalne testimine (NIPT) ema vereproovist on osutunud täpseks ja usaldusväärseks skriiningu tööriistaks. Kasutuses olevad NIPT protokollid põhinevad genoomi osalisel või täielikul sekveneerimisel, mis nõuab ressursimahukat bioinformaatilist võimekust. NIPT tehnoloogiad on küll usaldusväärsed, kuid nende ulatuslik kasutuselevõtt on piiratud eelkõige kõrge hinna tõttu.

Töö käigus simuleeriti ema- ja lootespetsiifilisi genotüübi andmeid, vastavalt rakuvaba DNA olemusele raseda veres, kasutati edaspidises analüüsis. Simulatsioonid ja teoreetilised kalkulatsioonid näitavad iseloomulikke alleelsete suhteid nii normaalse kui ka ema- või isapoolse trisoomiaga loote korral. Loote riskikromosoomide alleelsete suhete võrdlemisel oodatud alleelsete suhetega

Kirjeldatud analüütiline lähenemine põhineb riskikromosoomide alleelsete suhete ja oodatavate alleelsete suhete võrdlusel ning võimaldab tuvastada loote trisoomiat. Statistiliseks analüüsiks kasutati t -testi ja varjatud Markovi mudelit (HMM). Analüüsiks on võimalik kasutada keskmiselt 3/4 kõikidest kõrge polümorfisusega ühenukleotiidsetest polümorfismidest (SNP). T -testi korral läheb loote trisoomia tuvastamiseks vaja ligikaudu 300 kõrge polümorfisusega SNP-i, mis paikneksid nii riski- kui ka referentskromosoomides. Lisaks võimaldab HMM analüüs t -testust sõltumatult detekteerida loote trisoomiat ning selle vanemlikku päritolu.

Uudset ning kvantitatiivset SNP-põhist lähenemist kasutades on simuleeritud andmete põhjal teoreetiliselt võimalik tuvastada loote trisoomia olemasolu ja pärinevust. Laboratoorne meetodika on meie uurimisgrupi poolt hetkel väljatöötamisel ja läbinud edukalt esimese testfaasi. Lõpliku kindluse saavutamiseks on vaja analüüsi korrata reaalsetel katseandmetel.

Märksõnad

prenataalne skriining, mitte-invasiivne prenataalne testimine, ühenukleotiidne polümorfism, varjatud Markovi mudel, rakuvaba DNA, loote rakuvaba DNA, alleelne suhe, aneuploidia, trisoomia, Down'i sündroom

CERCS

B110 Bioinformaatika, meditsiiniinformaatika, biomatemaatika, biomeetrika

CONTENTS

Evaluation of the possibility to detect fetal chromosome trisomies based on a defined set of single nucleotide polymorphisms for non-invasive prenatal testing	2
Abstract.....	2
Keywords.....	2
CERCS	2
Mitte-invasiivse prenataalse skriiningu taustauuring – kas loote kromosoomi trisoomiaid on võimalik määrata ühenukleotiidsete polümorfismide valimiga?.....	3
Lühikokkuvõte.....	3
Märksõnad	3
CERCS	3
CONTENTS	4
ABBREVIATIONS	6
INTRODUCTION	7
1 LITERATURE REVIEW	8
1.1 Aneuploidy.....	8
1.1.1 Non-disjunction	8
1.1.2 Anaphase lag.....	10
1.1.3 Mosaicism.....	11
1.1.4 Parental origin	11
1.1.5 Maternal age effect	12
1.1.6 Trisomy.....	13
1.2 Current methods of prenatal diagnosis and screening	14
1.3 Non-invasive cell-free DNA based screening.....	15
1.3.1 Cell-free fetal DNA	15
1.3.2 Detection of aneuploidies	16
1.4 Computational methods for aneuploidy detection	17
1.4.1 Next-generation Aneuploidy Test Using SNPs	17

1.4.2	Fetal Fraction Optimized Risk of Trisomy Evaluation	17
2	STUDY	19
2.1	Aim of the study	19
2.2	Methods	19
2.2.1	Data simulation.....	19
2.2.2	Allelic distribution and informative SNPs.....	20
2.2.3	Allelic ratio	21
2.2.4	Fetal-to-maternal cell ratio	22
2.2.5	Analytical approach.....	23
2.3	Results.....	24
2.3.1	Detection of trisomy with <i>t</i> -test.....	26
2.3.2	Detection of trisomy with hidden Markov model	27
2.4	Discussion.....	29
	SUMMARY	32
	KOKKUVÕTE.....	33
	ACKNOWLEDGEMENTS	34
	REFERENCES	35
	NON-EXCLUSIVE LICENCE	40

ABBREVIATIONS

cfDNA – cell-free DNA

cffDNA – cell-free fetal DNA

CPM – confined placental mosaicism

DANSR™ – Digital Analysis of Selected Regions

DS – Down syndrome

FMCR – fetal-to-maternal cell ratio

FORTE™ – Fetal Fraction Optimized Risk of Trisomy Evaluation

HMM – hidden Markov model

MI – meiosis I

MII – meiosis II

NATUS™ – Next-generation Aneuploidy Test Using SNPs

NGS – next generation sequencing

NIPT – non-invasive prenatal testing

PMAP – pointwise maximal a posterior

PZM – post-zygotic (mitotic)

SNP – single nucleotide polymorphism

UPD – uniparental disomy

INTRODUCTION

Changes in fetal chromosome copy number constitute a large part of genetic disorders affecting more than half of early embryos (Taylor, Gitlin et al. 2014; McCoy, Demko et al. 2015). Prenatal screening and diagnostics aim to detect these problems in time, which may be essential for the future parents and the child.

Fetal chromosomal abnormalities can be detected either by invasive diagnostic tests or by non-invasive screening tests. The latter are considered less harmful for developing fetus, but in terms of detection rates are outperformed by invasive tests. Searching for the methods that can provide superior accuracy without the cost of safety has driven the field of prenatal genetic testing for years.

Recent developments in non-invasive prenatal testing have shown promising results and it is becoming a widely implemented alternative to conventional serum marker based screening. Several commercial platforms are currently available for common autosomal and sex chromosome aneuploidies detection. Unfortunately, the adaption of NIPT is limited by the high cost, which is due to the fact that current NIPT methods mostly use next-generation sequencing (NGS). To fulfill the gap in accurate and affordable methods, we evaluated the possibility to detect fetal chromosome trisomies using a highly cost-effective SNP-based approach.

This study covers the main mechanisms that lead to fetal aneuploidy, especially to trisomy, and proposes a theoretical way to calculate expected allelic ratios, which can be used to estimate fetal trisomy. Using simulated data, that mimics the cfDNA of a pregnant patient's blood, we estimate the underlying fetal condition.

The study was supported by Competence Centre on Health Technologies.

1 LITERATURE REVIEW

1.1 Aneuploidy

The search for the origin of chromosomal aneuploidy became possible due to the cytogenetic analyses of human oocytes in early 1970s (Pellestor et al. 2005). Studies have established that aneuploidy is the leading known cause of congenital birth defects. At least 10% of all clinically diagnosed pregnancies are trisomic or monosomic. It is assumed that many aneuploid conceptions are eliminated during the earliest stages of pregnancy (Nagaoka et al. 2012).

Aneuploidies are mainly caused by the faults made during oocyte development (Nagaoka et al. 2012). Two main differences between female and male gametogenesis, that have an effect on the errors made in meiosis, are timing and number of cells produced from one parent cell (Hassold et al. 2007).

The majority of errors are caused by different non-disjunctional mechanisms. Some errors are associated with failure to crossover, others with crossovers that occur too close or too far from the centromere. Also there are mechanisms that have nothing to do with recombination, but are attributable to abnormalities in other meiotic processes – loss of sister chromatid cohesion or defects in spindle assembly/disassembly, which may lead to mosaicism (Hassold et al. 2007; Sherman et al. 2006). The main mechanisms of chromosomal missegregation errors that result in aneuploidy or mosaicism – non-disjunction and anaphase lag – are described in detail in the next chapters.

1.1.1 Non-disjunction

Chromosomal non-disjunction is a state, where chromosomes in meiosis I (MI) or sister chromatids in meiosis II (MII) fail to separate equally into the daughter cells. This results in one cell having two chromosomes or chromatids that remained connected and a cell that has none (Lamb et al. 2005).

Up to five different patterns of abnormal meiotic segregation can be distinguished as described in Figure 1. One possible mechanism is the achiasmate non-disjunction, in which homologues fail to pair and/or recombine and segregate to the same pole due to the absence of chiasma (Figure 1B). The second possible mechanism involves premature separation of sister chromatids (PSSC), where sister chromatids segregate from a univalent in MI (Figure 1C). PSSC may be caused by premature biorientation of sister chromatid centromeres of univalents during MI to avoid spindle assembly checkpoint (Kouznetsova et al. 2007). The third mechanism is the “true” non-disjunction at MI, which probably originates from the joint segregation of a linear bivalent showing a single distal chiasma (Figure 1D). The presence of a

proximal chiasma can lead to a premature loss of centromeric cohesion between chromatids and their subsequent segregation at random (Rockmill et al. 2006). This has been observed in MII oocytes and is called a balanced PSSC (Figure 1E). The true non-disjunction at MII implies the segregation of the two sister chromatids to the same pole (Figure 1F). It is thought that this happens because of the inability to lose cohesion between centromeres of sister chromatids or a bad orientation in the metaphase plate (Templado et al. 2013).

Studies on human oocytes and sperms have repeatedly shown achiasmate non-disjunction and PSSC as the two main mechanisms generating aneuploidy in MI (Fragouli et al. 2011; Uroz & Templado 2012). The errors in MII usually result from the failure of sister chromatid separation (Hassold & Hunt 2001). Most of the remaining bivalent errors of MI had exchanges occurring near the telomeres. In contrast, errors of MII seemed to cluster at the pericentromeric region of the chromosome. It is suggested that since susceptibility to non-disjunction is associated with the distance between the centromere and the nearest exchange, errors of MII are not independent events, as widely believed, but errors of MI resolved in a later stage (Lamb 1997).

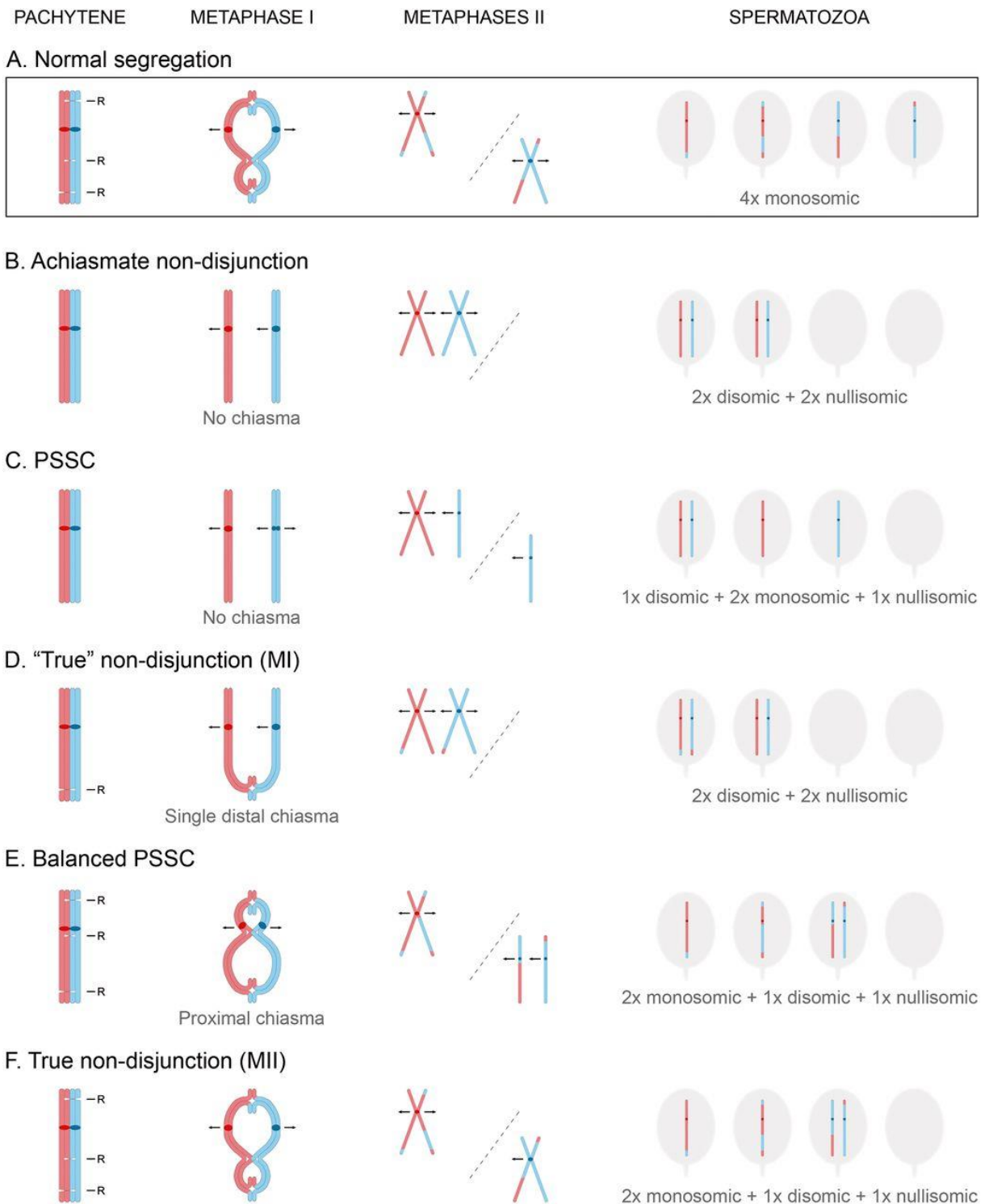


Figure 1. Meiotic non-disjunction on the example of spermatozoa. (A) Normal chromosome segregation during meiosis. (B) Achiasmate non-disjunction involves the random segregation of two homologous chromosomes at MI. (C) Premature separation of sister chromatids (PSSC) is caused by the sister chromatids being segregated from one another in MI. (D) "True" non-disjunction at MI means that homologues travel to the same pole. (E) Balanced PSSC is characterized by the random segregation of two separated sister chromatids at MII. (F) True non-disjunction at MII originates by the joint segregation of the sister chromatids that maintain their cohesion. R – recombination. Adjusted from Templado et al., 2013.

1.1.2 Anaphase lag

Chromosome non-disjunction and anaphase lagging are two different mechanisms that contribute to mono- and trisomies. Anaphase lagging is defined as the failure of a single

chromosome or chromatid to be incorporated into the nucleus despite the attachment to the spindle or caused by the failure to attach to the spindle. This results in a monosomy of the chromosome in one cell and a disomy of corresponding chromosome in the other cell (Coonen et al. 2004).

Aside from aneuploidies, anaphase lag can lead to a uniparental disomy (UPD) – a case where both copies of a chromosome originate from the same parent. This may occasionally occur when a trisomic conceptus loses one extra chromosome to regain normal chromosomal constitution by mitotic non-disjunction or anaphase lag. In one third of cases, the remaining homologues originate only from one parent (Nicolaidis & Petersen 1998).

1.1.3 Mosaicism

Another set of whole chromosome aneuploidies affecting embryos, which result from chromosome missegregation during the first mitotic cell divisions in cleavage stage, leads to mosaicism – a condition in which a person has two or more genetically distinct cell lines that originated from a single zygote (Delhanty et al. 1993).

Confined placental mosaicism (CPM) is a subtype of mosaicism, in which only fetal and placental tissues contain differences in chromosome number. CPM serves as a challenge for NIPT screening as a source of false positive results, as majority of the cfDNA originates from placental tissue (Bayindir et al. 2015). CPM appears in approximately 6% of all pregnancies. The prevalence of chromosomal mosaicism in placental tissue in case of an euploid fetus is thought to be 1–2% (Taylor et al. 2014). Out of all people with any kind of DS, mosaicism for trisomy 21 is estimated to be present in 1.3–5% (Papavassiliou et al. 2015).

1.1.4 Parental origin

There is a consensus that possibly more than 90% of aneuploidies originate from maternal meiotic errors (Robinson et al. 1995; Hassold & Hunt 2001). Several potential causes have been proposed for the considerably higher proportion of maternal errors. Most of the examples and evidence is based on trisomy 21 as the most frequent and studied autosomal trisomy. These include exceptionally long MI during oogenesis, that spans from fetal development up to menopause (Hassold et al. 2007), and higher stringency in the elimination of trisomy 21 cells during fetal testicular than ovarian development (Iwarsson et al. 2015). Another theory proposes an explanation through trisomy 21 mosaicism in fetal ovaries, which means that female embryos already have a trisomic cell population in their ovaries (Hultén et al. 2010).

Only less than 10% aneuploidy is assumed to originate from the paternal line, and therefore significantly less studies have been dedicated to the understanding of paternal origin (Hassold

& Hunt 2001). Male gametes are continuously produced from male germ cells that are mitotically dividing in the sexually mature adult before entering meiosis. Male gametes are therefore under much less temporal strain, and sperm production is maintained throughout lifetime. Most aneuploidies in male gametes concern the sex chromosomes because during male meiosis, XY chromosomes that harbor only a limited region of homology have to pair and segregate, in contrary to female meiosis where the XX chromosome pair does not create additional challenges (Hassold & Hunt 2001). As in maternal origin, the main mechanism causing meiotic segregation errors in human sperm is non-disjunction, followed by anaphase lag. Achiasmata non-disjunction and PSSC are found to be main contributors to non-disjunction in MI in work on infertile patient and fertile men (Uroz & Templado 2012).

Despite the unbalance towards maternal contribution, parental origin is found to be chromosome-dependent (Table 1). According to the data, trisomies of acrocentric chromosomes 13, 15, 21 and 22 originate from maternal MI, while MI and MII equally contribute to trisomy 14 (Hall et al. 2007). Distinguishing parental origin can be of importance in cases of pregnancy loss or IVF treatment. It has been reported that men with female partners experiencing recurrent pregnancy loss have increased sperm aneuploidy compared with controls (Ramasamy et al. 2015).

Table 1. Origin of trisomies. N – number of cases; MI – meiosis I; MII – meiosis II; PZM – post-zygotic mitotic. Adapted from Hassold et al. 2007.

Trisomy	N	Maternal		Paternal		PZM (%)
		MI (%)	MI (%)	MII (%)	MII (%)	
21	782	69.6	23.6	1.8	0.0	2.7
18	150	33.3	58.7	0.0	0.0	8.0
13	74	56.6	33.9	2.7	5.4	1.4

1.1.5 Maternal age effect

Studies of trisomies have demonstrated that the risk and incidence of aneuploidy is associated with increasing maternal age. The long time interval between meiotic arrest in the fetus and each ovulation cycle in the adult enable maternal age to affect aneuploidy incidence (Hassold & Hunt 2001). Theories of age-associated aneuploidies include recombination errors in early meiosis, deterioration of sister chromatid connection with age, a defective spindle assembly checkpoint, and maternal “genetic age” or telomere length (Chiang et al. 2012; Ghosh et al. 2010). The impact of maternal age is substantial – 2% of all pregnancies among women under the age of 25 are trisomic, but the value approaches to 35% among women over 40 years of

age. There is no known influence of race, geography, or socio-economic status on maternal-age-specific rates of trisomy (Hassold & Hunt 2001).

Since the vast majority of aneuploidies, which originating from maternal errors, are in fact meiotic, therefore are age-dependent (Chiang et al. 2012). This does not apply to all cases, as non-disjunction is a complex mechanism influenced by factors and events both age-dependent and independent (Ghosh et al. 2010). The two-step mechanism provides an explanation combining these two. First, a fragile chiasmatic configuration is formed in the fetal oocyte, which is the age-independent step. Secondly, this event is followed by abnormal processing of the vulnerable bivalent at MI, that depends on maternal age (Muller et al. 2000).

Maternal age is the most significant risk factor for non-disjunction resulting in DS (Sherman et al. 2006). Incidence of DS varies from about 1 in 1,200–1,600 to 1 in 20–30 in maternal age range 20–45 years, respectively (Cheslock et al. 2005). When evident relation between maternal origin of aneuploidy and advanced maternal age has been established, no conclusive connection has been demonstrated between paternal age and frequency of disomic sperm (Templado et al. 2013).

1.1.6 Trisomy

The most common aneuploidies in humans are trisomies, which represent approximately 0.3% of all live births. Trisomies are characterized by the presence of one additional chromosome. Usually trisomies do not appear to be compatible with life, representing about 35% of spontaneous abortions (Hassold & Hunt 2001). Analyses show that regardless of the chromosome, most trisomies originate during oogenesis – therefore are maternal. This observation makes sense when considering that human oocytes can be arrested in prophase I for several decades. Usually errors of maternal MI are more common than errors of maternal MII and the proportion of cases of maternal origin increases with maternal age (Hassold & Hunt 2001). However, chromosome-specific differences can occur as described in Table 1 (Hassold et al. 2007).

Trisomy 16 is the most common trisomy, occurring more than 1% of all pregnancies (Hassold et al. 1995). However, this condition results in spontaneous miscarriage in the first trimester, only fetuses with mosaic form may survive (Langlois et al. 2006). The most common autosomal trisomies that survive to birth are (Parker et al. 2010):

- Trisomy 21 (Down syndrome) – 1 in 700
- Trisomy 18 (Edwards syndrome) – 1 in 4,000
- Trisomy 13 (Patau syndrome) – 1 in 8,000

Individuals with DS are associated with physical growth delays, characteristic facial features, and mild to moderate intellectual disability, but patients routinely live to adulthood (Weijerman & de Winter 2010). In contrast, other autosomal trisomies have more severe effects and are rarely given birth alive. In fact, the only other autosomal trisomies, that are detected in any appreciable numbers, involve chromosomes 18 and 13, but newborns rarely survive beyond the first few months (Hassold & Hunt 2001).

1.2 Current methods of prenatal diagnosis and screening

Genetic prenatal diagnosis for fetal aneuploidies, such as trisomy 21, 18 and 13, has been an important part of prenatal medicine over 40 years now. Definitive prenatal diagnosis can be assessed only by sampling the fetal material, obtained through invasive testing, which is associated with a chance of fetal miscarriage (Chitayat et al. 2011). Therefore, to set a limit to invasive procedures, more convenient screening methods are used.

There are several screening methods in use to assess the risk of chromosomal abnormalities, which vary in terms of gestation time, cost and accuracy. A list of different methods of prenatal diagnosis and screening are shown in Table 2. Once the high risk is identified, an invasive test, such as chorionic villus sampling, amniocentesis or cordocentesis, is performed to obtain the fetal genetic material for a definitive diagnosis.

Table 2. Methods of prenatal diagnosis and screening. Sorted by invasiveness and time. Adopted and modified from Kotsopoulou et al. 2015.

Invasiveness	Method	Time
Non-invasive	Preimplantation genetic diagnosis	Before implantation
	Fetal cells in maternal blood	First trimester
	Cell-free fetal DNA in maternal blood	
	Ultrasound detection	First or second trimester
	Fetal heartbeat	
	Maternal serum screening	
	Embryoscopy and fetoscopy	After 20 weeks
Invasive	Chorionic villus sampling	After 10 weeks
	Amniocentesis	After 15 weeks
	Cordocentesis	After 20 weeks

One of the most widely used non-invasive screening method of fetal chromosomal pathologies is maternal serum screening with or without ultrasound scan. The combination of the two procedures is called the First Trimester Combined Test, which detects 78–90% of fetal aneuploidies and has a false positive rate around 5%. However, 95% of pregnant women, who result positive in the combined screening and follow an invasive procedure, are diagnosed with

no chromosomal pathology and therefore suffer an unnecessary stress and a risk of miscarriage (Russo & Blakemore 2014). Therefore, a more reliable and convenient method for prenatal diagnosis with a smaller risk of fetal loss or a more specific screening method with a lower rate of false positive results has long been searched for (Wright & Chitty 2009).

1.3 Non-invasive cell-free DNA based screening

Placenta has a two-way permeability – from fetus to mother and *vice versa*, having heavy traffic between the fetal and maternal sections (Lo et al. 1996). During pregnancy some cells of placenta undergo an apoptosis – a programmed cell death – which result in small fragments of DNA appearing in the maternal circulation. These fragments form cfDNA, part of which originates from fetus and is referred to as cffDNA (Alberry et al. 2007).

For screening analysis, cfDNA is extracted from maternal plasma and used to assess the fetal genetic material (Lo et al. 1999). In addition to conventional screening methods, which can generally detect only the most common autosomal trisomies, NIPT has been used in determination of fetal sex, fetal rhesus D genotyping, prediction of gestational complications and even detection of single gene disorders (Ayse Kirbas, Korkut Daglar 2016).

1.3.1 Cell-free fetal DNA

The cffDNA was first described in 1997 and it consists of small DNA pieces of fetal origin, which circulate freely in the mother's blood system until reaching to liver or kidneys for final utilization (Lo et al. 1997). These fragments are not within a cell, so they are unstable and have a short half-life up to 30 minutes (Lo et al. 1999). Studies of cfDNA have shown that the trophoblastic cells, which form the outer layer of a blastocyst and develop into a large part of the placenta, are the major source of cffDNA (Alberry et al. 2007). Only 3–13% of cfDNA in maternal circulation is fetal origin, the rest belongs to the mother (Wang et al. 2013).

CffDNA can be detected in maternal blood as early as 7 weeks of gestation (Lo et al. 1998). At 10 weeks the median cffDNA in total cfDNA is approximately 10% and the concentration increases 0.1% per week as the pregnancy advances (Wang et al. 2013). The percentage of cffDNA in total cfDNA, also known as fetal fraction, is similar to those in euploid pregnancies in trisomy 13 and 21, but noticeable depletion has been shown in trisomy 18. This decrease in fetal fraction has been linked to the smallness of placenta. Fetal fraction also decreases as maternal weight increases. Therefore, a cfDNA based aneuploidy screening may not be the best method for obese women (Palomaki et al. 2011).

1.3.2 Detection of aneuploidies

The cfDNA from maternal blood has been used in detecting fetal aneuploidies such as trisomies and monosomies of autosomal and sex chromosomes (Fan et al. 2008; Samango-Sprouse et al. 2013). A meta-analysis (Table 3) support that the cfDNA analysis of maternal blood can be used safely for fetal trisomy screening in singleton pregnancies. Trisomy 21 screening with NIPT has been shown to be superior compared to the combination of conventional methods – mother’s age, ultrasound results and biochemical analysis of the maternal serum. However, the cfDNA screening test performance for trisomy 18 and 13 was not as sensitive compared to the trisomy 21 (Gil et al. 2015).

Table 3. Summary of NIPT studies of common autosomal trisomies. Studies – number of studies; Cases – number of cases; DR – pooled weighted detection rate; CI – confidence interval; FPR – pooled weighted false positive rate. Adapted and modified from Gil et al. 2015.

Trisomy	Studies	Cases		DR (% (95% CI))	FPR (% (95% CI))
		Trisomy	Non-trisomy		
21	24	1,051	21,608	99.2 (98.5–99.6)	0.09 (0.05–0.14)
18	21	389	21,306	96.3 (94.3–97.9)	0.13 (0.07–0.20)
13	18	139	18,059	91.0 (85.0–95.6)	0.13 (0.05–0.26)

There are several different approaches to detect fetal aneuploidy from cfDNA. These methods are commonly referred as NIPT. The first NGS based approaches employed genome-wide random sequencing, which enables to assess the proportion of each chromosome by counting the uniquely aligned sequences on each chromosome. If the proportion of a chromosome is increased, then it is considered that the fetus has trisomy (Fan et al. 2008). In scientific communities, these methods are called as the first-generation NIPT.

The latter, next-generation NIPT approaches mostly use targeted sequencing, which focus on individual chromosomes or pre-defined regions in genome, requiring less cfDNA for analysis. Targeting involves the use of hybridization-based capture of the genomic regions of interest or the use of highly-multiplexed PCR to amplify SNPs followed by NGS (Liao et al. 2011; Zimmermann et al. 2012). Two of the next-generation NIPT approaches, that are widely used and commercially available in Estonia, are described in more detail in the next chapter (Ustav 2015).

1.4 Computational methods for aneuploidy detection

1.4.1 Next-generation Aneuploidy Test Using SNPs

In 2012, Natera, Inc.¹ in San Carlos, CA released Panorama™ Prenatal Test, which employs Next-generation Aneuploidy Test Using SNPs (NATUS) algorithm. The algorithm considers parental genotypes or, in the absence of a paternal sample, population allele frequencies and crossover frequency data² to calculate the expected allele distributions for 19,488 SNPs and possible fetal genotypes based on recombination sites in the parent chromosomes. It compares predicted allelic distributions to measured allelic distributions by employing a Bayesian-based Maximum Likelihood approach to determine the relative likelihood of chromosomal copy number – monosomy, disomy or trisomy – hypothesis. The likelihoods of each sub-hypothesis, which are based on recombination sites, are summarized to find out the hypothesis with the maximum likelihood (Samango-Sprouse et al. 2013).

NATUS takes into account a number of quality control indicators such as identifying sub-optimal lab or sequencing results, estimating the amount of total starting DNA, determining the fetal fraction and calculating the extent of expected distributions to which the measured cfDNA data has to be fitted (Samango-Sprouse et al. 2013).

1.4.2 Fetal Fraction Optimized Risk of Trisomy Evaluation

In 2012, Ariosa Diagnostics, Inc.³ in San Jose, CA developed a novel assay – Digital Analysis of Selected Regions (DANSR). The assay enables targeted amplification of specific regions under investigation (Sparks et al. 2012). DANSR in combination with Fetal Fraction Optimized Risk of Trisomy Evaluation (FORTE) algorithm, which is designed to account for an individual's age-related risks and the percentage of cfDNA in the sample to provide an individualized risk score for trisomy, are used in Harmony™ Prenatal Test (Stokowski et al. 2015).

DANSR approach is closely related to random whole genome sequencing, where the entire genome is evaluated using approximately 25 million raw reads per subject, which limits sequencing throughput to 4–6 samples per lane. The method includes an initial targeted amplification step in which 384 loci of each chromosome of interest are selectively amplified prior to NGS analysis. This targeted amplification results in an improvement in sequencing efficiency per target chromosome and compared to random genome-wide methods, DANSR uses 10 times less cfDNA and enables aneuploidy detection using approximately 1 million raw

¹ <http://www.natera.com/>

² <https://www.ncbi.nlm.nih.gov/probe/docs/projhapmap/>

³ <http://www.ariosadx.com/>

reads per subject, which allows analysis of 96 subjects per sequencing lane. As with all quantitative methods, the approach is dependent on chromosomes having low amplification variability, which may limit its diagnostic accuracy for some chromosomes (Norwitz & Levy 2013).

Although quantitation of cfDNA for fetal aneuploidy screening to date has mostly relied upon NGS, this approach has been demonstrated to accurately quantify the targeted DANSR products for NIPT using DNA microarrays (Stokowski et al. 2015). Thus, DANSR has an advantage over NGS in terms of sequencing cost and throughput (Juneau et al. 2014).

2 STUDY

2.1 Aim of the study

The aim of the study was to evaluate the possibility to develop a computational data analysis method for a SNP-based NIPT to determine fetal trisomies and their parental origin.

2.2 Methods

2.2.1 Data simulation

Three independent simulations were conducted to study the distribution of allelic ratios of SNPs in the cfDNA using R software⁴. As the cfDNA of a pregnant patient contains maternal and fetal genetic material, I started the simulation with the formation of chromosomes for each individual. Firstly, I created a pair of homologous chromosomes for either parent. For simplification each chromosome was generated as a vector of letters representing alleles of consecutive SNPs. Secondly, as homologous recombination takes place in meiosis, I switched half of the alleles between the copies of parental homologous chromosomes to simulate their offspring's possible fetal genotype. Thirdly, I picked the recombinant chromosomes to form a pair of fetal chromosomes. In case of a normal fetus, two recombinant chromosomes were selected, one from each parent. In case of a trisomy, one recombinant chromosome originated from one parent and two from the other, depending on the parental origin of the trisomy. Lastly, having simulated the maternal and the fetal chromosomes, I mixed them together according to the theoretical allele frequencies (Table 4) in case of 1/10 fetal-to-maternal cell ratio (FMCR), which corresponds to the median fetal fraction (Ashoor et al. 2013).

In case of normal fetus 9/10 of the observed chromosome is maternal and 1/10 fetal. In case of trisomy, fetal fraction of the observed chromosome is theoretically increased by one third, which shifts the overall proportion – 6/7 belongs to the mother and 1/7 to the fetus.

Table 4. Theoretical allele frequencies and chromosomal fractions in cfDNA. The maternal and fetal allele frequencies and chromosomal fractions are calculated for three different fetal genotypes – normal, maternal and paternal trisomy. The calculations are carried out in 1/10 FMCR.

Fetal genotype	Maternal			Fetal		
	Allele		Chromosomal fraction	Allele		Chromosomal fraction
	N	Frequency		N	Frequency	
Normal	2	9/20	9/10	2	1/20	1/10
Maternal trisomy		9/21	6/7	3	1/21	1/7
Paternal trisomy						

⁴ <https://www.r-project.org/>

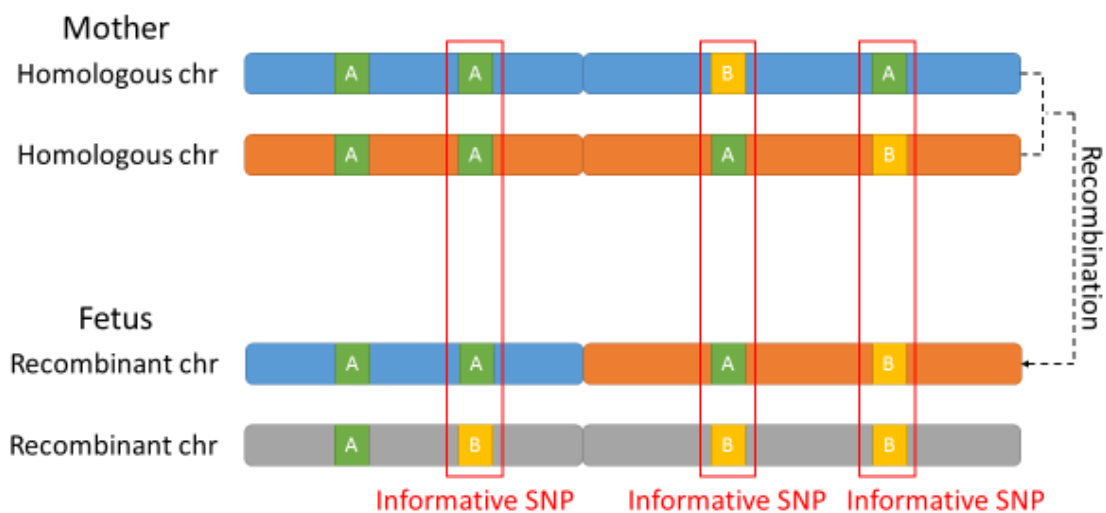
As a result, 440 000 000 reads were generated per simulation to cover 440 000 virtual SNPs, which makes an average of 1 000 reads per SNP.

2.2.2 Allelic distribution and informative SNPs

The simulations incorporate only highly polymorphic SNPs where minor allele frequency is equal to major allele frequency. If both alleles of a SNP are present in cfDNA, regardless whether they derive from the mother or the fetus, the SNP is defined as an informative SNP (Figure 2). In case of equally balanced alleles I expect that 1/2 of SNPs are heterozygous in mother and fetus. As allelic ratio can be calculated even if only one of the individuals is heterozygous, I would expect that 3/4 of all SNPs are heterozygous in case of normal fetus.

In case of maternally inherited trisomy, the expected proportion of heterozygous SNPs remains the same as in case of normal fetus – 3/4. In case of paternally inherited trisomy, the proportion of informative SNPs is higher as fetal alleles also express paternal heterozygosity. The opposite event, SNP being homozygous, occurs only when the mother and father are both homozygous and the alleles match – 1/8 of total SNPs. To get the probability of a SNP being informative in case of paternal trisomy, I subtract 1/8 from all possibilities which gives us 7/8.

A. Normal



B. Maternal trisomy



C. Paternal trisomy

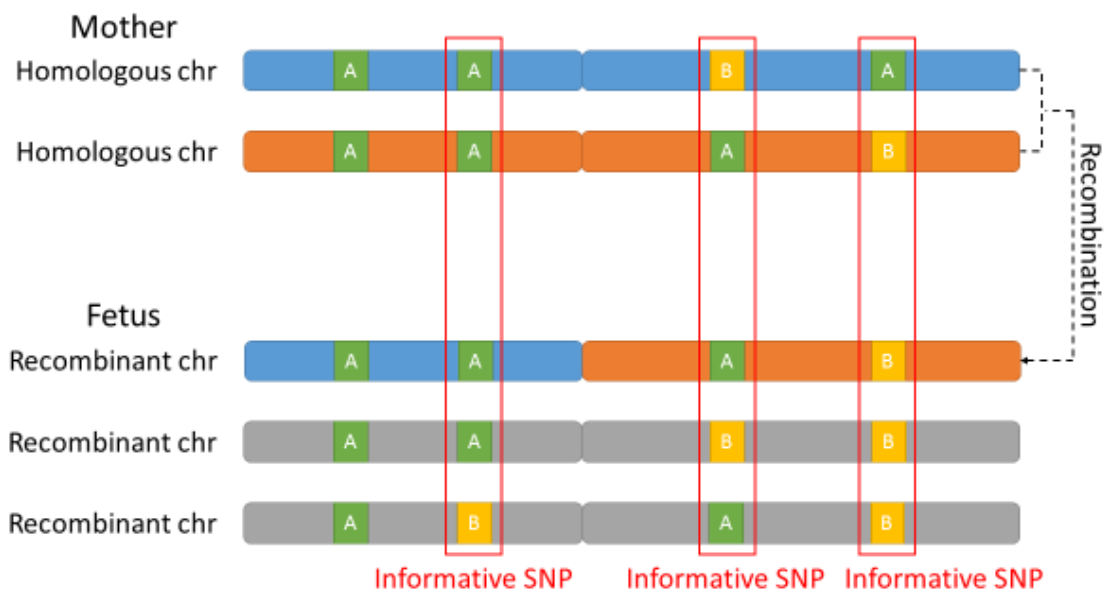


Figure 2. Chromosomal schemes of allelic distribution. (A) Normal fetus. (B) Maternal trisomy. (C) Paternal trisomy. Loci surrounded by red rectangle represent informative SNPs.

2.2.3 Allelic ratio

Using previously simulated data, the number of reads containing major and minor alleles were counted in each locus. Then, considering only the informative SNPs where both alleles were present, I calculated allelic ratios as follows:

$$allelic\ ratio_i = \frac{major\ allele\ frequency_i}{minor\ allele\ frequency_i}$$

where i represents the i^{th} informative SNP for studied sample.

For detection of fetal trisomy, I have to compared the measured allelic ratios with the expected allelic ratios specific to each possible combination of maternal and fetal genotype (Table 5). Before calculating the expected allelic ratios, FMCR must be known in advanced. The calculation of FMCR is described in more detail in the next chapter.

Table 5. Allelic ratio dependency on maternal and fetal zygosity and genotype. Formulas of expected allelic ratios by maternal and fetal zygosity in case of different fetal genotypes – normal, maternal and paternal trisomy. MA – maternal allele; FA – fetal allele; * – theoretically impossible.

Fetal genotype	Maternal zygosity	Fetal zygosity	Major allele	Minor allele	Allelic ratio
Normal	Homozygous	Homozygous	$2MA + 2FA$	-	-
		Heterozygous	$2MA + FA$	FA	$\frac{2MA + FA}{FA}$
Maternal trisomy		Homozygous	$2MA + 3FA$	-	-
		Heterozygous	$2MA + 2FA$	FA	$\frac{2MA + 2FA}{FA}$
Paternal trisomy		Homozygous	$2MA + 3FA$	-	-
		Heterozygous	$2MA + 2FA$	FA	$\frac{2MA + 2FA}{FA}$
Normal	Heterozygous	Homozygous	$MA + 2FA$	MA	$\frac{MA + 2FA}{MA}$
		Heterozygous	$MA + FA$	$MA + FA$	$\frac{MA + FA}{MA + FA}$
Maternal trisomy		Homozygous*	$MA + 3FA$	MA	$\frac{MA + 3FA}{MA}$
		Heterozygous	$MA + 2FA$	$MA + FA$	$\frac{MA + 2FA}{MA + FA}$
Paternal trisomy		Homozygous	$MA + 3FA$	MA	$\frac{MA + 3FA}{MA}$
		Heterozygous	$MA + 2FA$	$MA + FA$	$\frac{MA + 2FA}{MA + FA}$

2.2.4 Fetal-to-maternal cell ratio

Allele frequencies of a reference chromosome, that will unlikely be aneuploidy, were used to estimate the FMCR. For that I counted allele frequencies of the simulation of normal fetus and filtered the SNPs where mother is homozygous (allelic ratio >2.5). In this subset, the major allele frequency includes three sets of reads – two equal parts from the mother and one smaller part from the fetus. The minor allele frequency includes only one set of reads from the fetus that is equal with the fetal part of the major allele frequency (Table 5). Based on the previous theory, I extracted the estimated maternal and fetal allele frequencies from the major and the minor allele frequencies and calculate the estimated FMCR as follows:

$$FMCR = \frac{2 * \sum_{i=1}^n (\text{minor allele frequency}_i)}{\sum_{i=1}^n (\text{minor allele frequency}_i + \text{major allele frequency}_i)}$$

where i represents the observed locus and n represents the total number of observed SNPs. Using the estimated FMCR (~0.100), which was almost ideally confirmed by the FMCR set in simulations (0.1), I calculated the expected allelic ratios in case of every possible outcome. The expected allelic ratios in case of different FMCR are visualized on Figure 3.

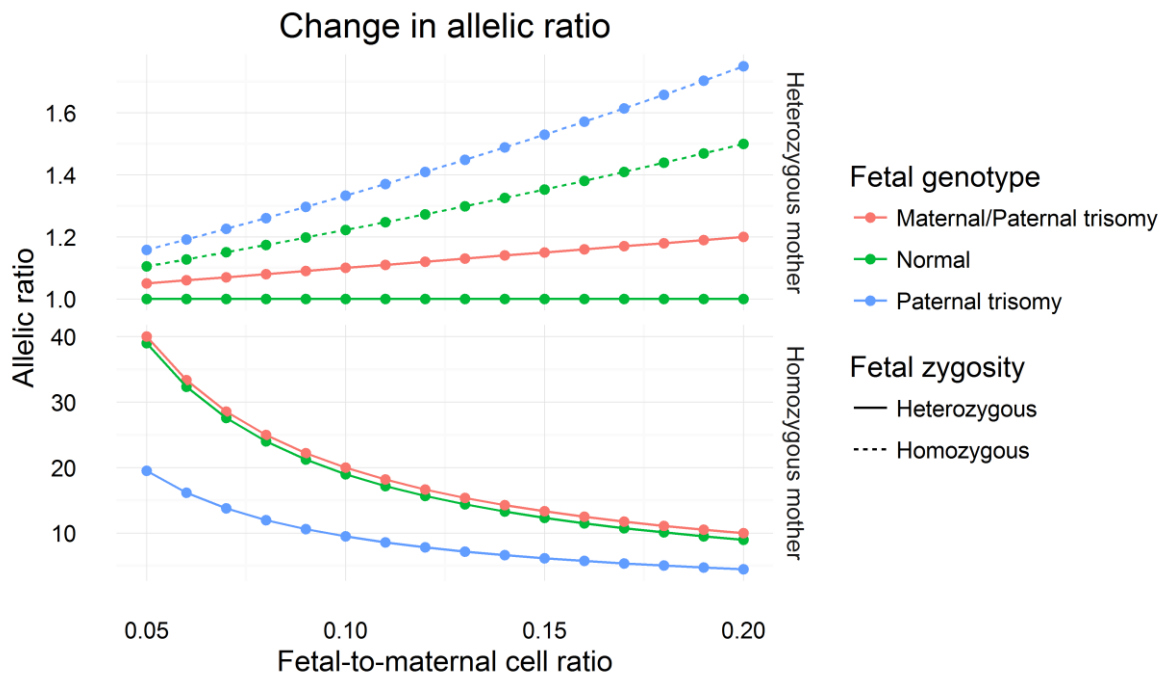


Figure 3. Relationship between allelic ratios of different fetal genotypes and FMCR. Upper panel represents maternal heterozygosity and lower panel represents maternal homozygosity. The dashed line represents fetal homozygosity and the solid line represents fetal heterozygosity.

2.2.5 Analytical approach

All calculations were performed in R software (version 3.3.0). For t -test based approach I used Welch’s t -test from “The R Stats Packages” (stats version 3.3.0) with default parameters which in case of two input vectors applies two-sided unpaired t -test expecting unequal variance and significance level of 0.05.

For HMM I used “Dependent Mixture Models – Hidden Markov Models of GLMs and Other Distributions in S4” (depmixS4 version 1.3-3) package. I created a 4-state HMM to separate allelic ratios by the combination of fetal zygosity and genotype – two hidden states for normal and two for trisomy (Figure 4). The HMM is designed for maternal heterozygosity only (Figure 3). The expected allelic ratios (Table 5) according to the estimated FMCR were used as the

means of observations associated with the states and the standard deviations of observations were fix to 0.05 in order to avoid excessive overlap. The transition probabilities of hidden states were symmetrical – 0.49 between two normal states or two trisomies and 0.01 between normal and trisomy. The initial hidden state probabilities were chosen according to the live birth prevalence of DS – 1 in 700 (Parker et al. 2010). Finally, Viterbi algorithm was used to find the most likely sequence of hidden states.

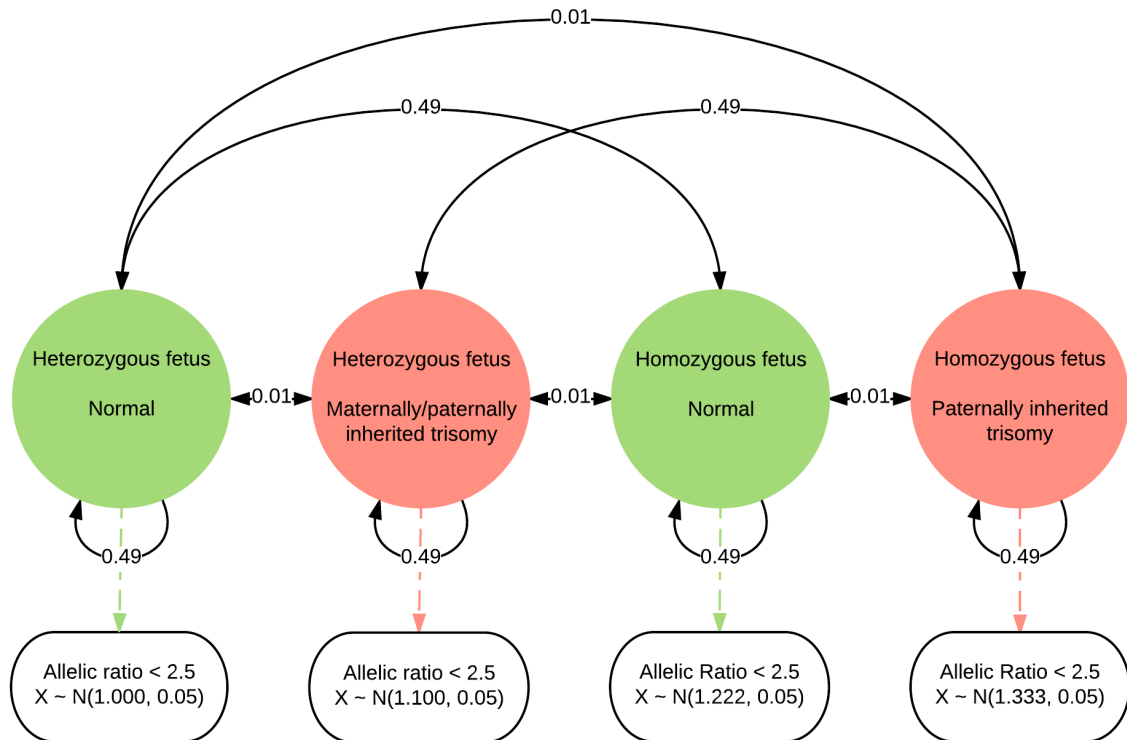


Figure 4. HMM. The circles represent hidden states and the ovals represent the distribution of expected allelic ratios. The black arrows with probabilities represent the transitions between the hidden states.

2.3 Results

Simulated data was generated per each fetal genotype to call informative SNPs (Table 6). Comparing the results of simulations with the theoretical calculations of informative SNPs described in chapter 2.2.2, we see that the proportion of informative SNPs out of the total number of SNPs differs very little from what is expected – approximately 1% at most. This assures that the simulations are conducted in accordance with the theory.

Table 6. Summary of simulated SNPs. Includes simulations of three fetal genotypes – normal, maternal and paternal trisomy.

Simulated fetal genotype	Number of SNPs in total	Informative SNPs					
		Maternal zygosity				Total	
		Homozygous		Heterozygous			
		N	%	N	%	N	%
Normal	440,000	110,115	25.03	219,458	49.88	329,573	74.90
Maternal trisomy		110,071	25.02	221,355	50.31	331,426	75.32
Paternal trisomy		164,975	37.49	219,882	49.97	384,857	87.47

Comparing the three simulations, I distinguish that the distributions of allelic ratios of informative SNPs have different patterns depending on the fetal genotype (Figure 5). Allelic ratio is mainly influenced by the maternal zygosity and less by the fetal zygosity, as most of the cfDNA in blood belongs to the mother (Wang et al. 2013). On that basis, allelic ratio can be divided into two distinct groups by maternal zygosity. In case of 1/10 FMCR, the allelic ratio of a SNP where mother is heterozygous is close to 1 (Figure 8). If the mother is homozygous and the heterozygosity derives from the paternally inherited chromosome(s) of the fetus, then the allelic ratio is nearly 10 times higher depending on the parental origin of the fetal trisomy (Figure 6).

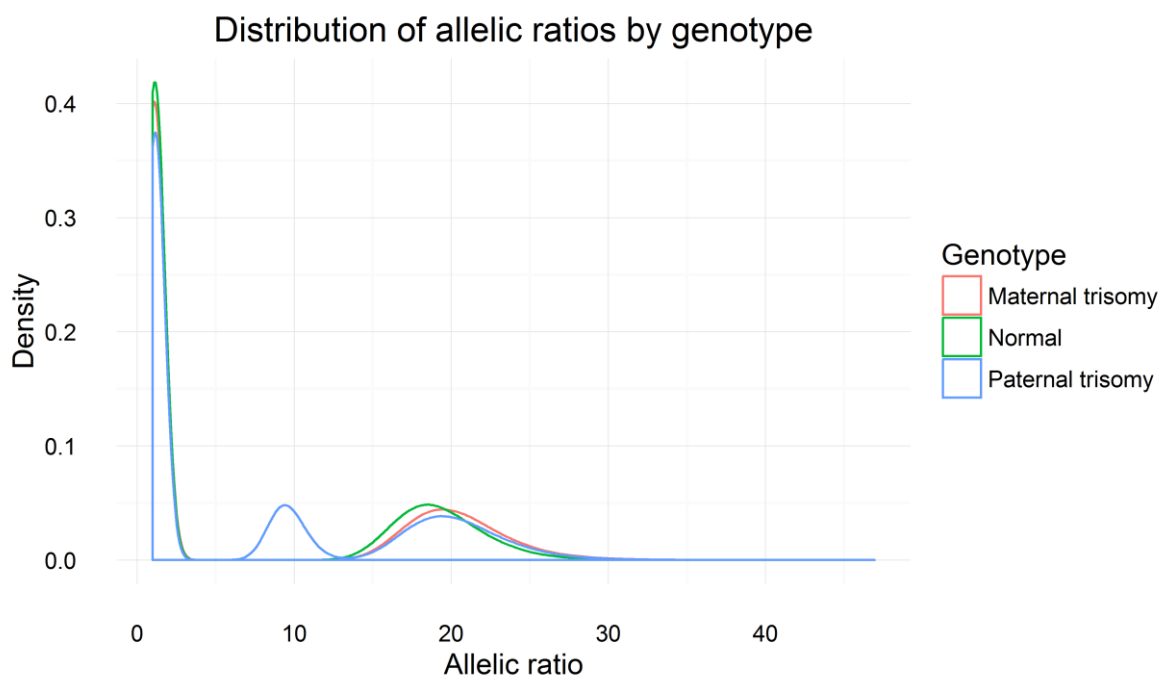


Figure 5. Distribution of allelic ratios. Simulated 440 000 SNPs per genotype and 1 000 reads per SNP in average. The SNPs of normal fetus, fetus with maternal trisomy and fetus with paternal trisomy were converted into respectively 329 573, 331 426 and 384 857 allelic ratios. Allelic ratios close to 1 represent maternal heterozygosity and allelic ratios close to 9.5, 19 and 20 represent maternal homozygosity. The simulations are conducted in 1/10 FMCR.

2.3.1 Detection of trisomy with *t*-test

The calculations based on the expected allelic ratios (Table 5) show that in case of homozygous mother and 1/10 FMCR (Figure 6) there is a ~5% difference between allelic ratio of normal fetus ($\mu = 19$) and fetus with maternally inherited trisomy ($\mu = 20$). The difference increases slowly with the growth of the FMCR (Figure 3). Paternally inherited trisomy, however, has two different scenarios. If only one paternally inherited fetal allele is different from the maternal alleles then the distribution of allelic ratio is identical to maternally inherited trisomy, if both alleles are different then the allelic ratio is 50% smaller ($\mu = 9.5$) than in case of a normal fetus.

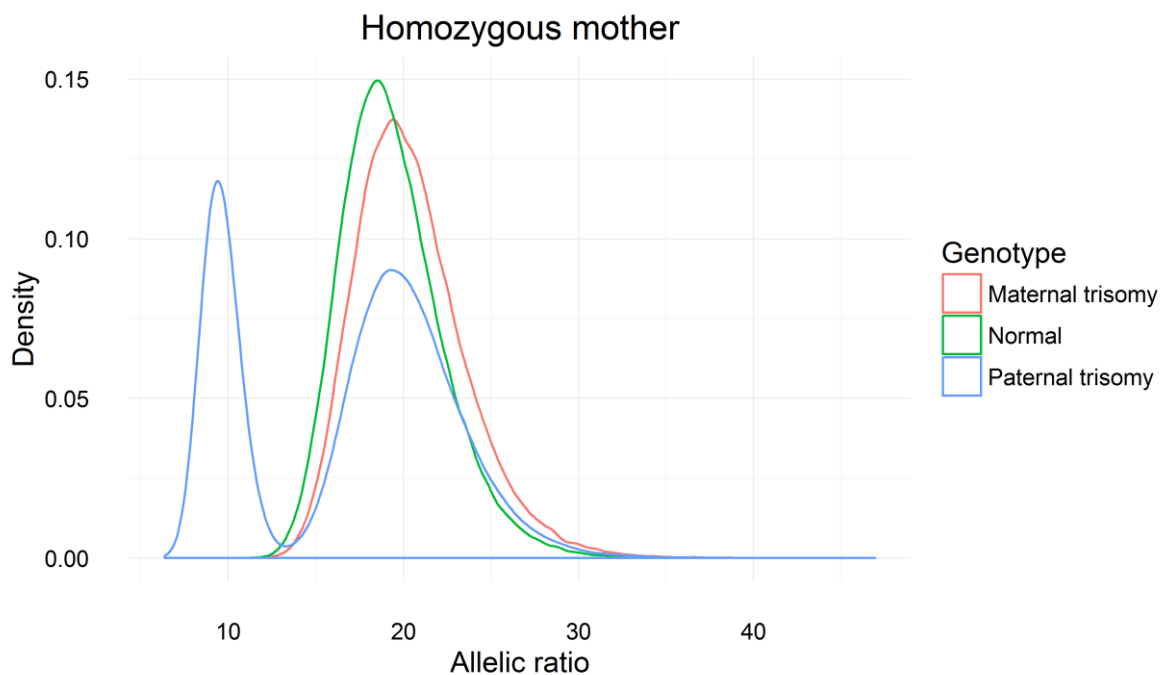


Figure 6. Distribution of allelic ratio typical to maternal homozygosity. Allelic ratio is over 2.5 in ~33% of all informative SNPs of fetus with normal genotype, ~33% of fetus with maternal trisomy and ~43% of fetus with paternal trisomy.

Based on the previously described differences of allelic ratios, I wanted to know how many SNPs are necessary for the detection of trisomy with two-sided Welch's *t*-test assuming heteroscedasticity. Using the simulated data, I compared the allelic ratios of normal fetus with the allelic ratios of fetus with maternally inherited trisomy by conducting a series of two-sample *t*-tests with variable number of informative SNPs. The statistical testing involved allelic ratios of informative SNPs typical to maternal homozygosity (Figure 6). As a result, I found that on average, at least 70 informative SNPs are necessary to detect full maternal trisomy (Figure 7). Considering the fact that approximately 1/4 of all SNPs are informative in case of homozygous mother and normal fetus or fetus with maternally inherited trisomy in our simulation (Table 6), then in total about 300 highly polymorphic SNPs are required to detect full maternal trisomy.

The allelic ratios of paternally inherited trisomy differentiate even more from the allelic ratios of normal fetus than the allelic ratios of maternally inherited trisomy, but as the distribution of the allelic ratios of paternally inherited trisomy is a bimodal distribution with two separate peaks, then the assumption of normal distribution is not met and using *t*-test is not optimal.

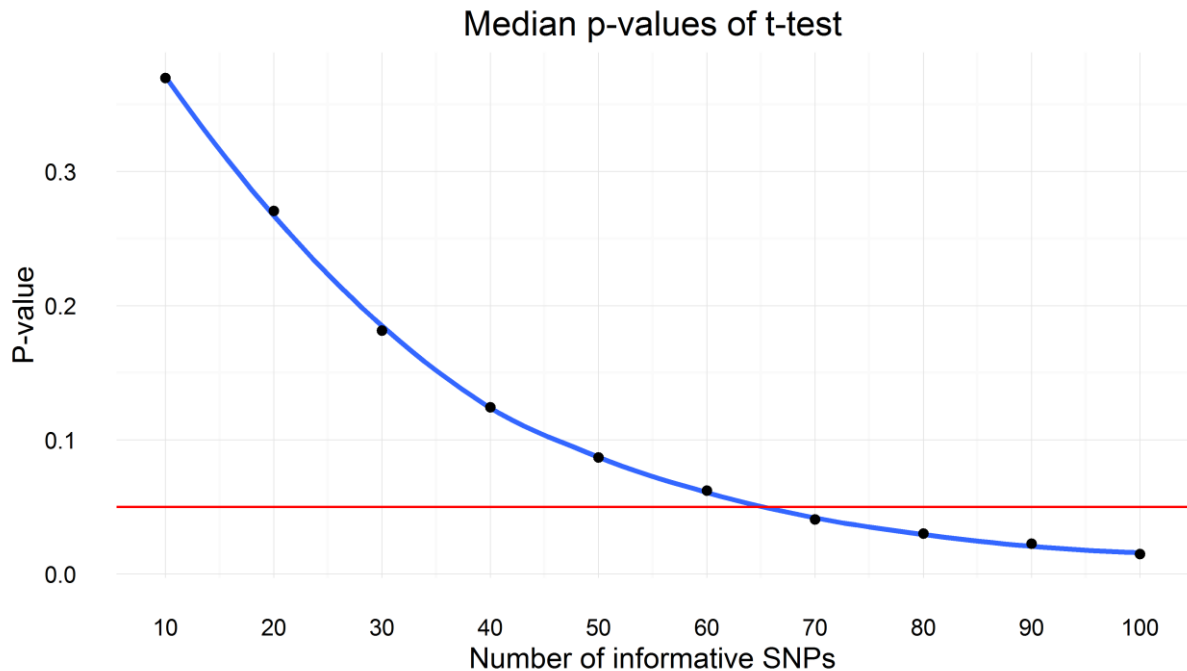


Figure 7. Relationship between number of informative SNPs and *t*-test p-value. Each point represents the median of p-values of 10 000 conducted *t*-tests between equal number of informative SNPs of allelic ratios of normal fetus and fetus with maternal trisomy. The blue line represents the local polynomial regression fitting (loess) curve and the red line represents the selected significance cut-off ($\alpha = 0.05$).

2.3.2 Detection of trisomy with hidden Markov model

The allelic ratios of informative SNPs in cfDNA are distributed between four distinct intervals in case of maternal heterozygosity (Figure 8). The four groups are formed by the combination of fetal zygosity and genotype as follows with expected allelic ratio:

- Heterozygous fetus with normal genotype ($\mu = 1$)
- Heterozygous fetus with maternal or paternal trisomy ($\mu = 1.1$)
- Homozygous fetus with normal genotype ($\mu = \sim 1.2$)
- Homozygous fetus with paternal trisomy ($\mu = \sim 1.3$)

The difference between the expected allelic ratios in adjacent groups are close to $\sim 10\%$ and increase with the increase of FMCR (Figure 3).

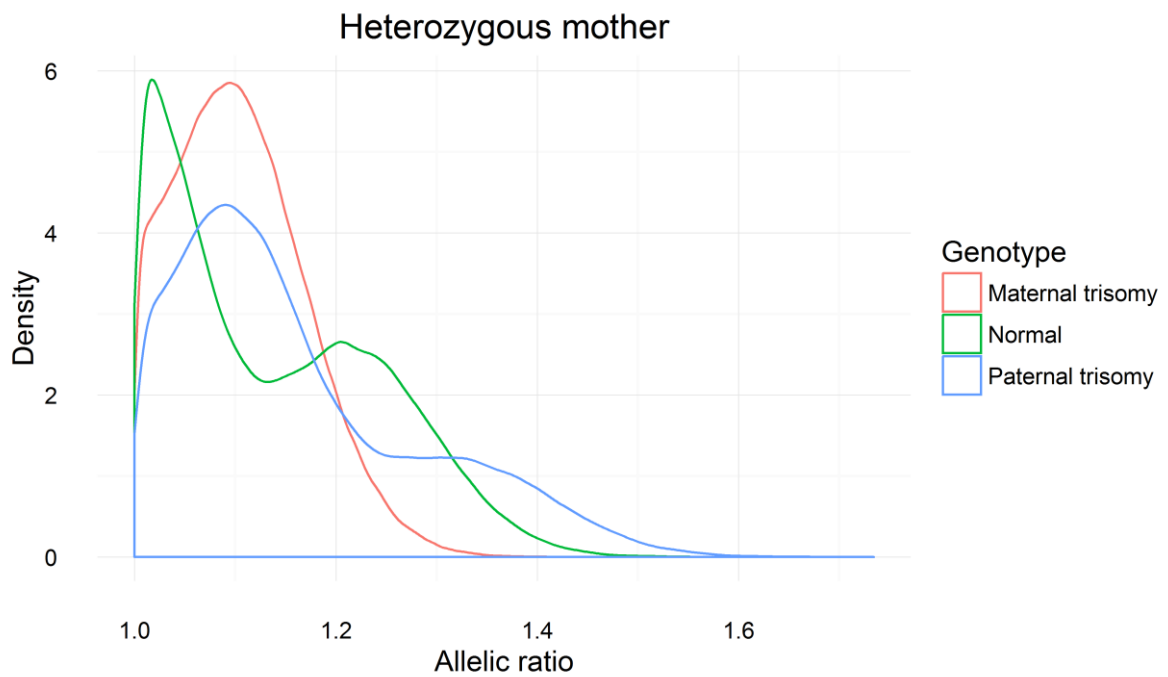


Figure 8. Distribution of allelic ratio typical to maternal heterozygosity. Allelic ratio is less than 2.5 in ~67% of all informative SNPs of fetus with normal genotype, ~67% of fetus with maternal trisomy and ~57% of fetus with paternal trisomy.

I used allelic ratios of simulated fetal genotypes (Figure 8) to estimate fetal disomy or trisomy by HMM. Close to 220 000 informative SNPs per simulation were divided between previously described states (Figure 4). The results reveal the true underlying genotypes behind the simulated fetal genotypes in all three occasions. In total, 63.37% of informative SNPs of normal fetal genotype, 83.99% of maternal trisomy and 96.48% of paternal trisomy are classified correctly (Table 7). As FMCR increases, the differences between the allelic ratios also increase which should ease the detection of underlying fetal genotype of SNPs (Figure 3).

Table 7. The estimated fetal genotype by HMM. Includes simulations of three fetal genotypes – normal, maternal and paternal trisomy.

Simulated fetal genotype	Number of SNPs	Estimated fetal genotype (%)			
		Normal	Trisomy		
			Maternal/Paternal	Paternal	Total
Normal	219,458	63.37	26.18	10.44	36.63
Maternal trisomy	221,355	11.58	83.99	4.43	88.42
Paternal trisomy	219,882	3.52	69.95	26.53	96.48

To illustrate the performance of the HMM analysis in patient-specific cases, I took the first 500 sequential informative SNPs and performed an analysis using the previously defined HMM. The results were similar to the previously obtained outcomes, confirming that the HMM

analysis with selected parameters performs as expected in case of realistic number of SNPs (Figure 9).

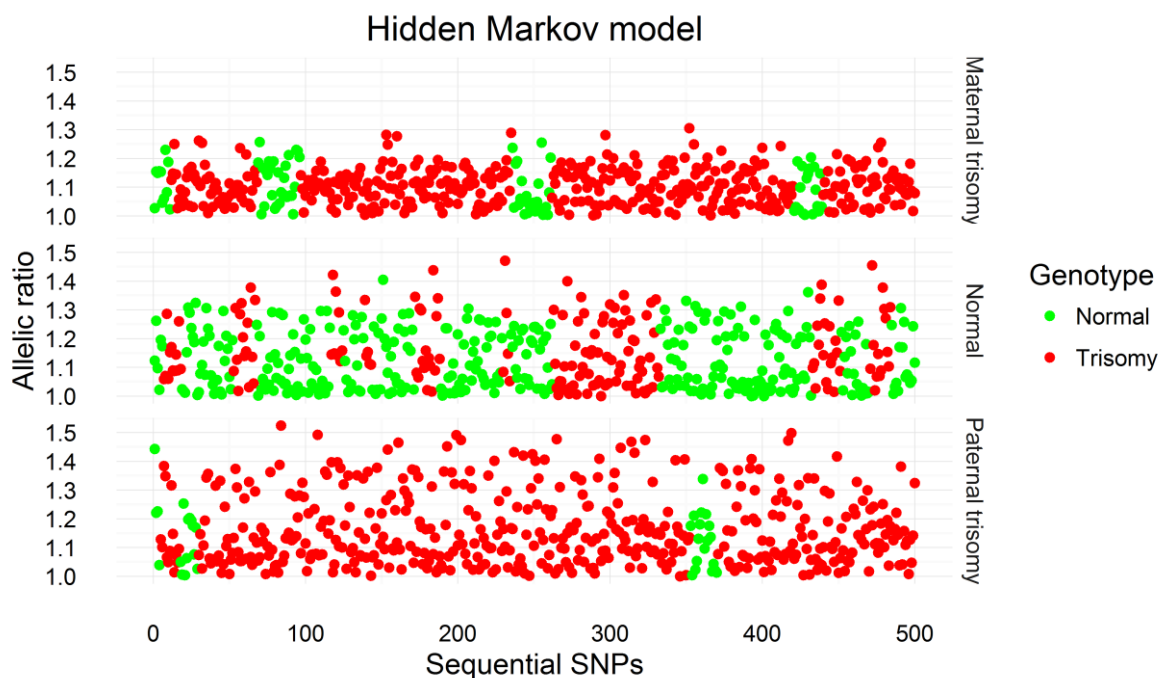


Figure 9. Estimated fetal genotypes of sequential SNPs by HMM. Estimation incorporated allelic ratios of 500 sequential SNPs per fetal genotype. Simulation of maternal trisomy (upper panel) resulted in ~17% (84) of normal genotype (green) and ~83% (416) of trisomy (red), simulation of normal fetus (center panel) resulted in 67% (335) of normal and 33% (165) of trisomy and simulation of paternal trisomy (lower panel) resulted in ~7% (34) of normal and ~93% (466) of trisomy.

2.4 Discussion

There are several NIPT methods available for prenatal screening of fetal aneuploidy. Compared with the first-generation NIPT algorithms, the SNP-based analytical approach has several advantages – it does not require sequencing of the whole genome nor does it require prior knowledge of parental genotypes to detect fetal trisomy (Norwitz & Levy 2013). Targeting only pre-defined regions, as methods described in chapter 1.4, enables to sequence multiple patients in parallel, which keeps the sequencing cost low and makes it a promising alternative to current methods. In addition, the described method can discover the parental origin of fetal trisomy, which simplifies the identification of underlying problem and allows to focus on finding a patient-specific solution.

There are also a few shortcomings. Firstly, the approach has only been tested on simulated data. In case of real data, greater variation in allele frequencies may be expected, which may fuse the difference between the allelic ratios of normal fetus and fetus with trisomy, or even encounter unexpected distribution of allelic ratios, which are difficult to interpret. Secondly, it is designed for detection of common autosomal trisomies which form a major proportion of aneuploidies.

Competing methods can also detect other chromosomal abnormalities like monosomies, sex chromosome disorders, microdeletions and microduplications (Samango-Sprouse et al. 2013). Some NIPT studies even claim to be able to detect fetal aneuploidies in twin pregnancies (Huang et al. 2014).

As this method is still in active development phase, there are several ideas how to improve the accuracy. At first, HMM parameters can be optimized according to the real data. Also, it is possible to use even chromosome-specific HMM models, which could take into account maternal age, prevalence of the disease in general or specific populations and other traits which are correlated or have been associated with chromosomal aneuploidies.

Finding a sufficient quantity of highly polymorphic SNPs that are evenly distributed over the whole studied chromosome is crucial for this approach. Approximately 3/4 of highly polymorphic SNPs are informative by random distribution of parental alleles and therefore useful for the analysis (Table 6). Lower amount of SNPs could lead to unreliable results due to the greater uncertainty. Possibly, this could be solved by increasing read depth of coverage in sequencing or by increasing the number of analyzed SNPs.

At the moment, the 4-state HMM was adjusted for maternal heterozygosity and used approximately 2/3 of the informative SNPs, which accounts for approximately 1/2 of all SNPs (Table 6). Adding three extra states to the HMM, which are typical to maternal homozygosity, would incorporate all the informative SNPs. The shortage of the 7-state model is that the difference between the allelic ratios of these extra states is smaller and it would be harder to differentiate fetal genotypes. Alternatively, it would be possible to use higher amount of SNPs and the 3-state model.

Viterbi algorithm is used by default to estimate the most likely fetal genotype underlying each SNP. Although being optimal for obtaining the maximum probability, Viterbi algorithm does not minimize the number of expected classification errors. Relatively fewer expected errors could result from using “pointwise maximal a posterior” (PMAP) estimator which has its own disadvantages. Having a false negative result is less preferable than obtaining a false positive result as such screening result is confirmed or disconfirmed later by a diagnostic method. Unfortunately, PMAP may have very low posterior probability. A solution could be a hybrid-estimator that combines the properties of previously described estimators (Lember & Koloydenko 2014). Further testing has to be carried out in this matter.

One way to improve the accuracy of such analytical approach may be to use a combination of different statistical tests. Combining the 3-state HMM, which fits for allelic ratios where mother is heterozygous, and the *t*-test, which is suitable for allelic ratios where mother is homozygous,

takes an advantage of a larger number of SNPs than either test separately. The combination of the two tests may provide a more reliable result in case of fewer informative SNPs.

Further, we plan to add new features to the current analytical approach. First, real data is needed. Next task is to expand the theoretical calculations of allele frequencies for monosomy and sex chromosome aneuploidies. If characteristics of allelic ratios that are unique for a certain condition can be described, similar concept may be applied.

The HMM is based on Markov chain, which in the model means that the next SNP depends only on the current SNP and not on the SNPs that preceded it. Having pre-defined SNPs, we can easily locate them on a chromosome and find the distance between SNPs. If two sequential SNPs are situated closer together, they are more likely to be in the same state, whether it is a normal or an aneuploidy. Relying on the previous statement, the plan is to detect partial chromosomal abnormalities where only a certain part of the chromosome is duplicated or deleted. This can be done by splitting the chromosome virtually into consecutive bins, followed by bin by bin analysis and chromosomal state determination.

As the approach described here is based on allelic ratios of observed parental alleles, then it should be also suitable for detection of UPD. In case of isodisomy, in which a single pair of duplicated chromosomes are inherited, we expect to see high level of fetal homozygosity in observed chromosome.

In conclusion, analysis of only hundreds of highly polymorphic SNPs over risk- and reference chromosomes has a high potential to make NIPT laboratory analysis and computational methods much more cost- and resource effective. It enables to reduce the price of NIPT, making it available for all pregnant women in an early phase of pregnancy. As my study was conducted on simulated data, which provided sufficient results for distinguishing normal fetus and fetus with maternally or paternally inherited trisomy, highly quantitative genotyping results from real trisomy cases are needed to prove the concept described in this study.

SUMMARY

Using NIPT to detect fetal chromosomal aneuploidies from mother's blood sample has proven to be an accurate and reliable method. In this study, the only aneuploidy we focused on was trisomy. Using three sets of simulated data, which consisted of a combination of maternal and one of the fetal genotypes – normal, maternally and paternally inherited trisomy. Comparing the simulated allelic ratios of informative SNP with the expected allelic ratios revealed that the detection of fetal trisomy is theoretically possible.

However, NIPT is considered as screening method, which means that for diagnosis, patient needs to undergo an invasive procedure to be fully confirmed. As the accuracy of NIPT is exceptional, decreasing number of pregnant women need to be mistakenly examined by invasive techniques, which pose ~1% risk of miscarriage. Also, accurate prenatal screening makes difficult decision in case of possible chromosomal pathology easier for parents.

The second aim of the study was to determine a parental origin if the trisomy is confirmed. Using the HMM, we managed to estimate parental origin even though the allelic ratios of maternally and paternally inherited trisomies overlap in some cases. Knowing the parental origin may be important in case of infertility problems by narrowing down the possible causes, which helps physicians to focus on finding a patient-specific solution.

In conclusion, the highly quantitative SNP-based approach, that is under development by our research group, has theoretically a good potential to meet the need of being a cost-effective alternative to current NIPT method. However, further testing has to be carried out with the real data.

KOKKUVÕTE

NIPT on tõestanud end kui täpse ja usaldusväärse meetodina loote kromosomaalsete aneuploidiate detekteerimiseks. Uurimuses keskendusime vaid trisoomia tuvastamisele, kasutades kolme simuleeritud andmestikku, mis koosnesid ema ja ühe võimaliku loote genotüübi – normaalne, ema- ja isapoolne trisoomia – andmetest. Simuleeritud informatiivsete SNP-de alleelsete suhete võrdlemine oodatud alleelsete suhetega näitas, et loote trisoomia detekteerimine on teoreetiliselt võimalik.

NIPT puhul on tegemist skriininmeetodiga, mis tähendab, et positiivse testitulemuse kinnitamiseks peab patsient läbima invasiivse diagnostilise protseduuri. Kuna NIPT on erakordselt täpne, siis järjest vähem rasedaid vajab ekslikult invasiivset kontrolli, millega kaasneb ~1% nurisünnituse risk. Lisaks lihtsustab täpne prenataalne skriining kõrge riskihinnangu korral vanematel langetada raskeid otsuseid.

Töö teine eesmärk oli trisoomia vanemliku päritolu määramine. HMM analüüsi rakendades õnnestus meil hinnata trisoomia vanemlikku päritolu hoolimata sellest, et ema- ja isapoolse päritoluga trisoomiate alleelisuhetes on osaline kattuvus. Trisoomia vanemliku päritolu teadmine võib osutada oluliseks viljakusprobleemide lahendamisel, vähendades võimalike põhjuste koguhulka, mis võimaldab arstidel keskenduda patsiendispetsiifiliste probleemide lahendamisele.

Kokkuvõtteks on antud SNP-põhisel lähenemisel, mida arendatakse hetkel aktiivselt meie grupi poolt, hea teoreetiline eeldus olla hinnasõbralikum alternatiiv praegustele NIPT meetoditele. Siiski on lõpliku kindluse saavutamiseks vaja analüüsi kontrollida reaalsete andmetega.

ACKNOWLEDGEMENTS

Epp Kaleviste

Anne Mari Roost

Maria Teder

Kristi Kuljus

Priit Palta

Priit Paluoja

Jüri Lember

Priit Adler

Kaarel Krjutškov

Prof. Ants Kurg

REFERENCES

- Alberry, M. et al., 2007. Free fetal DNA in maternal plasma in anembryonic pregnancies: confirmation that the origin is the trophoblast. *Prenatal Diagnosis*, 27(5), pp.415–418. Available at: <http://doi.wiley.com/10.1002/pd.1700> [Accessed May 24, 2016].
- Ashoor, G. et al., 2013. Fetal fraction in maternal plasma cell-free DNA at 11-13 weeks' gestation: relation to maternal and fetal characteristics. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 41(1), pp.26–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23108725> [Accessed April 13, 2016].
- Ayse Kirbas, Korkut Daglar, N.D., 2016. Non-Invasive Prenatal Testing for Aneuploidy: A Review of the Literature. *Medicine Science / International Medical Journal*, p.1. Available at: <http://www.scopemed.org/?mno=214006> [Accessed April 2, 2016].
- Bayindir, B. et al., 2015. Noninvasive prenatal testing using a novel analysis pipeline to screen for all autosomal fetal aneuploidies improves pregnancy management. *European journal of human genetics : EJHG*, 23(10), pp.1286–93.
- Cheslock, P.S. et al., 2005. The roles of MAD1, MAD2 and MAD3 in meiotic progression and the segregation of nonexchange chromosomes. *Nature genetics*, 37(7), pp.756–60.
- Chiang, T., Schultz, R.M. & Lampson, M.A., 2012. Meiotic origins of maternal age-related aneuploidy. *Biology of reproduction*, 86(1), pp.1–7.
- Chitayat, D. et al., 2011. Prenatal Screening for Fetal Aneuploidy in Singleton Pregnancies. *Journal of Obstetrics and Gynaecology Canada*, 33(7), pp.736–750.
- Coonen, E. et al., 2004. Anaphase lagging mainly explains chromosomal mosaicism in human preimplantation embryos. *Human reproduction (Oxford, England)*, 19(2), pp.316–24.
- Delhanty, J.D.A. et al., 1993. Detection of aneuploidy and chromosomal mosaicism in human embryos during preimplantation sex determination by fluorescent in situ hybridisation, (FISH). *Human Molecular Genetics*, 2(8), pp.1183–1185.
- Fan, H.C. et al., 2008. Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proceedings of the National Academy of Sciences of the United States of America*, 105(42), pp.16266–71. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18838674> [Accessed May 24, 2016].
- Fragouli, E., Wells, D. & Delhanty, J.D.A., 2011. Chromosome abnormalities in the human oocyte. *Cytogenetic and genome research*, 133(2-4), pp.107–18. Available at: <http://www.karger.com/doi/10.1159/000323801> [Accessed April 26, 2016].
- Ghosh, S. et al., 2010. Telomere length is associated with types of chromosome 21

- nondisjunction: a new insight into the maternal age effect on Down syndrome birth. *Human genetics*, 127(4), pp.403–9.
- Gil, M.M. et al., 2015. Analysis of cell-free DNA in maternal blood in screening for fetal aneuploidies: updated meta-analysis. *Ultrasound in obstetrics & gynecology : the official journal of the International Society of Ultrasound in Obstetrics and Gynecology*, 45(3), pp.249–66.
- Hall, H.E. et al., 2007. The origin of trisomy 22: Evidence for acrocentric chromosome-specific patterns of nondisjunction. *American Journal of Medical Genetics Part A*, 143A(19), pp.2249–2255.
- Hassold, T. et al., 1995. Recombination and maternal age-dependent nondisjunction: molecular studies of trisomy 16. *American journal of human genetics*, 57(4), pp.867–74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7573048> [Accessed May 25, 2016].
- Hassold, T., Hall, H. & Hunt, P., 2007. The origin of human aneuploidy: where we have been, where we are going. *Human Molecular Genetics*, 16(R2), pp.R203–R208.
- Hassold, T. & Hunt, P., 2001. To err (meiotically) is human: the genesis of human aneuploidy. *Nature Reviews Genetics*, 2(4), pp.280–291.
- Huang, X. et al., 2014. Noninvasive prenatal testing of trisomies 21 and 18 by massively parallel sequencing of maternal plasma DNA in twin pregnancies. *Prenatal Diagnosis*, 34(4), pp.335–340. Available at: <http://doi.wiley.com/10.1002/pd.4303> [Accessed May 22, 2016].
- Hultén, M.A. et al., 2010. On the origin of the maternal age effect in trisomy 21 Down syndrome: the Oocyte Mosaicism Selection model. *Reproduction (Cambridge, England)*, 139(1), pp.1–9.
- Iwarsson, E., Kvist, U. & Hultén, M.A., 2015. Disomy 21 in spermatozoa and the paternal origin of trisomy 21 Down syndrome. *Molecular cytogenetics*, 8, p.67.
- Juneau, K. et al., 2014. Microarray-Based Cell-Free DNA Analysis Improves Noninvasive Prenatal Testing. *Fetal Diagnosis and Therapy*, 36(4), pp.282–286. Available at: <http://www.karger.com?doi=10.1159/000367626>.
- Kotsopoulou, I. et al., 2015. Non-invasive prenatal testing (NIPT): limitations on the way to become diagnosis. *Diagnosis*, 2(3), pp.141–158. Available at: <http://orcid.org/0000-0002-5876-2599> [Accessed May 21, 2016].
- Kouznetsova, A. et al., 2007. Bi-orientation of achiasmatic chromosomes in meiosis I oocytes contributes to aneuploidy in mice. *Nature genetics*, 39(8), pp.966–8.
- Lamb, N., 1997. Characterization of susceptible chiasma configurations that increase the risk for maternal nondisjunction of chromosome 21. *Human Molecular Genetics*, 6(9),

- pp.1391–1399.
- Lamb, N.E., Sherman, S.L. & Hassold, T.J., 2005. Effect of meiotic recombination on the production of aneuploid gametes in humans. *Cytogenetic and genome research*, 111(3-4), pp.250–5.
- Langlois, S. et al., 2006. Postnatal follow-up of prenatally diagnosed trisomy 16 mosaicism. *Prenatal Diagnosis*, 26(6), pp.548–558. Available at: <http://doi.wiley.com/10.1002/pd.1457> [Accessed May 25, 2016].
- Lember, J. & Koloydenko, A.A., 2014. Bridging Viterbi and Posterior Decoding: A Generalized Risk Approach to Hidden Path Inference Based on Hidden Markov Models. *Journal of Machine Learning Research*, 15, pp.1–58.
- Liao, G.J.W. et al., 2011. Targeted Massively Parallel Sequencing of Maternal Plasma DNA Permits Efficient and Unbiased Detection of Fetal Alleles. *Clinical Chemistry*, 57(1), pp.92–101. Available at: <http://www.clinchem.org/cgi/doi/10.1373/clinchem.2010.154336> [Accessed May 25, 2016].
- Lo, Y.M. et al., 1999. Rapid clearance of fetal DNA from maternal plasma. *American journal of human genetics*, 64(1), pp.218–24.
- Lo, Y.M. et al., 1996. Two-way cell traffic between mother and fetus: biologic and clinical implications. *Blood*, 88(11), pp.4390–5.
- Lo, Y.M.D. et al., 1997. Presence of fetal DNA in maternal plasma and serum. *The Lancet*, 350(9076), pp.485–487.
- Lo, Y.M.D. et al., 1998. Quantitative Analysis of Fetal DNA in Maternal Plasma and Serum: Implications for Noninvasive Prenatal Diagnosis. *The American Journal of Human Genetics*, 62(4), pp.768–775.
- Muller, F. et al., 2000. Parental origin of the extra chromosome in prenatally diagnosed fetal trisomy 21. *Human Genetics*, 106(3), pp.340–344.
- Nagaoka, S.I., Hassold, T.J. & Hunt, P.A., 2012. Human aneuploidy: mechanisms and new insights into an age-old problem. *Nature reviews. Genetics*, 13(7), pp.493–504.
- Nicolaidis, P. & Petersen, M.B., 1998. Origin and mechanisms of non-disjunction in human autosomal trisomies. , 13(2).
- Norwitz, E.R. & Levy, B., 2013. Noninvasive prenatal testing: the future is now. *Reviews in obstetrics & gynecology*, 6(2), pp.48–62. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24466384> [Accessed May 23, 2016].
- Palomaki, G.E. et al., 2011. DNA sequencing of maternal plasma to detect Down syndrome: an international clinical validation study. *Genetics in medicine: official journal of the*

- American College of Medical Genetics*, 13(11), pp.913–20.
- Papavassiliou, P. et al., 2015. Mosaicism for trisomy 21: a review. *American journal of medical genetics. Part A*, 167A(1), pp.26–39.
- Parker, S.E. et al., 2010. Updated National Birth Prevalence estimates for selected birth defects in the United States, 2004-2006. *Birth defects research. Part A, Clinical and molecular teratology*, 88(12), pp.1008–16. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20878909> [Accessed May 12, 2016].
- Pellestor, F., Anahory, T. & Hamamah, S., 2005. The chromosomal analysis of human oocytes. An overview of established procedures. *Human reproduction update*, 11(1), pp.15–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15569701> [Accessed May 26, 2016].
- Ramasamy, R. et al., 2015. Fluorescence in situ hybridization detects increased sperm aneuploidy in men with recurrent pregnancy loss. *Fertility and sterility*, 103(4), pp.906–909.e1.
- Robinson, W.P. et al., 1995. Molecular studies of chromosomal mosaicism: relative frequency of chromosome gain or loss and possible role of cell selection. *American journal of human genetics*, 56(2), pp.444–51.
- Rockmill, B., Voelkel-Meiman, K. & Roeder, G.S., 2006. Centromere-proximal crossovers are associated with precocious separation of sister chromatids during meiosis in *Saccharomyces cerevisiae*. *Genetics*, 174(4), pp.1745–54.
- Russo, M.L. & Blakemore, K.J., 2014. A historical and practical review of first trimester aneuploidy screening. *Seminars in Fetal and Neonatal Medicine*.
- Samango-Sprouse, C. et al., 2013. SNP-based non-invasive prenatal testing detects sex chromosome aneuploidies with high accuracy. *Prenatal Diagnosis*, 33(7), pp.643–649.
- Sherman, S.L., Lamb, N.E. & Feingold, E., 2006. Relationship of recombination patterns and maternal age among non-disjoined chromosomes 21. *Biochemical Society transactions*, 34(Pt 4), pp.578–80.
- Sparks, A.B. et al., 2012. Selective analysis of cell-free DNA in maternal blood for evaluation of fetal trisomy. *Prenatal Diagnosis*, 32(1), pp.3–9.
- Stokowski, R. et al., 2015. Clinical performance of non-invasive prenatal testing (NIPT) using targeted cell-free DNA analysis in maternal plasma with microarrays or next generation sequencing (NGS) is consistent across multiple controlled clinical studies. *Prenatal Diagnosis*, 35(12), pp.1243–1246. Available at: <http://doi.wiley.com/10.1002/pd.4686> [Accessed May 23, 2016].
- Zimmermann, B. et al., 2012. Noninvasive prenatal aneuploidy testing of chromosomes 13, 18, 21, X, and Y, using targeted sequencing of polymorphic loci. *Prenatal diagnosis*, 32(13),

- pp.1233–41. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/23108718> [Accessed May 25, 2016].
- Taylor, T.H. et al., 2014. The origin, mechanisms, incidence and clinical consequences of chromosomal mosaicism in humans. *Human Reproduction Update*, 20(4), pp.571–581. Available at: <http://humupd.oxfordjournals.org/cgi/doi/10.1093/humupd/dmu016> [Accessed April 15, 2016].
- Templado, C., Uroz, L. & Estop, A., 2013. New insights on the origin and relevance of aneuploidy in human spermatozoa. *Molecular human reproduction*, 19(10), pp.634–43.
- Uroz, L. & Templado, C., 2012. Meiotic non-disjunction mechanisms in human fertile males. *Human Reproduction*, 27(5), pp.1518–1524.
- Ustav, E., 2015. Uus võimalus sünnieelseks loote kromosoomihaiguste sõeluuringuks – loote rakuvaba DNA määramine ema verest. , 94(6), pp.352–357.
- Wang, E. et al., 2013. Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. *Prenatal Diagnosis*.
- Weijerman, M.E. & de Winter, J.P., 2010. Clinical practice. *European Journal of Pediatrics*, 169(12), pp.1445–1452. Available at: <http://link.springer.com/10.1007/s00431-010-1253-0> [Accessed May 25, 2016].
- Wright, C.F. & Chitty, L.S., 2009. Cell-free fetal DNA and RNA in maternal blood: implications for safer antenatal testing. *BMJ (Clinical research ed.)*, 339(jul06_2), p.b2451.

NON-EXCLUSIVE LICENCE

Non-exclusive licence to reproduce thesis and make thesis public

I, Hindrek Teder (date of birth: 19.01.1989),

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to:
 - 1.1. reproduce, for the purpose of preservation and making available to the public, including for addition to the DSpace digital archives until expiry of the term of validity of the copyright, and
 - 1.2. make available to the public via the web environment of the University of Tartu, including via the DSpace digital archives until expiry of the term of validity of the copyright,

Evaluation of the possibility to detect fetal chromosome trisomies based on a defined set of single nucleotide polymorphisms for non-invasive prenatal testing

supervised by Kaarel Krjutškov and Prof. Ants Kurg,

2. I am aware of the fact that the author retains these rights.
3. I certify that granting the non-exclusive licence does not infringe the intellectual property rights or rights arising from the Personal Data Protection Act.

Tartu, 27.05.2016